MJFAS
Malaysian Journal of Fundamental and Applied Sciences

RESEARCH ARTICLE

Geometric Morphometric and Pattern Discrimination of Handwritten Numeral Characters Based on Local Ethnicities and Native Linguistic Disparities in Malaysia for Forensic Applications

Wan Nurul Syafawani Wan Mohd Taufek[a], Helmi Mohd Hadi Pritam[a], Wan Nur Syuhaila Mat Desa[a], Dzulkiflee Ismail[a*], Naji Arafat Mahat[b,c,d,e]

[a]Forensic Science Programme, School of Health Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia; [b]Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia; [c]Investigative and Forensic Sciences Research Group, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia; [d]Centre for Sustainable Nanomaterials, Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia; [e]Centre of Research for Fiqh Forensics and Judiciary, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

Abstract Handwriting can be influenced by factors like culture and native language disparities; however, the specific research to discriminate handwritten numeral characters for forensic questioned document investigation involving the diverse ethnicities (Malay, Chinese and Indian) in Malaysian population remains unreported. Despite its application in forensic anthropology, utilization of the landmark based analysis using the Geometric Morphometric (GMM) technique for examining handwritten numeral character specimens appears sparse. Therefore, this present research that attempted to discriminate ethnicities background of 390 consenting adult participants in Malaysia based on the handwritten numeral characters (0-9) using GMM technique, merits forensic and scientific considerations. Results revealed that while there were universal ones, ethnic-specific handwritten numeral characters that may possibly be influenced by culture and native language (Malay, Chinese and Tamil languages) also prevailed. Significant differences ($p < 0.0001$) in ethnic-specific characters for the Malay, Chinese and Indian participants who attended the different national primary school settings in Malaysia were also observed. In conclusion, the discrimination of handwritten numeral characters for forensic examination of questioned documents among Malay, Chinese and Indian authors for the development of forensic intelligence is empirically supported.

Keywords: Forensic science, numeral handwriting, geometric morphometric, landmark-based analysis, forensic document examination.

***For correspondence:**
dzulkiflee@usm.my

Received: 13 June 2024
Accepted: 21 Aug. 2024

Introduction

Every individual's handwriting style is unique and variations can be influenced by multifaceted factors (1). The peculiarity of handwriting characteristics is analogous to the peculiarity of genetic markers or the uniqueness of ridge patterns on the fingertip in a particular individual (1). In addition, handwriting analysis could be used to complement other individual characteristic evidence like fingerprints and DNA. It has been indicated that handwriting habits and styles are strongly influenced by one's culture, native language and surroundings, given that handwriting style is developed during childhood (2–4). This is because of the cultural norms, prejudices and biases among ethnicities, resulting in diverse consequences in their handwriting systems (2). The culture plays a critical role in influencing a child's cognitive and motor developments related to behavioral skill such as handwriting, varying greatly during development in different culture of ethnicities (2). A person's handwriting comprises several individual and class characteristics that when combined, assigning a person to a particular group of ethnicities may prove feasible (5). As such, utilizations of handwriting class characteristics among mixed populations, such as Hispanics (born in Latin America) but now living in the United States (USA) (6) or Singapore (a multiculturalism country that comprised of Chinese, Malay and Indians where the cultures, ideologies, religions and passions meet) and recognized diversity (7) have been reported. In this context, embellishments, pen lifts, a broad range of letter forms and shape, letter spacing and writing movements have been seen in both Hispanic and Singaporean communities (6,7). Despite using the quantitative approach to accentuate the statistically significant prevalent of class characteristics of handwriting in a particular population, a specific study addressing inequities in class characteristics that may arise across generation gaps, particularly in groups who have migrated as well as raised children in various native languages and educational systems, remains unreported (6).

In Malaysia, the primary education systems are divided based on the medium of instruction used in teaching and learning, namely the Malay-medium national schools, vernacular schools of Chinese and Tamil, as well as Islamic religious schools (3). The general subjects taught at government-funded national primary schools are Malay, English, Chinese and Tamil as well as Religious Education. As a result, copybook writing style could be adapted throughout primary levels, and their influence on the creation and formation of one's handwriting, known as class characteristics, is of considerable relevance for forensic intelligence relating to handwritten evidence (3). The exploration of how educational systems in secondary schools could impact the development and quality of numeral handwriting skills is also limited in the study. However, it is important to consider the secondary school's influence on handwriting skills as secondary schools also emphasize structured handwriting instruction as part of their curriculum which tends to contribute positively to students' graphic maturity. The legibility, speed, fluency, and overall formation of characters in handwriting can adapt to different styles, particularly when students come from diverse cultural backgrounds (8). However, the incorporation of technology tools for handwriting skills in modern secondary schools, such as digital platforms or applications providing interactive exercises and assessments, can alter students' innate handwriting styles. This shift in handwriting style due to technological intervention is a significant reason why this present study only focused on primary schools' criteria alone.

In the past twenty years, handwritten character analysis has played an important role in forensic document examination (9). Handwriting analysis makes extensive use of the identification of class characteristics amongst populations of individuals for cultural discrimination and identification (2,4). A sample of a person's handwriting might reveal their ethnicities or nationalities (6). Authors from different ethnicities or nationalities, with different native languages, knowledges and taught writing systems could exhibit discriminatory handwriting characteristics (3,10). Despite the fact that handwriting style analysis has recently gained attention, specific research on accents, native languages and cultural influences on writing styles has been relatively limited (2,4). It is uncommon to discover studies on handwriting style that focus on finding class characteristics that are representative of a particular group of authors within a large mixed-population country like Malaysia, Brunei, Indonesia and Singapore, comprising predominantly by Malays, Chinese and Indians (2,10). Still, identifying an author's ethnicity (via their handwriting by relying on their copybook writing style), has been made challenging by the abundance of variances in handwriting class as well as individual characteristics in different countries, which were frequently confused with writing styles from other countries (3). Furthermore, forensic document examiners (FDEs) undertake expertise to offer opinions regarding the authorship of questioned handwriting, although their conclusions may vary based on their training and experience (11). Although the identification and discrimination of the author's ethnicity is beyond the scope of scientific context, the possibility that authors of different ethnicities can be discriminated against each other based on the uniqueness of class characteristics that may exist among them cannot be ruled out.

In this context, conducting identification or discrimination tasks by quantifying certain aspects of handwriting (11) may prove imminent. Geometric morphometrics (GMM) is a mathematical study of shapes which originally focused on Cartesian landmark coordinates for specific areas of structural, functional, and developmental importance (12). Recently, there has been a growing interest in using GMM to forensic anthropology. The popularity of this methodology might be attributed to its complexity as an acquisition method, which allows for a comprehensive characterization of shape and provides a rigorous toolkit for allometry (13). Despite its possible applications in forensic investigations, there have been no reports on its use for other study domains, particularly in question document analysis.

Hence, this present study was carried out to compare the class characteristics of the handwritten numeral characters among participants from different ethnic groups, residing in the same geographical region while attending different primary school systems in Malaysia. The main goal of this study was to demonstrate the application of the novel GMM approach, for the first time, in analyzing the handwritten numeral characters between Malay, Chinese and Indian participants based on primary school systems that they experienced in Malaysia. Throughout this study, the researchers effectively identified and discriminated the participants according to their ethnicities by evaluating their handwritten numeral characters. This was achieved by highlighting the different and complementary possibilities that the GMM approach could offer for accurately describing the shape variations of numeral characters among Malay, Chinese and Indian participants. Furthermore, this study also facilitated the quantitative assessment of morphological variations using principal component analysis (PCA), Procrustes analysis of variance (ANOVA) and wireframe graph. This type of analysis can potentially offer better insights into how the handwritten numeral characters can be used to identify and discriminate an individual's origin or ethnic group.

Materials and Methods

Participants Recruitment

Upon obtaining the ethical approval from the Human Research Ethics Committee of Universiti Sains Malaysia (USM) (JEPeM USM Code: USM/JEPeM/19110832), numeral handwriting specimens from 390 consented Malaysian participants aged 18 to 60 years old, were obtained, representing equal number of participants for each gender as well as ethnicities (Malay, Chinese and Indian). The sampling was done at the three USM campuses. The choice of these campuses was strategic, allowing the researcher to efficiently manage the logistics of recruiting participants prior to participate in this study from the diverse student populations available at these campuses.

In Malaysia, primary education is compulsory for all Malaysian children, and in this study, the participants must have attended one of the public national primary schools. The schools considered included Sekolah Kebangsaan (SK), Sekolah Jenis Kebangsaan Cina (SJKC) and Sekolah Jenis Kebangsaan Tamil (SJKT) with Bahasa Malaysia, Chinese and Tamil as their core subjects, respectively. Specifically, for Malay participants, they must have attended a government-funded Islamic school, known as Sekolah Rendah Agama (SRA) where Arabic language, Jawi and Imlak were taught. The decision to include only Malay participants whom have attended both SK and SRA was made by considering that learning Bahasa Malaysia alongside Jawi and Imlak may influence the construction of their handwriting, including both alphabetical and numerical characters. Moreover, there is currently no existing study focusing on such an aspect. Therefore, this present research was conducted to provide empirical evidence on how these ethnicities and native language can impact on the numeral handwriting characteristics by comparing the class characteristics of the handwritten numeral characters among participants from different ethnic groups. While participants must be fluent of their native languages to participate, those with learning or writing difficulties, as well as physical or mental disabilities, were excluded. The distribution of participants by ethnicity, gender and types of primary school attended is shown in Table 1.

Table 1. The distribution of participants by ethnicities, gender and types of primary school group

Ethnicities of participant	Gender		Types of primary school attended		
	Male	Female	SK & SRA	SJKC	SJKT
Malay	65	65	130	-	-
Chinese	65	65	-	130	-
Indian	65	65	-	-	130

Numeral Handwriting Specimen Sampling and Pre-Processing

Each participant was provided a consent and information form that included attributes such as name, age, types of primary school education, gender and ethnicity. Participants were also provided a blank sheet of A4 white paper and a blue ballpoint pen of the same brand and model (Faber-Castell, 0.7). Participants were required to write numeral characters from 0 until 9 in their most natural handwriting, without force, for seven times to ensure reliability and repeatability in handwritten numeral characters analysis (14,15). It allows for the identification and determination of consistent class characteristics, reduction errors, accounted for natural variations and provides a larger and more robust dataset for analysis (14,15).

Writing the same numeral characters for seven times would help in establishing consistency in handwriting patterns, enabling suitable evaluation to be made (14), especially on individual's consistency in reproducing their handwriting style by repeating the writing specimen. However, only the handwritten numeral characters from the fourth time line were accounted for and extracted since the fourth time of writing represented the most optimum handwriting style (14). This consistency is important for handwriting analysis reliability. Next, everyone's form was labelled with a unique code for identification and archiving purposes.

Pre-processing refers to image processes at the most basic level of abstraction (2), as a preparatory work for specimen sampling where some particular parts of a larger datasets are chosen prior to analysis (16). The goal of pre-processing is to improve the image data by suppressing undesired distortions or enhancing certain image features that are useful for further processing (2). Scanning, smoothing and noise removal were all part of this stage and following sampling, each of the handwritten numeral character specimen obtained was processed. Prior to the processing stage, all the handwritten numeral character specimens were scanned at 300 dots per inch (dpi), and converted into scanned digital images using an Epson TX120 scanner. After scanning all of the handwritten numeral characters, smoothing and noise removal were performed with Adobe Photoshop filters in order to remove artefacts or extra pixels created during the scanning process.

Geometric Morphometric (GMM) Analysis

MorphoJ software version 1.06d was utilized to assess length, depth and width ratio of the handwritten numeral characters, prior to the GMM analysis. Geometry, multivariate morphometric, computer science and imaging techniques, were used in this GMM approach of assessing the handwritten numeral characters (16). Formally, the GMM is defined as a set of approaches for multivariate statistical analysis of Cartesian coordinate data, which traditionally limited to landmark point locations (16). Following pre-processing of the handwritten numeral character specimens, all scanned images were digitized using the tpsUtil software. The landmark configurations on the handwritten numeral characters were then completed by utilizing the tpsDig2 software for further processing.

Landmark configuration is the heart of the GMM approach, in which the landmark precisely positioned over the handwritten numeral character specimens and correlated in a one-to-one manner amongst Malay, Chinese and Indian participants in this study. However, because there were no established landmark configurations on the handwritten numeral characters, the configuration's landmark points were designed in this study. The landmarks configurations for this analysis were defined by selecting specific points on the objects of interest (in this case, two-dimensional (2D) handwritten numeral character images) that were anatomically meaningful and capable of capturing the shape variation. In order to comprehend and choose the landmark that was most likely to capture meaningful shape variations by considering the research goals, the process of determining landmarks included a combination of understanding the handwritten numeral characters morphology, by reviewing relevant studies and consultation with experts. Besides, the number of landmarks used and their placement on the handwritten numeral characters were also determined by considering the relevant variation in morphology that could exist within numeral characters among ethnic groups.

The designated landmarks must be defined consistently across the handwritten numeral character specimens, and must be clearly identifiable for the GMM analysis. Furthermore, the designated landmarks were probably less prone to manual digitization error as the same landmarks can be easily identified in the same structure covering the specimen's shape comprehensively. In this regard, the definition of landmark configurations of each the handwritten numeral characters suggested by the previous researchers (17) was utilized for performing and discussing GMM analyses.

The generalized Procrustes analysis (GPA) with full-tangent space projection was performed on the resulting 2D coordinate configurations (n=390). This was accomplished using the MorphoJ software. This Procrustes procedure was performed in order to eliminate the irrelevant information from the handwritten numeral character specimens, prior to morphological shape comparison. Each designated landmark was translated at the origin superimposed by subtracting the coordinates of its centroid from the corresponding X or Y coordinates (16). GPA then generated a new matrix of Procrustes coordinates before rescaling and rotating to the mean centroid size, which was calculated as the square root of the summed squared distances of each landmark from the centroid (16).

This Procrustes superimposition removed sized-based effects, leaving only shaped-based differences. In this study, the principal component (PC) scores were generated since Procrustes coordinates can be equally applied on PC scores. Prior to PCA and Procrustes ANOVA, the MorphoJ software performed a complete Procrustes fit, which was more conservative and resistant to shape outlier's variation of the analysis (16). PCA and Procrustes ANOVA were performed in order to test the differences, examining the existing variations of numeral characters written by Malay, Chinese and Indian participants.

Principal Component Analysis (PCA)

Traditional morphometric analysis, also known as multivariate morphometric analysis, involves applying a number of morphological variables in order to describe the morphological shape variation within and among groups of objects (18,19). Typically, linear distance measures were utilized, but other methods such as counts, ratios, and angles were also commonly used. Moreover, co-variation in the morphological measurements, as well as patterns of variation within and among groups, can be quantified using these methods (18,19). MorphoJ software generated PCA using the covariance matrix, which derived from the group means distributions and simplified those variations to make them easier to visualize and comprehend. PCA helps in reducing the overwhelming number of dimensions images while still capturing the core of the dataset. PCA was utilized on morphological shape patterns in order to determine which handwritten numeral characters can be used to discriminate Malay, Chinese and Indian participants. Furthermore, in this present research, PCA was also used as a graphical method in order to visualize patterns of shape variation of the handwritten numeral characters among Malay, Chinese and Indian participants, and aided in the comprehension of shape variations for the handwritten numeral character specimens. It also demonstrated the data clustering with equivalent characteristics between specimens.

In this present research, PCA generated scatterplots that reflected the dispersion of shapes in tangent space and gave a visual representation of the relative shape variation of the handwritten numeral characters between participants from different ethnic groups. The PCA scatterplot was generated by applying the principal components (PCs) that possess a high variance value. On the other hand, the relative separation of groups in a PCA scatterplot does not allow one to conclude significant differences (or its absence). Thus, to effectively interpret the results and support the PCA's findings, additional descriptive tools such as Procrustes ANOVA and wireframe graphs were evidently necessary.

Procrustes Analysis of Variance (ANOVA) and Wireframe Graph

Procrustes analysis of variance (ANOVA) is a statistical shape analysis that used to evaluate the distribution of a set of shapes and determine whether or not there is a statistically significant difference between two or among more sets of specimens based on the Procrustes coordinates or distances (20, 21). In this present study, Procrustes ANOVA was used to compare the variations in the handwritten numeral character specimens among Malay, Chinese and Indian participants, as well as to assess the correlation of similarity and dissimilarity among individuals in the given specimens. Furthermore, 10,000 randomized permutations of landmark reshuffling were used in conjunction with Procrustes ANOVA in order to test the null hypothesis (**H_0**)**,** implying no significant difference between specimens obtained, concerning morphological shape differences and variations amongst Malay, Chinese and Indian participants (n = 390).

The p-value represents the likelihood of discovering that the observed, or the **H_0** of the study question, was true or vice versa. Hence, the p-value represented any statistically significant differences (22), in the handwritten numeral characters written by Malay, Chinese and Indian participants. After running the Procrustes ANOVA, the p-value that was considered following the analysis was $p<0.0001$, as suggested by previous researchers (23, 24). When the p-value is lower than 0.0001, the alternative hypothesis (**H_A**) shall not be rejected, whereby highly significant differences in the handwritten numeral characters between Malay, Chinese and Indian specimens can be construed.

The derived Procrustes coordinates were visualized using the wireframe graphs feature, which allowed all the landmarks connected for each geometric shape of specimens (25), and the significant shape differences between each handwritten numeral character to be described. Wireframe graphs used previously constructed wireframes to connect all of the landmarks with straight lines for the starting and target shapes, displaying the relative displacements of landmarks from a mean shape (26).

In MorphoJ software, wireframe graphs were easily generated for 2D data. The wireframe visualization of the first three PCs in this study (accountable for > 60% of the total variance) revealed that the handwritten numeral characters from 0 to 9 differed in the morphological shape among Malay, Chinese and Indian participants. While the first three PCs accounted for over 60% of the total variance, it is important to note that this was just one aspect of this present study. The PCs were chosen based on their initial dominance in explaining the variance; however, further exploration/refinement of this aspect proves necessary.

Wireframe graphs illustrated the shape changes and variations along the PC1, PC2 and PC3 axes by comparing the shape changes and variations of each numeral character written by Malay, Chinese and Indian participants. The light-blue lines represented the mean shape while the dark-blue lines represented the shape change in positive and negative directions (27,28) for each handwritten numeral character. The wireframe graphs supplemented the findings of the PCA and Procrustes ANOVA, revealing differences in the morphological shape of the numeral characters written by Malay, Chinese and Indian participants.

Results and Discussion

Several statistical and artificial intelligence (AI) techniques have been employed to study and compare the handwritten numeral character specimens among individuals from various geographic locations (2,29,30). When the GMM approach (based on Procrustes methods, including the PCA, Procrustes ANOVA and wireframe graphs) was used, this present study yielded promising results in identifying and discriminating handwritten numeral characters among Malaysians. It was discovered that certain handwritten numeral characters can be utilized to classify and discriminate, as well as to correlate the participants to particular ethnicities. The findings of the PCA, Procrustes ANOVA and wireframe graphs suggested that there were differences in the handwritten numeral characters of 2, 3, 4, 5, 6, 7, 8 and 9 among Malay, Chinese and Indian participants. In addition, using the GMM approach, significant variations for the most of the analyses among the three ethnicities prevailed. However, when these three ethnic groups were compared using this approach, the handwritten numeral characters of 0 and 1 did not yield statistically significant differences (p>0.0001), indicating that the observed shape variations were not significant enough to be considered as statistically meaningful among Malay, Chinese and Indian participants. Besides, when visualizing the data on the PCA scatterplot, no distinct clustering can be observed for the handwritten numeral characters of 0 and 1, which suggested that the data points associated with these values shared similar characteristics and exhibited minimal dispersion. Given these insights, we chose to only focus on the findings of the handwritten numeral characters of 2, 3, 4, 5, 6, 7, 8 and 9, which demonstrated promising results that were more likely can be used to identify and discriminate the different ethnicities of the participants in Malaysia.

The first three PCs were extracted for each handwritten numeral character (viz. 2, 3, 4, 5, 6, 7, 8 and 9) accountable for more than 60% of the total variance among Malay, Chinese and Indian participants. Considering the shape of handwritten numeral characters, the morphology of the numeral characters written by Malay, Chinese and Indian participants was compared using a relative percentage of total variance. Typically, the relative percentage of PCs, comprised of PC1 and PC2, should supposedly be equal to or greater than 60% (31). However, in this research, we considered PC3 that exhibited the next largest variation, when looking at the trend from the Eigenvalue graph (Figure 1a to Figure 1h).

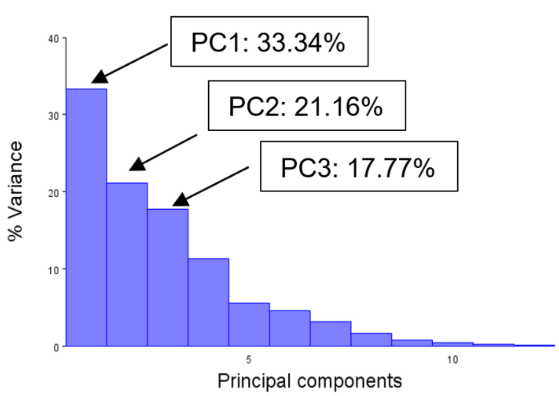


Figure 1a. Eigenvalue graph of Malay, Chinese, and Indian participants ' handwritten numeral character of 2

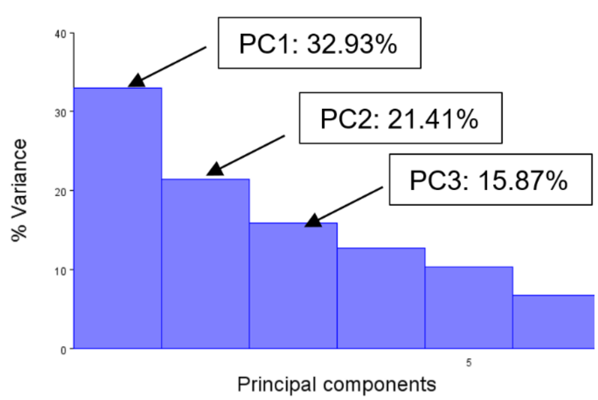


Figure 1b. Eigenvalue graph of Malay, Chinese, and Indian participants ' handwritten numeral character of 3

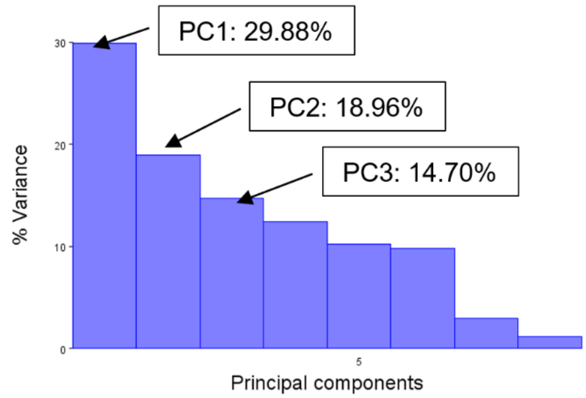


Figure 1c. Eigenvalue graph of Malay, Chinese, and Indian participants ' handwritten numeral character of 4

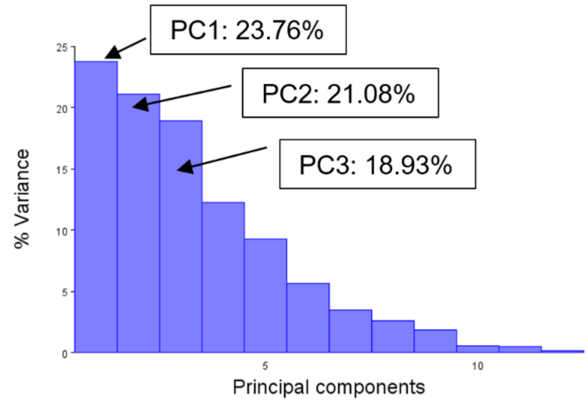


Figure 1d. Eigenvalue graph of Malay, Chinese, and Indian participants ' handwritten numeral character of 5

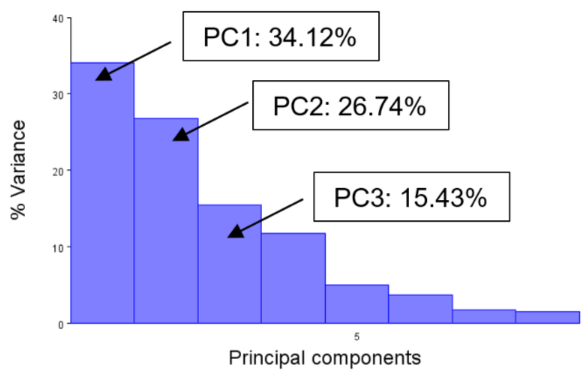


Figure 1e. Eigenvalue graph of Malay, Chinese, and Indian participants ' handwritten numeral character of 6

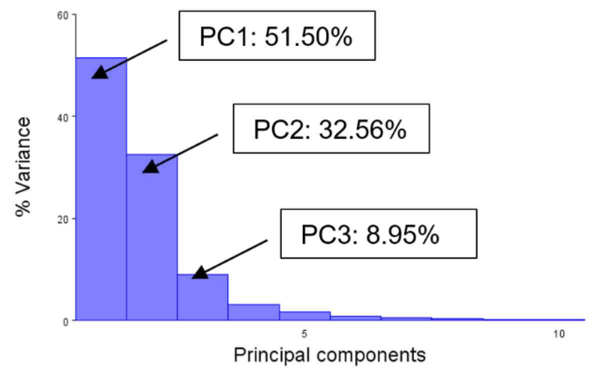


Figure 1f. Eigenvalue graph of Malay, Chinese, and Indian participants ' handwritten numeral character of 7

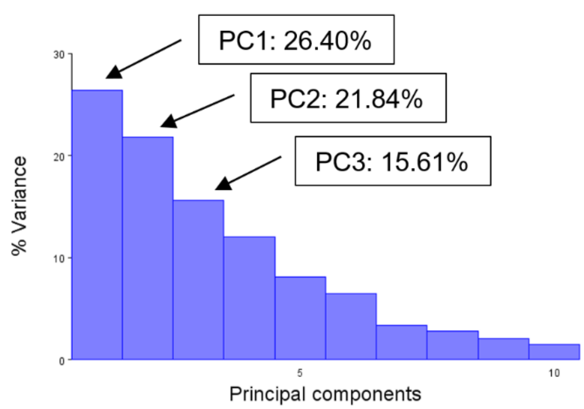


Figure 1g. Eigenvalue graph of Malay, Chinese, and Indian participants ' handwritten numeral character of 8

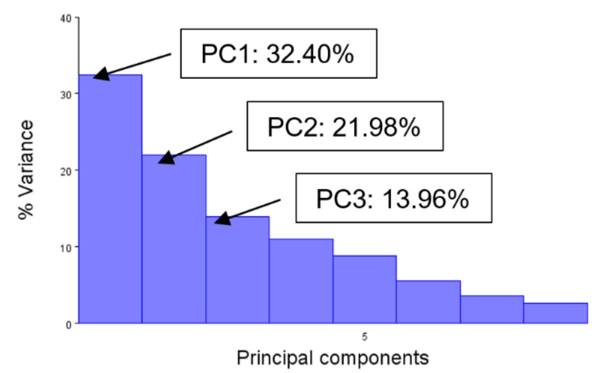


Figure 1h. Eigenvalue graph of Malay, Chinese, and Indian participants ' handwritten numeral character of 9

According to the findings, the relative percentage of PC for the handwritten numeral characters of 2 to 9 demonstrated a satisfactory value to represent the shape variation in the handwritten numeral character. This demonstrated that the chosen defined landmarks were capable to capture the shape variation of each handwritten numeral character among these three ethnicities, showing a valuable shape variation of each handwritten numeral character. Furthermore, PCA scatterplots generated by the MorphoJ software (Figure 2a until Figure 2h) revealed ethnic clusters based on the morphological similarity and dissimilarity of the handwritten numeral character specimens. PCA scatterplots identified three clusters, represented by blue, red, and green, which corresponded to the Malay, Chinese, and Indian ethnic groups, respectively.

These clustering results revealed considerable significant differences in the handwritten numeral characters of 2 to 9 between numeral among Malay, Chinese and Indian participants. However, the PCA scatterplots revealed an overlap among the three participant ethnic groups, suggesting that the discrimination task was indeed complex due to the shared similar characteristics of the handwritten numeral characters among ethnicities. This observation highlighted that the PCA-based analyses provided a preliminary glimpse into the possibility for the participant identification and discrimination according to their ethnicities. To improve the accuracy of the attribution model, more sophisticated analytical approaches that account for such overlap and intricate relationships by exploring advanced statistical techniques and strategies tailored to forensic cases were required. This included examining the features beyond PCA that may better capture participant -specific attributes while considering the nuances of cross-group similarities in the handwritten numeral characters.

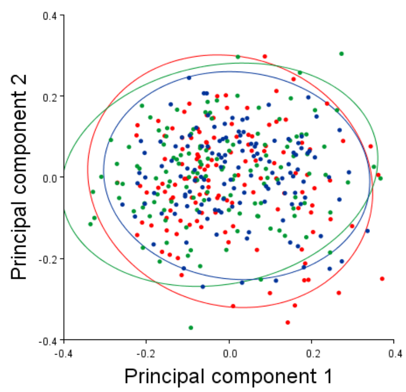


Figure 2a. PCA scatterplot of the handwritten numeral character of 2 among Malay vs Chinese vs Indian participants

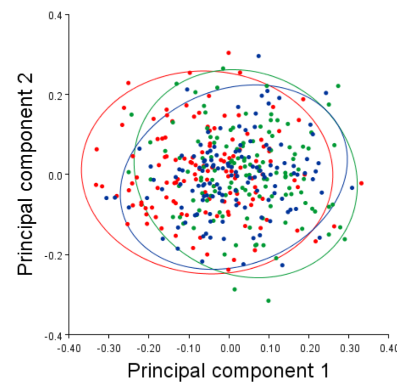


Figure 2b. PCA scatterplot of the handwritten numeral character of 3 among Malay vs Chinese vs Indian participants

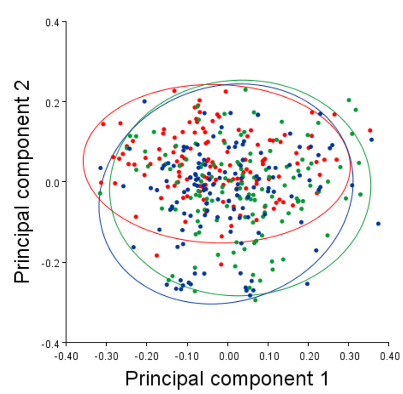


Figure 2c. PCA scatterplot of the handwritten numeral character of 4 among Malay vs Chinese vs Indian participants

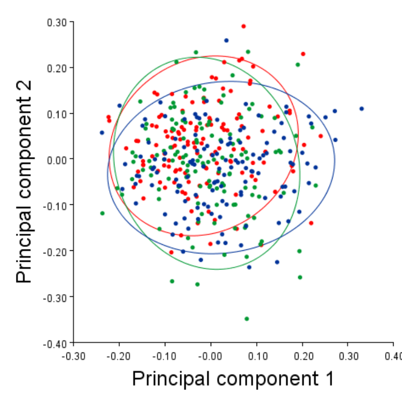


Figure 2d. PCA scatterplot of the handwritten numeral character of 5 among Malay vs Chinese vs Indian participants

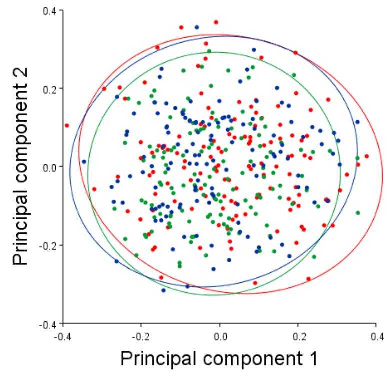


Figure 2e. PCA scatterplot of the handwritten numeral character of 6 among Malay vs Chinese vs Indian participants

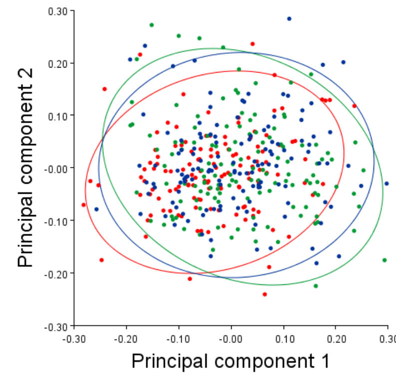


Figure 2f. PCA scatterplot of the handwritten numeral character of 7 among Malay vs Chinese vs Indian participants

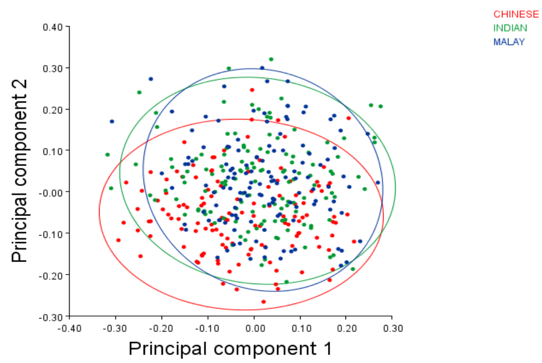


Figure 2g. PCA scatterplot of the handwritten numeral character of 8 among Malay vs Chinese vs Indian participants

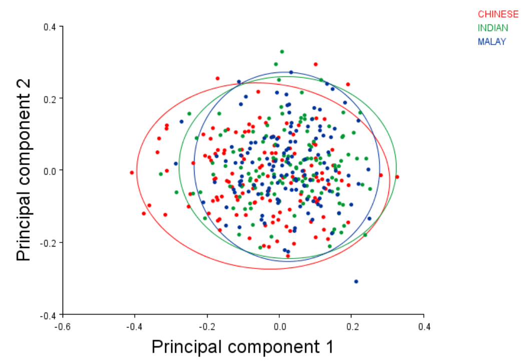


Figure 2h. PCA scatterplot of the handwritten numeral character of 9 among Malay vs Chinese vs Indian participants

The findings of the Procrustes ANOVA are presented in Table 2, revealing a significant difference (p<0.0001) in the morphological shape for each handwritten numeral character among Malay, Chinese and Indian participants with high Goodall's F statistic (F) values. Besides, when performing the Procrustes ANOVA, the degrees of freedom (df) ranged between 12-24, representing the number of shape variations and reflects the nature of the datasets when the comparing Malay, Chinese and Indian participants' handwritten numeral characters. The df value was closely tied to the underlying assumptions of the statistical test used in this study for determining the p-value.

Table 2. Procrustes ANOVA of handwritten numeral characters among Malay, Chinese and Indian participants

Handwritten Numeral Characters	SS	MS	Df	F	*p*-values
2	0.8304	0.0346	24	6.86	<0.0001
3	0.8621	0.0718	12	9.36	<0.0001
4	0.8797	0.0550	16	7.58	<0.0001
5	0.6608	0.0275	24	7.52	<0.0001
6	0.7182	0.0449	16	5.72	<0.0001
7	0.2505	0.0125	20	5.63	<0.0001
8	0.9195	0.0460	20	8.77	<0.0001
9	0.7358	0.0460	16	7.49	<0.0001

Note: SS: sums of squares; MS: mean squares; df: degrees of freedom; F: statistics and parametric p-values are provided for each effect

The *p*-value was used to describe the test statistic's probability distribution under the assumption of $\mathbf{H_0}$. The smaller the *p*-value, the higher the significant differences. Because the *p*-value was only used as an aid after the specimen collection, to assess whether the observed statistic is a simply random event or indeed belongs to a unique phenomenon fitting the researchers' scientific hypothesis, the *p*-value considered was p <0.0001 due to its highly significant inference (32). As a result, for the shape variations on each numeral character, the differences were small yet significant. According to the Procrustes ANOVA results, the handwritten numeral characters of 2, 3, 4, 5, 6, 7, 8 and 9 were highly recommended to be used, prior to discriminate the numeral handwriting among Malay, Chinese and Indian participants i.e. when FDEs need to examine a pool of numeral handwriting evidence involving people from these ethnicities. There were highly significant differences in the handwritten numeral characters among these three ethnicities, as evidenced by the fact that their *p*-values for each handwritten numeral character being less than 0.0001. One possible explanation for this finding can be attributed to the differences in the native language writing styles, possibly shaped by the exposure they had during primary school. There were numerable differences in native language writing systems across different regions, including differences in the make-up of characters, the native language structure, how handwriting is taught in primary school and the importance placed on it (3,10 ,33)

It was hypothesized that the handwritten numeral characters among Malay, Chinese and Indian exhibit the potential to help in the participants identification and discrimination according to their respective ethnicities. As such this research was the first that supported the potential to break new ground in the realm of forensic questioned documents analysis, while a detailed quantitative assessment of accuracy rates is another part of our ongoing analysis. All these findings indicated that ethnicities can be a predictive factor of numeral handwriting, which differed across ethnicity groups. Though the differences among Malaysia's three main ethnicities were minor, they corresponded to our expectations when comparing Malay, Chinese and Indian participants who live together in a diverse society in urban or rural areas. A plural society is one in which several ethnic groups coexist in discrete geographical and sociocultural enclaves, only coming together in the marketplace or under complex government (34,35). When Malaysia gained independence in 1957, it became a largely harmonious nation of diverse ethnic groups. Since then, Malaysia has been well known as a diverse country, consisting many different ethnicities, languages and religions, whereby each ethnicity retained their ancestors' cultural customs and norms for several generations (34,35).

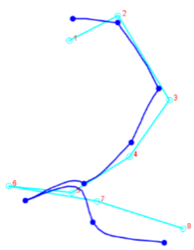
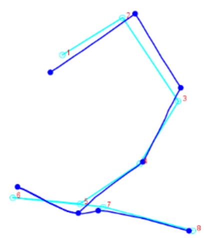
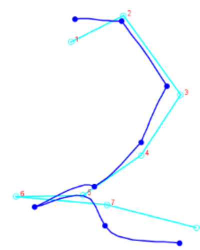
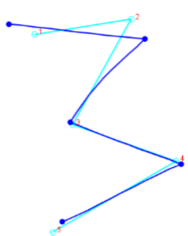
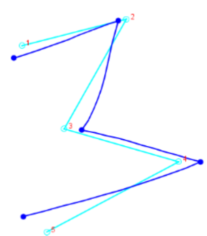
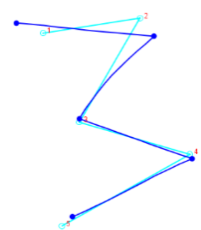
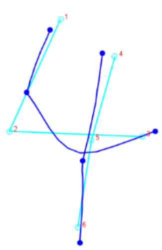
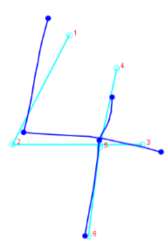
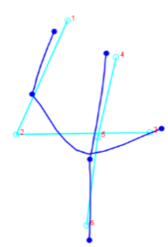
Malaysia's multiculturalism does not change the fact that their national language, Bahasa Malaysia is not the mother tongue language of the Chinese and Indian communities. In order to communicate with one another, they must also learn their mother tongue languages; Mandarin and Tamil as well as English, because of its importance in modern society. With such an early awareness of cultural and native linguistic differences, it is possible that this would influence their numeral handwriting style when they learn it during their primary education. This may reason out the concern for ethnic identity and group formation among Malaysians since they favored their children to attend vernacular and religious schools, as well as able to speak and read in their own mother tongue language (3,35).

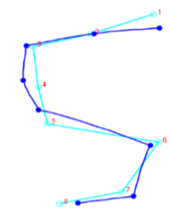
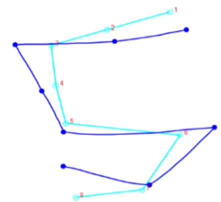
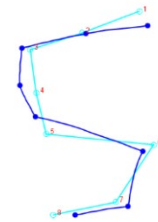
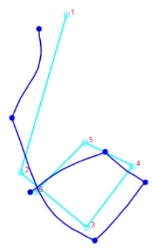
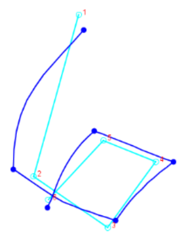
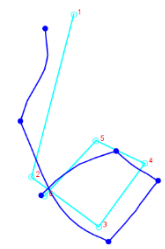
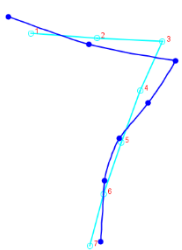
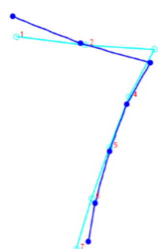
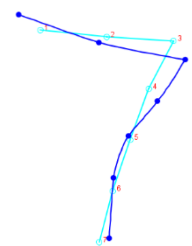
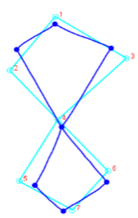
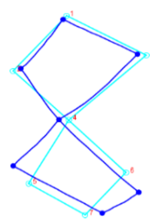
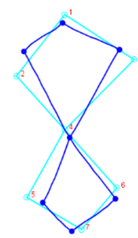
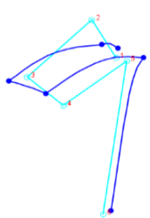
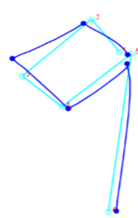
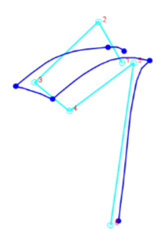
Furthermore, Malaysia also does not have a pattern of geographical population distribution by ethnic groups akin to the Canada and Switzerland (34), based on their handwriting especially on numeral handwriting. In this case, the Procrustes ANOVA suggested a minor but discernible variation in morphological shape of numeral characters written by Malay, Chinese and Indian participants in order to identify and determine the significant differences in each handwritten numeral character across these three ethnicities. To the best of our knowledge, this was the first study in Malaysia that used this approach to identify and discriminate handwritten numeral character between Malay, Chinese and Indian participants based on the primary school systems prior learning. This study can be acknowledged and applied to analyze and determine the ethnic background of participants if FDEs come across a questioned document at a crime scene involving persons of different ethnicities. Exploring and varying level of depths about the handwritten numeral characters among these three main ethnic groups; Malay, Chinese and Indian, may prove be beneficial for the future of forensic questioned document examination.

Visualization of morphological shape variation is a key component of GMM approach for data exploration, hypothesis formulation and testing, as well as reporting. As a result, the visualization of shape variation for the handwritten numeral characters must appear as more realistic representations, in order to provide additional information prior to aid the researcher in correlating the PCA scatterplot and Procrustes ANOVA's findings to the structures under this study. Since there are no "right" or "wrong" visualizations approaches, there are several different available ways to indicate the morphological context of the landmarks such as utilizing wireframe graphs to depict the target shapes side by side (36). In this study, a pair of superimposed wireframe graphs that connected the landmarks with straight lines for the from the beginning point of the target shapes was performed prior to visualizing how the landmark points shift from the beginning to the terminal point of the target shape. It is crucial for viewers to understand that each shift at any landmark can be related to all previous shifts (36).

Wireframe graphs of mean morphological shape variations from Table 3 demonstrated a visual assessment of these numeral characters written by Malay, Chinese and Indian participants. According to these wireframe graphs, the average morphological shape for numeral characters of 2 until 9 written by Malay, Chinese and Indian participants, corresponded to a significant difference in their immediate morphological context. The visualization of the wireframe graphs associated with different ethnicities, correlated to the morphological features, allowing for a clear diagnosis of the morphology for each handwritten numeral character. The comparison of the mean shape in positive and negative directions can be observed in the wireframe graphs in Table 3, where the positive and negative directions (purple lines) varied and were wider than the mean shape (light-blue lines). Wireframe graphs of the mean for the morphological handwritten numeral character of 2 until 9 and their displacement landmarks described the significant shape differences, which corresponding to the PCA scatterplot and Procrustes ANOVA's findings.

Table 3. Wireframe graphs of the handwritten numeral characters between Malay, Chinese and Indian authors

Wireframe graph of handwritten by Malay vs Chinese participants	Wireframe graph of handwritten by Malay vs Indian participants	Wireframe graph of handwritten by Chinese vs Indian participants
		
		
		

Wireframe graph of handwritten by Malay vs Chinese participants	Wireframe graph of handwritten by Malay vs Indian participants	Wireframe graph of handwritten by Chinese vs Indian participants
		
		
		
		
		

The relative aspect of landmark points movements is shared by all visualizations based on the superimpositions, regardless of the graphical techniques used to depict the morphological context of handwritten numeral characters (36). Although the wireframe graphs do not depict a physical or biological reality, only an imagined aid for visualization, the algorithms used to compute wireframe graphs were based exclusively on the geometric criteria of numeral characters written by Malay, Chinese and Indian participants. Besides, researchers have to note that alternative registrations of the identical beginning and terminal points of the target shape might result in displays with fundamentally different appearances, depending on the application that has been used (36). Therefore, the interpretation of variations could be simplified if the researchers strive to create clear visualizations and communicate the complex shape variations as a means of discovering and disseminating patterns of the morphological variation for the handwritten numeral characters in full morphometric context, while minimizing the risk of misunderstanding in sharing insight from their morphometric analyses (36).

The substantial divergence in morphological variance for each handwritten numeral character is a result of the participants' cultural and native linguistic backgrounds. The native language and cultural influences on the handwritten numeral characters are related, resulting in morphological variation between in Malay, Chinese, and Indian participants. Except for the handwritten numeral characters of 0 and 1, significant variances were discovered in the remainder of the handwritten number characters. Overall, the differences in morphometric contexts that broadly associated with written numeral features, were the primary source of the handwritten numeral character variations observed among ethnicities studied in this present research. Since the shape variations observed in this study have limited practical use in terms of discriminating participants' ethnicity, further studies involving larger specimen sizes as well as more sophisticated statistical approaches prove necessary.

In order to indicate the difference in Malay, Chinese, and Indian handwritten numeral characters, the GMM analysis was performed. However, if there is only a minor variation in how these ethnicities write the numeral characters, the GMM approach may not be able to identify, resulting in most of the specimens being positioned close to each other. In addition, this approach was also incapable of differentiating and distinguishing between the handwritten numeral characters based on subtle variations or features such as the slant, stroke, angle and direction of the numeral characters themselves. Yet, there were certain clusters of the handwritten numeral characters that overlapped slightly among the three ethnic groups that were studied.

The overlapping clusters can be due to the participants' posture or the type of pen used to write the numeral characters during the writing process. Aside from that, when it comes to writing, every individual may experience and generate natural variation depending on the weather, surroundings, type of surface and the participants' emotions when they are about to write (37,38). Handwriting styles can naturally differ from one another due to the factors such as numeral characters composition, pen-lifts, spacing, and embellishments (38). Nonetheless, the findings of this research supported the idea that cultural and native linguistic differences (among Malay, Chinese, and Tamil speakers) might explain the morphological shape variations in the handwritten numeral characters between Malay, Chinese and Indian participants.

Successful study conducted by the previous researchers (8) on Malay, Chinese and Indian participants in Singapore, confirmed that their native language writing systems did have an influence on their daily handwriting. They further revealed that because of the peculiar and uniqueness of their native linguistics, Malay, Chinese and Indian participants tend to write in a variety of styles. A Malay's handwriting is always smooth and curved, in contrast to a Chinese's handwriting, which is normally angular and sharp stroke in appearance, differently when compare to an Indian's handwriting which is broader and curved than the norm. Because of the influence of learning their mother tongue language, the participants of these three ethnic groups might employ different natural forms for their handwritten numeral characters (8).

Interestingly, the findings presented in this present research also revealed that the participants' primary educational and cultural background could play a greater role in influencing their handwritten numeral characters. Consequently, as suggested by the previous researchers (2), a participant's primary school may influence his or her handwritten numeral characters. It is reasonable to claim that the development of basic skill of handwriting has an influence on the nature appearance of the handwritten numeral characters. From the early phases of handwriting skills until mastering the basic handwriting skills, an author is influenced by his or her cultural peers, relatives, family and teachers, which leads in the transmission of graphical skill vertically, horizontally and obliquely (2). The authors from various nationalities or ethnicities, as well as different native language and writing systems that they were exposed could exhibit discriminative of class and individual handwriting characteristics (3), and these differences arise as a result due to the ethnic differences in the intonation, rhythm, and pronunciation styles of the taught native linguistics (4).

Conclusions

The use of landmark configurations in GMM analysis helped to investigate how ethnicity and native linguistic disparities among Malay, Chinese, and Indian participants could influence their handwritten numeral characters, particularly for forensic document examination. This study found that the GMM approach can effectively identify and discriminate the participants based on shape variations into different ethnicities, especially for numeral character of 2 until 9. The findings also highlight the impact of culture and native linguistic disparities on numeral handwriting, suggesting potential applications in forensic investigations and supporting the use of illegible handwriting as evidence. Individuals handwriting styles vary greatly, making it impossible to utilize the handwritten numeral characters prior to determine someone's ethnicity, origin or nationality. Nonetheless, this study with the GMM approach can pave the new way for future studies in which integration of the findings with those of other unsupervised and supervised machine learning methods in order to improve the precision of ethnicity identification and discrimination procedure, can be promoted. It is expected that the results presented here can serve as a call to action for other researchers to explore the feasibility of the GMM approach prior to visually convey information about the morphological shape diversity in the handwritten numeral character among participants of different ethnicities. It would be intriguing to examine how the handwritten numeral characters can contribute to the fundamental laws of handwriting analysis.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgement

The authors acknowledged the financial assistances given by PHA Handwriting Analysis Sdn. Bhd. and HFDE Services Pte Ltd (Singapore) through the industry grant initiative. The authors would also like to thank all the participants and to those who have directly and indirectly involved in this study.

References

[1] Sahani, P. (2021). Distribution of various factors affecting handwriting among North and North-Eastern population groups of India. *International Journal of Scientific Research, 10*(1), 1–3.

[2] Saini, M., & Kapoor, A. K. (2018). The comparative influence of culture and schooling environment on handwriting features. *Arab Journal of Forensic Sciences & Forensic Medicine, 1*(7), 842–58.

[3] Gannetion, L., Wong, K. Y., Lim, P. Y., Chang, K. H., & Abdullah, A. F. L. (2022). An exploratory study on the handwritten allographic features of multi-ethnic population with different educational backgrounds. *PLOS ONE, 17*(10).

[4] Ramaiah, C., Porwal, U., & Govindaraju, V. (2012). Accent detection in handwriting based on writing styles. In *Proceedings – 10th IAPR International Workshop on Document Analysis Systems, DAS 2012* (pp. 312–6).

[5] Tyagi, N. (2018). Qualitative analysis of class characteristics of individuals belonging to Northern India of handwriting samples written in Devanagari script. *Journal of Forensic Sciences & Criminal Investigation, 9*(3), 1–5.

[6] Turner, I. J., Sidhu, R. K., & Love, J. M. (2008). A preliminary study investigating class characteristics in the Gurmukhi handwriting of 1st and 2nd generation Punjabis. *Science and Justice, 48*(3), 126–32.

[7] Cheng, N., Lee, G. K., Yap, B. S., Lee, L. T., Tan, S. K., & Tan, K. P. (2005). Investigation of class characteristics in English handwriting of the three main racial groups: Chinese, Malay, and Indian in Singapore. *Journal of Forensic Sciences, 50*(1), 1–8.

[8] Harne, P., Mishra, M. K., & Sodhi, G. S. (2018). Analysis of handwriting characteristics based on diverse ethnic distribution. *International Journal of Advanced Trends in Computer Application, 5*(2), 1–6.

[9] Zhang, B., Srihari, S. N., & Lee, S. (2003). Individuality of handwritten characters. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)* (pp. 1086–90).

[10] Marquis, R., Bozza, S., Schmittbuhl, M., & Taroni, F. (2011). Handwriting evidence evaluation based on the shape of characters: Application of multivariate likelihood ratios. *Journal of Forensic Sciences, 256*(SUPPL. 1), 238–42.

[11] Johnson, A., Singh, S., Thomas, A., & Chauhan, N. (2021). Geometric morphometric analysis for sex determination using lateral cephalograms in Indian population: A preliminary study. *Journal of Oral and Maxillofacial Pathology, 25*(2), 364–7.

[12] Viacava, P., Blomberg, S. P., & Weisbecker, V. (2023). The relative performance of geometric morphometrics and linear-based methods in the taxonomic resolution of a mammalian species complex. *Ecology and Evolution, 13*(3).

[13] Hakimie, N. (2020). Classification of individuals using handwritten numeral with geometric morphometric techniques (MSc. thesis). Universiti Sains Malaysia.

[14] Taufek, W. N. S. W. M., Pritam, H. M. H., Desa, W. N. S. M., & Ismail, D. (2023). Identification of writers' ethnicity using handwritten numeral characters in combination with novel geometric morphometric (GMM) technique. *AIP Conference Proceedings, 2896*(1).

[15] Manacorda, C. A., & Asurmendi, S. (2018). Arabidopsis phenotyping through geometric morphometrics. *Gigascience, 7*(7), 1–20.

[16] Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the 'revolution'. *Italian Journal of Zoology, 71*(1), 5–16.

[17] Taufek, W. N. S. W. M., Pritam, H. M. H., Desa, W. N. S. M., & Ismail, D. (2024). Evaluation of geometric morphometric approach for ethnicities discrimination using handwritten numeral characters. *Jurnal Teknologi, 86*(4), 139–150.

[18] Viscosi, V., & Cardini, A. (2011). Leaf morphology, taxonomy, and geometric morphometrics: A simplified protocol for beginners. *PLOS ONE, 6*(10).

[19] Klingenberg, C. P., Barluenga, M., & Meyer, A. (2007). Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. *Evolution, 56*(10), 1909–1920.

[20] Sanfilippo, P. G., Hewitt, A. W., Mountain, J. A., & Mackey, D. A. (2013). A geometric morphometric assessment of hand shape and comparison to the 2D:4D digit ratio as a marker of sexual dimorphism. *Twin Research and Human Genetics, 16*(2), 590–600.

[21] Medrano, M. G. T., & Jumawan, J. C. (2016). Landmark-based geometric morphometric analysis of body shape variation within population of Ibis fish, *Ambassis interrupta*, collected from Masao River, Butuan City, Philippines. *International Journal of Advances in Chemical Engineering and Biological Sciences, 3*(2).

[22] Dalton, H. A., Wood, B. J., Widowski, T. M., Guerin, M. T., & Torrey, S. (2017). An analysis of beak shape variation in two ages of domestic turkeys (*Meleagris gallopavo*) using landmark-based geometric morphometrics. *PLOS ONE, 12*(9), 1–25.

[23] Nelson, E., Hall, J., Randolph-Quinney, P., & Sinclair, A. (2017). Beyond size: The potential of a geometric morphometric analysis of shape and form for the assessment of sex in hand stencils in rock art. *Journal of Archaeological Science, 78*, 202–213.

[24] Von Cramon-Taubadel, N., Frazier, B. C., & Lahr, M. M. (2007). The problem of assessing landmark error in geometric morphometrics: Theory, methods, and modifications. *American Journal of Physical Anthropology, 134*(1), 24–35.

[25] Sauer, F. G., Jaworski, L., Erdbeer, L., Heitmann, A., Schmidt-Chanasit, J., & Kiel, E. (2020). Geometric morphometric wing analysis represents a robust tool to identify female mosquitoes (Diptera: Culicidae) in Germany. *Scientific Reports, 10*(1).

[26] Woon, C. K., Abu Jamal, N. A., Ilmi Mohd Nooi, M. N., Abdullah, S. M., Ibrahim, N. M., & Norman, N. H. (2019). Erratum: Geometric morphometric analysis of malocclusion on lateral cephalograms in Malaysian population. *Anatomy & Cell Biology, 52*, 397–405.

[27] Shin, J. Y., Alias, A., Chung, E., Ng, W. L., Wu, Y. S., & Gan, Q. F. (2021). Identification of race: A three-dimensional geometric morphometric and conventional analysis of human fourth cervical vertebrae in adult Malaysian population. *Journal of Clinical and Health Sciences, 6*(1), 17.

[28] Johnson, M. E., Vastrick, T. W., Boulanger, M., & Schuetzner, E. (2017). Measuring the frequency occurrence of handwriting and handprinting characteristics. *Journal of Forensic Science, 62*(1), 142–63.

[29] Nag, S., Shivakumara, P., Wu, Y., Pal, U., & Lu, T. (2018). New COLD feature based handwriting analysis for ethnicity/nationality identification. In *Proceedings of the 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (pp. 523–527).

[30] Foorman, B. R., & Petscher, Y. (2018). Decomposing the variance in reading comprehension to reveal the unique and common effects of language and decoding. *Journal of Visualized Experiments, 11*(140).

[31] Zhu, W. (2016). Opinion P < 0.05, < 0.01, < 0.001, < 0.0001, < 0.00001, < 0.000001, or ... *Journal of Sport and Health Science, 5*(1), 77–9.

[32] Bai, R., Hu, G., & Gu, P. Y. (2013). The relationship between use of writing strategies and English proficiency in Singapore primary schools. *The Asia Pacific Education Researcher, 23*, 355–365.

[33] Baharuddin, S. A., & Noor, M. M. (2006). Multiculturalism in Malaysia: The need for local knowledge to grapple with identity and ethnicity. *Demokrasi, 2*, 187–204.

[34] Winskel, H. (2020). Learning to read in multilingual Malaysia: A focus on Bahasa Melayu, Tamil, and Chinese. *GEMA Online Journal of Language Studies, 20*(1), 1–15.

[35] Klingenberg, C. P. (2013). Visualizations in geometric morphometrics: How to read and how to make graphs showing shape changes. *Hystrix, 24*(1), 15–24.

[36] Harralson, H. H., & Miller, L. S. (2018). *Huber and Headrick's handwriting identification: Facts and fundamentals* (2nd ed.). New York, NY: CRC Press.

[37] Morris, R. (2020). *Forensic handwriting identification* (2nd ed.). United Kingdom: Academic Press.

[38] Stewart, L. F. (2017). The process of forensic handwriting examinations. *Forensic Research & Criminology International Journal, 4*(5), 139–41.