**Analyzing Handwriting Characteristics of Children with Autism by Chinese Characters and Mandarin Phonetic Symbols**

# Abstract

Autism is a neurodevelopmental disorder that typically manifests symptoms during childhood, characterized by difficulties in social interaction and repetitive behaviors. In addition to the primary signs of autism, motor impairments are also commonly observed. Specifically, autistic children may have difficulties in converting sequential actions into integrated movements. These difficulties in fine motor skills can impact their daily lives, fostering frustration. Hand-eye coordination deficits may affect activities requiring precision and accuracy, including handwriting. This study delves into the handwriting characteristics of autistic children through a meticulous analysis of handwriting Chinese characters. Building upon existing research, our study has two main objectives. First, we incorporate phonetic notation data to investigate whether its inclusion improves the model's classification performance between autistic children and typically developing children. Second, we design a neatness label to distinguish whether each Chinese character is written neatly. The aim of the present study is to understand the neatness of handwriting for both groups of children. Additionally, we explore the more challenging task of training a classification model using only neatly-written Chinese characters.

The dataset used in this study is directly sourced from elementary school students' workbooks, providing a direct reflection of children's real-life situations. CAM (class activation map) technology is employed to analyze handwriting features. To overcome the limitations of manual observation, we also designed a method of encoding the CAM results. Through encoding the CAM results, we objectively and swiftly observe the handwriting differences between autistic children and typically developing children. Through expanding the handwriting dataset, introducing the neatness label, and utilizing the CAM results in a coded manner, this paper contributes to the field by suggesting handwriting analysis as a potentially valuable tool for detecting tendencies towards autism in children.

# Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that is typically identified in childhood. ASD is characterized by signs such as social challenges, language difficulties, repetitive behaviors, restricted interests, and oversensitivity or insensitivity to sensory stimuli [1-3]. In addition to the cardinal symptoms of autism, autistic children often experience challenges in motor coordination and planning, resulting in clumsiness [4, 5]. Autistic children may exhibit difficulties in various daily life movements, including maintaining posture [6-8], playing sports [9, 10], and grasping or manipulating materials with different weights and shapes [11-13].

Specifically, autistic children often experience challenges in foreseeing motor outcomes, conceiving goal-directed motor acts, integrating the sensory stimuli followed by movements, conducting sequential movements, and forming coordinated motor responses [14]. Moreover, deficiencies in hand-eye coordination can affect autistic children’s learning performance which requires motor precision and accuracy, such as reading and handwriting [15].

Numerous studies have explored the handwriting difficulties experienced by autistic children [16-18]. Previous research has indicated various deviated handwriting performances among autistic children. As illustrated in Figure 1, (A) shows a template, and (B) demonstrates more disconnected strokes, inconsistent letter sizes, and irregular shapes than (C) (both are from autistic children). It is therefore important to note that there is a large within-group difference among the autism population in their handwriting abilities and performance.

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

Figure 1: A is the template, and B and C are both written by individuals with ASD [16]

The differences between Chinese characters and English letters span multiple aspects, with a significant distinction in their glyph structures [19]. For instance, the structure of English letters is relatively simple, typically composed of a small number of curves and straight lines. In contrast, Chinese characters are characterized by rich strokes and intricate structures, incorporating various substructures and organizational patterns that deeply impact the meaning and context of the characters. On the other hand, the common curves and circles found in English letters are relatively less prevalent in Chinese characters, where the predominant elements are straight lines and horizontal strokes. Such distinctions also affect the motor planning and coordination in the process of handwriting, presenting a potential challenge for the learners, particularly for individuals with autism. Therefore, investigating the specific manifestations of Chinese handwriting is a challenging yet meaningful task for research in autistic children.

In the context of Chinese character writing, we collected a dataset from elementary school students' exercise books. These exercise books were provided by a total of 6 autistic children (average age: 10.5 years; all boys) and 17 typically developing (TD) children (average age: 8.67 years, eight boys, nine girls). Our first work (Yen et al. [20]) reveals that autistic children exhibit variations in arcs and spatial distribution for Chinese character writing. Although phonetic notations are also presented in the exercise books, they were not utilized. In Taiwan, elementary school students not only need to learn how to write Chinese characters, but also need to practice a phonetic system simultaneously in order to familiarize themselves with the pronunciation of the Chinese characters. The most frequently used phonetic system in Taiwan is the Mandarin Phonetic Symbols I (MPS I), also known as Zhuyin [21]. As shown in Figure 2, the system comprises 37 notations and 4 tonal marks. When practicing their handwriting, students need to write the Chinese characters and the corresponding phonetic symbols side by side. The writing style of phonetic notations falls between English letters and Chinese characters, embodying characteristics of both writing systems [22]. Similar to English letters, phonetic notations have fewer strokes, a relatively simple structure, and no distinct substructures. However, like Chinese characters, phonetic notations are primarily written vertically and consist of vertical, horizontal, and diagonal strokes. This combination of features from two writing systems makes phonetic notations unique in both shape and structure. Therefore, incorporating phonetic notations may contribute to providing distinctive features lacking in Chinese characters and enhancing the comparability of these features with those of English writing for analyzing handwriting characteristics of autistic children.

In this study, we first examine how neatly the Chinese characters were written. Here, the *neatness* refers to the cleanliness of handwriting, encompassing the consistency of character shapes and the organized arrangement of characters. In educational settings, teachers often place significant emphasis on the neatness of handwriting when assessing a child's writing performance. Neatness is not only associated with the aesthetic aspect of writing, it is also crucial for readability. If a child's handwriting is neat, with well-organized text and consistent character forms, the reading experience becomes smoother, and comprehension becomes more accessible. In this research, we have developed a classification model designed specifically to evaluate the neatness of Chinese character handwriting. By introducing the neatness label, the model learns to distinguish between neat and non-neat Chinese characters. If the model's predictions align with the assessment of the neatness from teachers, it can then be used to assist teachers in grading students' writing assignments. This makes the model a practical tool, supporting teachers in more effectively evaluating and guiding students' writing abilities, ultimately enhancing learning outcomes.



Figure 2: The Manual of the Phonetic Symbols of Mandarin Chinese [21]

Writing is a crucial aspect of language learning for children, and Chinese character writing places higher demands on hand-eye coordination and motor control. To delve further into this matter, we chose to analyze and compare the Chinese handwriting of autistic and TD children. We developed a classification model to decide whether a Chinese handwriting is from autistic or TD children with the following three objectives. Our first objective is to incorporate phonetic notation data into this study. We aim to investigate whether the inclusion of phonetic notation data improves the model's performance in classifying autistic and TD children, as Yen et al. [20] did not utilize phonetic notation data.

Our second objective is to design a neatness label to distinguish whether a Chinese character is written in a neat manner. It is a binary label (yes, no) annotated through manual labeling. The purpose of this is to track the neatness of handwriting for both autistic and TD children. By using only neatly written Chinese characters, we plan to train the classification model for the writing from autistic and TD children. This constitutes a more challenging task, as distinguishing the handwriting characteristics between autistic and TD children becomes even more intricate when all Chinese characters are written neatly. The purpose of this approach is to further assess the model's ability to differentiate the handwriting styles of autistic children from TD children when only neatly written Chinese characters are considered.

In Yen et al. [20], the classification activation map (CAM) technology [23] was employed to analyze the handwriting characteristics of autistic children. CAM is a visualization technique that assists in understanding the focus of the model. This method allows us to gain insights into which parts of the input contribute significantly to the model's decision, offering a more interpretable perspective on how the model processes and distinguishes writing characteristics from the autistic children. However, direct observation of the results of applying CAM for thousands of handwriting Chinese characters is both time-consuming and subjective. To address this issue, in our third objective we designed a method of encoding the results of CAM such that the differences between autistic and TD children’s handwritings can be objectively and swiftly observed. This approach enhances the efficiency of analyzing Chinese handwriting characteristics associated with the autistic and TD children.

The remainder of the paper is organized as follows. Section 2 presents the related work on using handwriting data to detect ASD. Section 3 introduces our dataset and the neatness label. In Section 4, the flowchart of our work and the various models proposed are provided. Section 5 covers the experiments and the results, and the CAM analyses of the handwriting characteristics. The conclusion is finally given in Section 6.

# Related Work

The integration of various data modalities has been a longstanding strategy in the detection of ASD tendencies. This approach stems from the intricate nature of ASD symptoms, which manifest across diverse domains. In the realm of social interactions, individuals diagnosed with ASD often face challenges in establishing and sustaining eye contact during natural communications. Consequently, researchers explored the use of eye-tracking technologies as a potential tool for detecting ASD [24, 25]. ASD, being a neurodevelopmental disorder, is associated with developmental alterations in both brain structure and facial tissues [26]. Accordingly, studies have delved into utilizing brain imaging techniques [27-29] and facial image analysis [26, 30] to contribute to ASD detection efforts. Moreover, ASD frequently coexists with language impairments, prompting investigations into linguistic aspects for diagnostic insights. Researchers have leveraged speech spectrograms to discern patterns indicative of ASD [31], and natural language processing techniques have been applied to analyze narrative expressions for potential ASD markers [32]. This multifaceted approach recognizes the heterogeneity of ASD symptoms and aims to capture the disorder's varied manifestations through a combination of data sources and analytical methods. These modalities offer valuable insights into neurodevelopmental patterns and cognitive processes associated with ASD. Over the years, researchers have extensively explored the potential of these advanced technologies to enhance our understanding of ASD and contribute to early and accurate diagnosis.

Autistic children encounter difficulties initiating actions that result in subsequent movements or ultimate goals, making them notably challenged in tasks involving sequential motor skills [9]. Handwriting is one such task. For example, a deviation in the initial stroke can impact the entire word's positioning. Handwriting analysis offers insights into a child's motor control, coordination, and cognitive processes during such a task. The subtleties in stroke irregularities, letter shapes, and overall handwriting style may serve as potential indicators of neurodevelopmental variations that are characteristics of ASD. By focusing on this relatively observable aspect of a child's behavior, our study aims to contribute to the growing body of research exploring unconventional yet promising avenues for ASD detection. In the study conducted by Beversdorf et al. [17], an examination of handwriting samples from adults with and without ASD revealed a noteworthy difference in letter size between the ASD group and the control TD group. Specifically, individuals with ASD tended to produce significantly larger letters compared to their counterparts. Another relevant study by Johnson et al. [18] focused on autistic children and shed light on spatial aspects of their handwriting. The research disclosed that autistic children exhibited poorer spatial arrangement in their handwriting when compared to TD children. This indicates challenges in organizing and spacing characters appropriately on the writing surface, suggesting that spatial arrangement might be a distinctive feature in the handwriting of individuals with ASD.

Hendr et al. [33] collected data from 104 participants, comprising of 51 individuals with ASD and 53 without ASD. Each participant completed a total of 18 tasks, encompassing shapes, numbers, and English, as illustrated in Figure 3. The tasks were divided into three sets, each set focusing on different instructions. The initial three tasks instructed participants to draw circles, triangles, and squares within provided outlines. Subsequently, tasks 4 to 6 required participants to draw circles, triangles, and squares without the presence of outlines. Moving on, tasks 7 to 9 instructed participants to write the numerals 1, 2, 3, 4, and 5 using ruled lines. Tasks 10 to 12 maintained a similar theme without ruled lines. Likewise, tasks 13 to 15 and tasks 16 to 18 instructed participants to write the phrases “cat and dog” with and without the guidance of ruled lines, respectively.

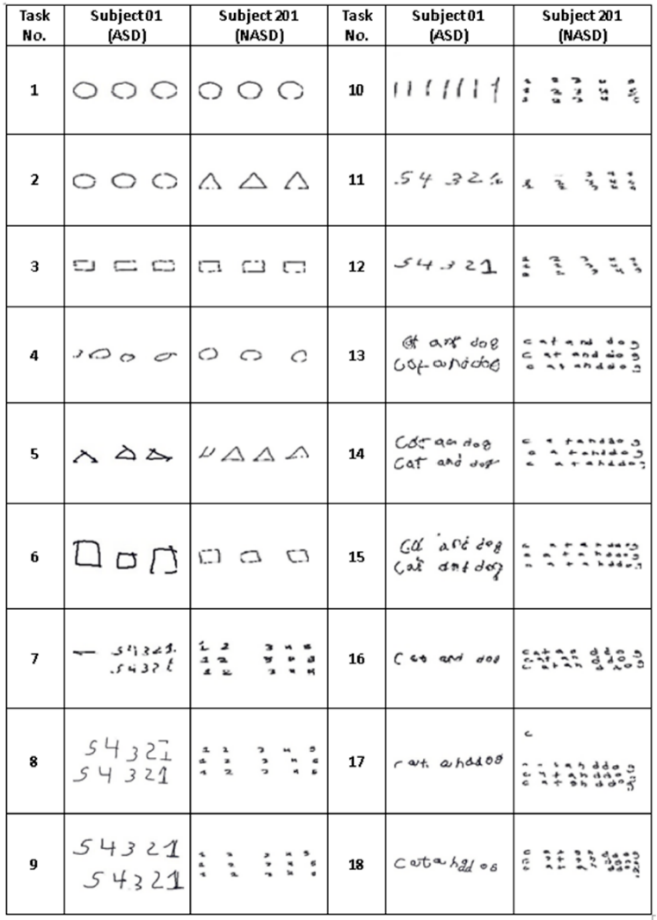


Figure 3: A total of 18 tasks for the ASD and non-ASD individuals [33]

After pre-processing steps such as cropping and rescaling, these handwriting pictures were input into a CNN model corresponding to each task. The final outcome was determined by the majority output of the 18 tasks. As the study made predictions on a participant-by-participant basis, the testing set included 18 participants and achieved an accuracy of 90.48%.

Yen et al. [20] collected a handwriting dataset for autistic children that included both traditional Chinese characters and phonetic notations. However, the phonetic notations data were not utilized. The data were input into the classification model in Chinese character units, aiming to classify whether the handwriting characters were produced by individuals with ASD or TD. Due to the uneven distribution of data (more data from TD individuals than ASD individuals), there was an issue of data imbalance. The study addressed data imbalance through undersampling, achieving an F1-score of 0.954. However, the approach of performing undersampling before splitting the dataset into training and testing sets resulted in inconsistent testing sets across different experiments. In our study, we aim to investigate whether incorporating phonetic notation data contributes to improved model performance. Moreover, we split the dataset into training and testing sets and applied both undersampling and oversampling exclusively on the training set to address the issue of inconsistent testing sets.

Additionally, Yen et al. [20] applied CAM [23] to observe the model's classification results and identify handwriting characteristics associated with ASD, such as turning strokes, alignment nuances, and spatial distribution. However, direct observation of the results of applying CAM is often time-consuming and subjective. In this study, we designed an encoding method for the results of applying CAM to objectively and swiftly observe the differences between the handwriting strokes of the ASD and TD individuals.

# Datasets

This study used the same dataset collected and introduced in Yen et al. [20]. We collaborated with a local elementary school and an association of autism to recruit participants. Rather than conducting an experiment to collect children’s handwriting as was done in Hendr et al. [33], we asked students to provide their handwriting workbooks in which they practiced Chinese characters and phonetic notations in class and at home. This way, we were able to collect handwriting that is relatively natural, providing us with potentially more accurate insights into the handwriting differences between ASD and TD children. In total, the dataset comprises handwritings from 23 children: 6 ASD children (average age: 10.5 years; all boys) and 17 TD peers (average age: 8.67 years, eight boys and nine girls). Among them, there are a total of 4 children with mild autism, 1 individual with moderate autism, and 1 individual with severe autism. Compared with Yen et al. [20], this study includes one additional autistic child. This child only provided phonetic notation data and was therefore not included in Yen et al. [20].

We used three types of datasets in the analysis: (1) the Chinese character-only dataset, (2) the phonetic notation-only dataset, and (3) the Chinese character + phonetic notation dataset where each Chinese character was written with its corresponding phonetic notation.

3.1 Chinese Character-Only Dataset

The Chinese character-only dataset comprises 17,950 words, with 14,173 from TD children and 3,777 from autistic children. We considered neatness of writing Chinese characters by defining the neatness criteria based on relevant literature [34, 35] and discussions with experienced elementary school teachers.

3.1.1 Neatness criteria

The neatness criteria were divided into two levels: stroke and component, and three factors: position, size, and correctness. The division into stroke and component levels allows for a more nuanced and comprehensive assessment of neatness. The stroke level pertains to the individual strokes that constitute a character, examining their position, size, and correctness. On the other hand, the component level considers the overall arrangement and coherence of the entire character, providing a holistic perspective on neatness. Specifically, at this level, size and position considerations extend beyond individual strokes to encompass the relative size and position between components. This broader view ensures that the evaluation captures not only the precision of each stroke but also the proportional relationships and overall structure of the characters. This multi-level approach ensures a detailed and nuanced evaluation, facilitating a more precise understanding of handwriting neatness in the context of our study. Therefore, we accessed the following six aspects: *stroke position, stroke size, stroke correctness, component position, component size*, and *component correctness* for each word. The neatness is labeled as “1” if the individual word satisfies five or more aspects, and “0” if it satisfies less than five aspects. Figure 4 exemplifies how we evaluated and labeled the neatness of each word.

Figure 4(a) demonstrates aspects not satisfied in *stroke position* and *component correctness*. The horizontal stroke in the component “女” is shifted downward, and the “日” component is miswritten. Therefore, It is labeled as 0. Figure 4(b) demonstrates only one aspect not satisfied in *component position* where the rightmost component of the character is shifted downward, and is labeled as 1. Figure 4(c) demonstrates an aspect not satisfied in *stroke size,* as there is an overly long horizontal stroke at the top, and another aspect not satisfied in *component position* where the “比” component is displaced. It is labeled as 0. Figure 4(d) demonstrates only one aspect not satisfied in *component size* where the “口” component is written too large, and is labeled as 1. Figure 4(e) demonstrates an aspect not satisfied in *stroke correctness* in the upper-left horizontal stroke and the right vertical stroke. Moreover, another aspect is not satisfied in *stroke position* since some strokes in Figure 4(e) exceed the grid. Therefore, it is labeled as 0. Finally, Figure 4(f) demonstrates an aspect not satisfied in *component correctness* where the “口” is written as a circle, and another aspect not satisfied in *stroke position* since strokes in Figure 4(f) exceed the grid. It is labeled as 0.



Figure 4: Example aspects not satisfied by the neatness criteria

3.1.2 Evaluation of neatness criteria

To assess the reliability of the above criteria, we employed the Fleiss' Kappa value [36], which is used to measure the degree of agreement between raters. In our study, we randomly selected 3600 pictures, which constitute 20% of the Chinese character-only dataset. The labels were made by three raters according to the neatness criteria mentioned in Section 3.1.1, and a majority vote was taken. The label was then compared to the author's label. The final calculated Kappa value was 0.7354, falling within the range of 0.61 to 0.80, indicating substantial agreement [37].

In conclusion, the author labeled all the pictures in the Chinese character-only dataset with 14,840 having a neatness label of 1 and 3,110 having a neatness label of 0. Detailed information is presented in Table 1. The value of the Chi-square test is 8824.91, with a *p*-value < 0.01, which shows TD children were more likely to produce neat Chinese characters than autistic children.

Table 1: Detailed information of the Chinese character-only dataset

|  |  |  |  |
| --- | --- | --- | --- |
|  | TD | ASD | Total |
| Neatness = 1 | 13,659 | 1,181 | 14,840 |
| Neatness = 0 | 514 | 2,596 | 3,110 |
| Total | 14,173 | 3,777 | 17,950 |

3.2 Phonetic Notation-Only Dataset

Our second dataset was the Phonetic notation-only dataset. In pursuit of our first objective of assessing the potential contribution of phonetic notations to the task at hand, we evaluated the impact of incorporating phonetic notations in the classification of the handwritings. In alignment with the established methodologies, we implemented the same data processing procedure as delineated in Yen et al. [20] to meticulously extract phonetic notations from the grids within the workbooks. The resultant dataset, denoted as the Phonetic notation-only dataset, encompassed a total of 18,833 images. Within this dataset, 14,943 instances pertained to TD children, while 3,890 instances were from autistic children. The exploration of the Phonetic notation-only dataset held unique significance in our study. While the discussion of the Chinese character-only dataset provided a detailed analysis of the handwriting characteristics of ASD and TD children, examining the Phonetic notation-only dataset was an essential corresponding part. By introducing the phonetic notations into the classification task, we aimed to identify unique patterns that might have emerged with the inclusion of phonetic data.

3.3 Chinese Character + Phonetic Notation Dataset

Our third dataset was the Chinese character + Phonetic notation dataset. The Chinese character + Phonetic notation dataset aggregated a total of 17,126 images. Among these, 13,687 instances pertained to TD children, while 3,439 instances to ASD children. This dataset served as a comprehensive collection that incorporates both Chinese characters and their corresponding phonetic notations. Table 2 provides detailed number of images in each of the three datasets. The Chinese character + Phonetic notation dataset occupied a unique position within our study, bridging the characteristics of the Chinese character-only dataset and Phonetic notation-only dataset. It acted as a crossroads, offering insights into the interplay and potential synergies between handwriting Chinese characters and their associated phonetic notations. The reduced number of instances in the Chinese character + Phonetic notation dataset came from the fact that there exist data with only Chinese characters (Figure 5(a)), and only phonetic notations (Figure 5(b)).

一張含有 筆跡, 行, 字型, 圖表 的圖片

自動產生的描述

Figure 5: (a) Only handwriting Chinese character, (b) Only handwriting phonetic notation

Table 2: Statistics of the three datasets

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | TD | ASD | Total |
| Chinese character-only | 14,173 | 3,777 | 17,950 |
| Phonetic notation-only | 14,943 | 3,890 | 18,833 |
| Chinese character + Phonetic notation | 13,687 | 3,439 | 17,126 |

# Method

Our flowchart of building classification models is divided into two types: a 5-fold flowchart, as shown in Figure 6, and the CAM flowchart, depicted in Figure 7. The 5-fold flowchart signifies our utilization of 5-fold cross-validation during model training. We used the average of the results from the 5 folds for evaluation. On the other hand, the CAM flowchart employed the results of applying CAM for identifying the handwriting characteristics. This required an individual model rather than an average of 5 models. Below, we introduce each step in the flowchart in order.

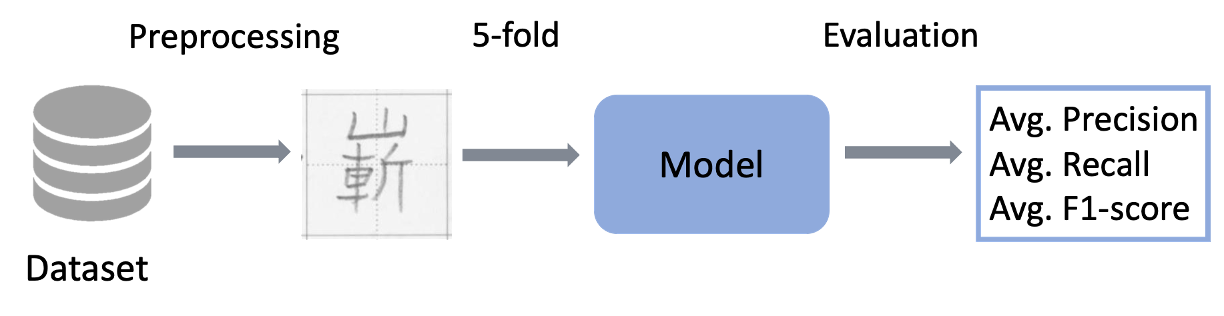


Figure 6: The 5-fold flowchart

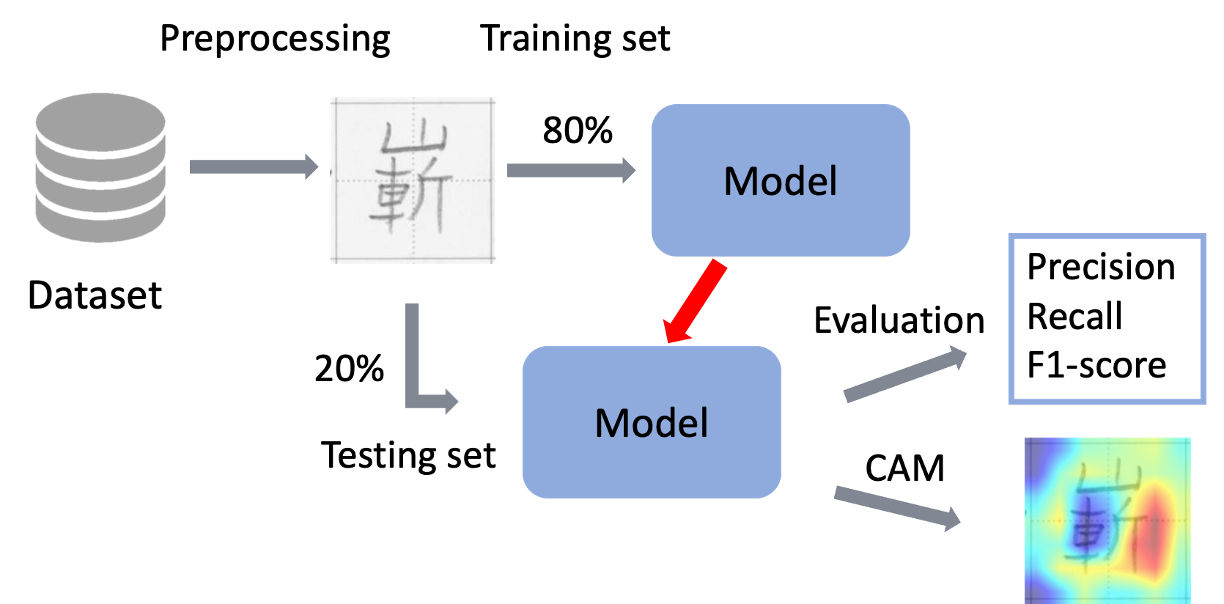


Figure 7: The CAM flowchart

We used Precision, Recall, and F1-score as metrics, as shown in formulas (1), (2), and (3). TP stands for True Positive, FP for False Positive, and FN for False Negative. Precision represents the accuracy of the model's positive predictions. Recall is the proportion of positives correctly predicted by the model out of all actual positives. The F1-score is the harmonic mean of Precision and Recall, providing a comprehensive metric for model evaluation.

(1)

(2)

(3)

## Data Preprocessing

Both flowcharts employed the same data preprocessing. First, we converted the image to grayscale. The second step involved resizing the image to 224x224. This is because ResNet-18 [38] requires an input image of size 224x224.

## Dataset Balance

As seen in Table 1, our data was imbalanced. To address this issue, we applied oversampling and undersampling techniques to the data in the training set. In the case of the 5-fold flowchart, oversampling or undersampling was applied to the training set in each fold.

When undertaking undersampling, we randomly selected a number of the majority dataset to match the size of the minority dataset. As depicted in Figure 8, there are 14,360 images in the training set, comprising 11,338 TD and 3,022 ASD images. We randomly selected 3,022 images from the 11,338 TD images. Consequently, the final undersampled training set consists of a total of 6,044 images.

When conducting oversampling, we employed duplicate oversampling, wherein we duplicated images from the minority dataset to a multiple close to, but not exceeding, the majority dataset. As illustrated in Figure 8, there are 14,360 images in the training set, consisting of 11,338 TD and 3,022 ASD images. In this scenario, we duplicated the ASD images to three times their original size. Consequently, the final oversampled training set comprised a total of 20,404 images.

一張含有 文字, 螢幕擷取畫面, 圖表, 設計 的圖片

自動產生的描述

Figure 8: Example of undersampling and oversampling

## Model

In terms of models, we employed support vector machine (SVM) [39], decision tree (DT) [40], K-nearest neighbor (KNN) [41], and logistic regression (LR) [42]—commonly used models in machine learning that are applicable to classification tasks. SVM operates by identifying a hyperplane that best separates the data into different classes while maximizing the margin between them. Its ability to handle complex decision boundaries and adapt to high-dimensional spaces makes SVM a valuable choice for our study. DT is an intuitive model that makes decisions based on a tree-like structure of conditional statements. Each node in the tree represents a decision based on a particular feature, leading to subsequent nodes until a final decision is reached. It captures non-linear relationships within the data and is easily interpretable, facilitating insights into the decision-making process. KNN is a non-parametric and instance-based learning algorithm. It classifies new instances based on the majority class of their k-nearest neighbors in the feature space. It is particularly useful for capturing local patterns and is effective in scenarios where instances of the same class tend to cluster together in the feature space. Despite its name, LR is a classification algorithm commonly used for multiclass classification tasks. It models the probability of an instance belonging to a particular class using the logistic function. It is computationally efficient, interpretable, and provides probabilities for class membership, making it a staple in classification tasks.

Additionally, we utilized the ResNet-18 [38] model, a widely adopted architecture in the field of computer vision. Its innovative residual connection design addresses the challenge of training deep convolutional neural networks by mitigating issues associated with excessive layer depth. The ResNet architecture's skip connections enable the direct flow of information, facilitating the training of deeper networks without suffering from degradation issues. In this study, we used a ResNet-18 model pretrained on ImageNet [43].

## Class Activation Map

Class Activation Map (CAM) [44] serves as a crucial visualization tool that facilitates a deeper understanding of a model's focal points. This technique involves establishing a connection with the Global Average Pooling (GAP) layer after the final convolutional layer. Following this connection, it captures the weights associated with the GAP layer output and linearly combines them with the corresponding feature map to generate the results. The conventional approach outlined above mandates the utilization of the GAP layer, imposing constraints on the overall flexibility of the network architecture.

To overcome this limitation, Grad-CAM [23] introduces an innovative solution by incorporating the partial differential of the feature map in the relevant category to supplant the weight output derived from the GAP layer. This modification enables Grad-CAM to be applied across a broader spectrum of CNN architectures, and therefore offers enhanced adaptability. In the context of this study, we adopted the Grad-CAM approach, aligning with the methodology employed by Yen et al. [20]. This ensured consistency and enabled us to leverage the proven effectiveness of Grad-CAM in visualizing and interpreting the focus areas of our model.

# Experiments

The neatness labels utilized in this study were manually annotated by human coders, which is a time-consuming process. In this section, we first presented the performance of a model built for automatic neatness labeling. If the model performs well, it can be used to automatically do the labeling in the future.

Next, we presented experiments on ASD/TD classification, which constituted the primary focus of this study. We investigated the effectiveness of distinguishing between Chinese characters written by ASD children and TD children using only neat Chinese characters. Additionally, we explored whether adding phonetic notations could enhance the model's performance. We also attempted domain adaptation to assess the model's ability to accurately predict data from different datasets.

Finally, we conducted the analyses of the handwriting characteristics. After encoding the CAM results, we statistically observed the trend of the model's focus. We aimed to analyze whether there were differences in focus between the training set in Chinese character-only dataset and neat Chinese character-only dataset.

## Neatness Classification

In the neatness classification experiments, we utilized SVM, DT, KNN, LR, and ResNet-18 models, each with three methods to deal with the data imbalance problem: undersampling, oversampling, and X which means neither undersampling nor oversampling techniques were applied. All results were presented in Table 3, and the best performance of each model was summarized in Table 4.

Observing Table 3, the undersampling practice increased the proportion of “Neatness = 0” in the training set, making the model more inclined to predict “Neatness = 0.” This led to an improvement in the Recall compared to the imbalanced approach. However, the Precision experienced a significant drop due to the bolder prediction of “Neatness = 0.” On the other hand, oversampling exhibited higher stability in the Precision than undersampling. Table 4 revealed that ResNet-18 with oversampling yielded the best performance in this task.

Table 3: All results from neatness classification

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Balance | 5-fold | Precision | Recall | F1-score |
| SVM | X | V | 0.6637 | 0.6948 | 0.6787 |
| SVM | Undersampling | V | 0.4795 | 0.8291 | 0.6075 |
| SVM | Oversampling | V | 0.6624 | 0.6955 | 0.6782 |
| DT | X | V | 0.6927 | 0.7448 | 0.7174 |
| DT | Undersampling | V | 0.4744 | 0.8337 | 0.6046 |
| DT | Oversampling | V | 0.7378 | 0.7314 | 0.7346 |
| KNN | X | V | 0.6138 | 0.7683 | 0.6823 |
| KNN | Undersampling | V | 0.3212 | 0.8847 | 0.4710 |
| KNN | Oversampling | V | 0.3574 | 0.8621 | 0.5050 |
| LR | X | V | 0.8438 | 0.6714 | 0.7476 |
| LR | Undersampling | V | 0.5333 | 0.8153 | 0.6447 |
| LR | Oversampling | V | 0.6425 | 0.7689 | 0.6998 |
| ResNet-18 | X | V | 0.8544 | 0.7309 | 0.7804 |
| ResNet-18 | Undersampling | V | 0.7798 | 0.7955 | 0.7793 |
| ResNet-18 | Oversampling | V | **0.8281** | **0.7799** | **0.7997** |

Table 4: The best performance of each model in the neatness classification

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Balance | 5-fold | Precision | Recall | F1-score |
| SVM | X | V | 0.6637 | 0.6948 | 0.6787 |
| DT | Oversampling | V | 0.7378 | 0.7314 | 0.7346 |
| KNN | X | V | 0.6138 | 0.7683 | 0.6823 |
| LR | X | V | 0.8438 | 0.6714 | 0.7476 |
| ResNet-18 | Oversampling | V | **0.8281** | **0.7799** | **0.7997** |

Based on the experiment results presented in Table 4, it is evident that the ResNet-18 model with oversampling performed the best among all models and methods, achieving an F1-score of 0.7997. This result indicated that the ResNet-18 model, when trained with oversampling to address data imbalance, excelled in both precision and recall compared to other models and methods tested.

Overall, the results suggested that the ResNet-18 model with oversampling is most suitable for classifying neatness labels in handwriting. More experiments and analyses may be needed to further improve the performance of the model for a practical use.

## ASD/TD Classification

We introduced the notations of Ch\_All, Ch\_Neat, Ch\_Mild, Ph\_All, and Ch+Ph in Table 5. These notations represent different training sets. In the subsequent experiments, we employed ResNet-18 as the model. Under different training sets Ch\_All, Ch\_Neat, Ch\_Mild, Ph\_All, and Ch+Ph, the corresponding testing sets were obtained from the respective datasets. Specifically, the testing set for Ch\_All, Ch\_Neat and Ch\_Mild came from the Chinese character-only dataset, the testing set for Ph\_All came from the Phonetic notation-only dataset, and the testing set for Ch+Ph came from the Chinese character + Phonetic notation dataset. This ensures that Ch\_All, Ch\_Neat and Ch\_Mild have the same testing set and can make fair comparisons.

Table 5: The meanings of Ch\_All, Ch\_Neat, Ch\_Mild, Ph\_All, and Ch+Ph

|  |  |
| --- | --- |
|  | Meanings |
| Ch\_All | The training set in Chinese character-only dataset |
| Ch\_Neat | The neat training set in Chinese character-only dataset |
| Ch\_Mild | The training set in Chinese character-only dataset from the mild ASD children |
| Ph\_All | The training set in Phonetic notation-only dataset |
| Ch+Ph | The training set in Chinese character + Phonetic notation dataset |

5.2.1 ASD/TD classification using Ch\_All

Table 6 presented the results of ASD/TD classification using the Chinese character-only dataset, along with the results from Yen et al. [20]. The three employed methods for dealing with imbalance data were X, undersampling, and oversampling. Notably, the undersampling without 5-fold approach achieved the highest F1-score in Yen et al. [20].

From Table 6, it was evident that both our X and undersampling approaches surpassed the results of Yen et al. [20]. Additionally, the oversampling approach outperformed the best performance reported in Yen et al. [20]. Consequently, we used the oversampling approach in our implementation as the benchmark for the analyses of the handwriting characteristics.

Table 6: The results of ASD/TD classification using Ch\_All

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data Balance | 5-fold | Approach | Precision | Recall | F1-score |
| X | V | Ours | 0.9780 | 0.9224 | 0.9490 |
| Undersampling | V | Ours | 0.9388 | 0.9891 | 0.9629 |
| Oversampling | V | Ours | **0.9807** | **0.9637** | **0.9720** |
| X | V | Yen et al. | 0.982 | 0.762 | 0.856 |
| Undersampling | V | Yen et al. | 0.956 | 0.911 | 0.932 |
| Undersampling | X | Yen et al. | 0.934 | 0.975 | 0.954 |

5.2.2 ASD/TD classification using Ch\_Neat

Unlike Ch\_All, Ch\_Neat excluded data with “Neatness = 0,” retaining only neat Chinese characters. The results of using the Ch\_Neat training set were displayed in the lower three rows in Table 7, while the upper three rows presented the results of using the Ch\_All training set for a comparison.

Using Ch\_Neat fell short of surpassing Ch\_All in X, undersampling, and oversampling. This is because Ch\_Neat excluded data with “Neatness = 0.” Despite the reduced amount of data, Ch\_Neat still achieved similar results compared to Ch\_All in oversampling. This demonstrated the feasibility of classifying ASD/TD using only neat Chinese characters.

Table 7: The results of ASD/TD classification using Ch\_All and Ch\_Neat

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Balance | 5-fold | Precision | Recall | F1-score |
| Ch\_All | X | V | 0.9780 | 0.9224 | 0.9490 |
| Ch\_All | Undersampling | V | 0.9388 | 0.9891 | 0.9629 |
| Ch\_All | Oversampling | V | **0.9807** | **0.9637** | **0.9720** |
| Ch\_Neat | X | V | 0.9804 | 0.8203 | 0.8881 |
| Ch\_Neat | Undersampling | V | 0.9638 | 0.9281 | 0.9447 |
| Ch\_Neat | Oversampling | V | 0.9739 | 0.9584 | 0.9658 |

5.2.3 ASD/TD classification using Ch\_Mild

The Ch\_Mild training set selectively retained the data from the mild ASD children while keeping all the data from the TD children. As depicted in Table 8, undersampling demonstrated the best model performance. We attributed the suboptimal performance of oversampling to the limited data from ASD children. Oversampling copied a larger amount of data from the mild ASD children, potentially leading to overfitting.

Table 8: The results of ASD/TD classification using Ch\_Mild

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Balance | 5-fold | Precision | Recall | F1-score |
| X | V | 0.9828 | 0.7652 | 0.8604 |
| Undersampling | V | **0.9690** | **0.8992** | **0.9318** |
| Oversampling | V | 0.9777 | 0.8444 | 0.9049 |

5.2.4 ASD/TD classification using Ph\_All

As indicated in Table 9, the model trained using Ph\_All failed to surpass Ch\_All under X, undersampling, and oversampling. Despite having the largest amount of data, relying solely on phonetic notations did not enhance the model's performance in this classification task.

Table 9: The results of ASD/TD classification using Ph\_All

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Balance | 5-fold | Precision | Recall | F1-score |
| X | V | 0.9520 | 0.8878 | 0.9171 |
| Undersampling | V | 0.8983 | 0.9702 | 0.9294 |
| Oversampling | V | **0.9250** | **0.9658** | **0.9424** |

5.2.5 ASD/TD classification using Ch+Ph

Despite having the smallest amount of data, the results achieved were similar to those of using Ch\_All, as demonstrated in Table 10. Combining the results from Sections 5.2.4 and 5.2.5, we concluded that adding phonetic notations did not enhance the performance of the model.

Table 10: The results of ASD/TD classification using Ch+Ph

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Balance | 5-fold | Precision | Recall | F1-score |
| X | V | 0.9780 | 0.9074 | 0.9406 |
| Undersampling | V | 0.9408 | 0.9624 | 0.9495 |
| Oversampling | V | **0.9624** | **0.9714** | **0.9663** |

5.2.6 Domain adaptation

In the experiments with domain adaptation, we aimed to evaluate how well the model predicts across datasets. First, we presented the performance on the respective datasets in Table 11. Similar to the previous results using 5-fold verification, the model performance from high to low is Chinese character-only, Chinese character + Phonetic notation, and Phonetic notation-only.

For the domain adaptation experiments, we first divided the training and testing sets for the Chinese character + Phonetic notation dataset at a ratio of 8:2. For the Chinese character-only (and Phonetic notation-only) dataset, we put the Chinese characters (and Phonetic notations) which appeared in the Chinese character + Phonetic notation dataset in the corresponding training and testing sets. For those Chinese characters (and Phonetic notations) in the character-only (and Phonetic notation-only) dataset, which did not appear in the Chinese character + Phonetic notation dataset, we divided them into the training and testing sets at the same ratio of 8:2, and combined them with the previously formed training and testing sets, respectively. This meticulous approach ensured that no data appeared in both the training and testing sets simultaneously during cross-dataset classifications.

Table 11: Results of model training and testing on the three datasets

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data Balance | 5-fold | Training set | Testing set | Precision | Recall | F1-score |
| Oversampling | X | Chinese character-only | Chinese character-only | **0.9737** | **0.9893** | **0.9814** |
| Oversampling | X | Chinese character + Phonetic notation | Chinese character + Phonetic notation | 0.9867 | 0.9709 | 0.9788 |
| Oversampling | X | Phonetic notation-only | Phonetic notation-only | 0.9414 | 0.9742 | 0.9575 |

Table 12 shows the results of training and testing on different datasets. It was evident from rows 1 and 3 in Table 12 that the training and testing sets are from the Chinese character-only and Chinese character + Phonetic notation datasets exhibited better performance. If either the training set or testing set is from the Phonetic notation-only dataset, the results dropped significantly. This is attributed to the fact that Chinese characters are more complex (and therefore contain more information) compared to phonetic notations. Additionally, when the training set was from the Chinese character + Phonetic notation dataset, the performance on the testing set from the Chinese character-only dataset was better than the testing set from the Phonetic notation-only dataset. This again indicates that the model tended to learn more effectively from the Chinese characters.

Table 12: Domain adaptation on the three datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Training set | Testing set | Precision | Recall | F1-score |
| Chinese character-only | Chinese character + Phonetic notation | 0.9642 | 0.9782 | 0.9711 |
| Chinese character-only | Phonetic notation-only | 0.8239 | 0.6039 | 0.6969 |
| Chinese character + Phonetic notation | Chinese character-only | 0.9817 | 0.9332 | 0.9568 |
| Chinese character + Phonetic notation | Phonetic notation-only | 0.8650 | 0.4877 | 0.6238 |
| Phonetic notation-only | Chinese character-only | 0.3392 | 0.9479 | 0.4996 |
| Phonetic notation-only | Chinese character + Phonetic notation | 0.3757 | 0.9666 | 0.5411 |

## Identifying Handwriting Characteristics

In this subsection, we aimed to explore the differences between the performance using the Ch\_All and Ch\_Neat training sets. The performance of these two training sets was detailed in Table 13, and following the trend observed in Section 5.2.2, using Ch\_All slightly outperformed Ch\_Neat.

Table 13: Results of using Ch\_All and Ch\_Neat without 5-fold verification

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Training set | Balance | 5-fold | Precision | Recall | F1-score |
| Ch\_All | Oversampling | X | 0.9789 | 0.9841 | 0.9815 |
| Ch\_Neat | Oversampling | X | **0.9865** | **0.9656** | **0.9759** |

Our testing set comprised a total of 3590 images. For each image, there were four possible prediction results: both correct, both wrong, Ch\_All correct and Ch\_Neat wrong, and Ch\_All wrong and Ch\_Neat correct as illustrated in Figure 9.

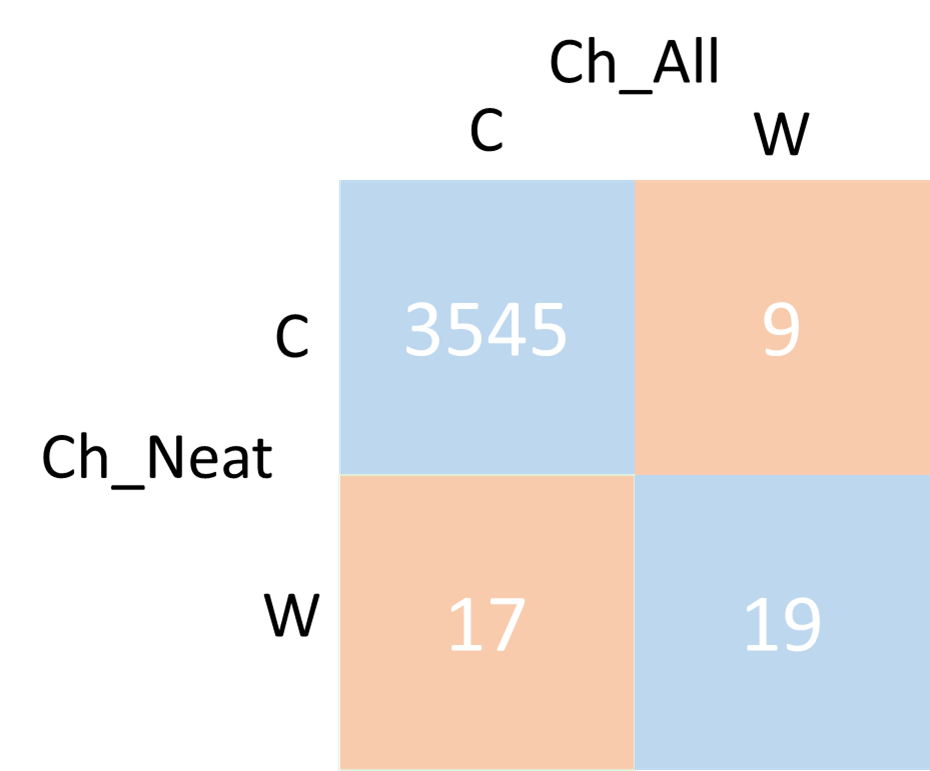


Figure 9: The prediction of using Ch\_All and Ch\_Neat

Out of the 3590 images, 3545 were both correctly predicted, 19 were both incorrectly predicted, and only 26 showed different prediction results. Despite having different training sets, their prediction results were very similar. In the subsequent analysis, we examined both models from a CAM perspective.

In order to solve the time-consuming and the subjective problems of manual observation of the CAM results, we processed these results in two steps. The first step was to specify a color area to divide the CAM result into two parts, as shown in Figure 10. We chose the red and orange regions (the two most focused regions) in the CAM result to form this color area. The second step was to use formula (4) to encode the result of the first step, as shown in Figure 11. In formula (4), each image (224\*224 pixels per image) was divided into 16 blocks (56\*56 pixels per block). If the block is conformed to formula (4), it was coded as 1, otherwise it was 0. The significance of this step was to encode the model focus since it likely refers to the area or features of the input data the model pays much attention to when making a prediction. Encoding the model focus involves converting the model focus into a format that can be visualized and analyzed.

一張含有 螢幕擷取畫面, 鮮豔 的圖片

自動產生的描述

Figure 10: First step of processing the CAM results

一張含有 螢幕擷取畫面, 圖表, 正方形 的圖片

自動產生的描述

Figure 11: Second step of processing the CAM results

(4)

5.3.1 Handwriting characteristics of not-centered images

We define a *not-centered image* as one where the surrounding 12 blocks were marked 1 or 0 and the middle four blocks were marked 0, as illustrated in Figure 12. The number of not-centered images was presented in Table 14. Among the not-centered images, the majority of them came from the TD children. Examples of the TD not-centered images were provided in Figure 13. This result is quite surprising. It is generally believed that Chinese characters written by TD children are of moderate size and centered, so the model's focus area would be more centralized [20]. On the other hand, Chinese characters written by ASD children tend to have offsets or be written too large, potentially leading the model's focus area to be on the periphery. However, the results show that for Chinese characters written by TD children, the model's focus area is not in the center. This indicates that the model has indeed found unexpected features to do the prediction.

一張含有 螢幕擷取畫面, 正方形, Rectangle, 鮮豔 的圖片

自動產生的描述

Figure 12: Illustration of a not-centered image

Table 14: Number of not-centered images

|  |  |  |  |
| --- | --- | --- | --- |
| Training set | Total | TD | ASD |
| Ch\_All | 2,273 | 2,200 | 73 |
| Ch\_Neat | 2,386 | 2,278 | 108 |

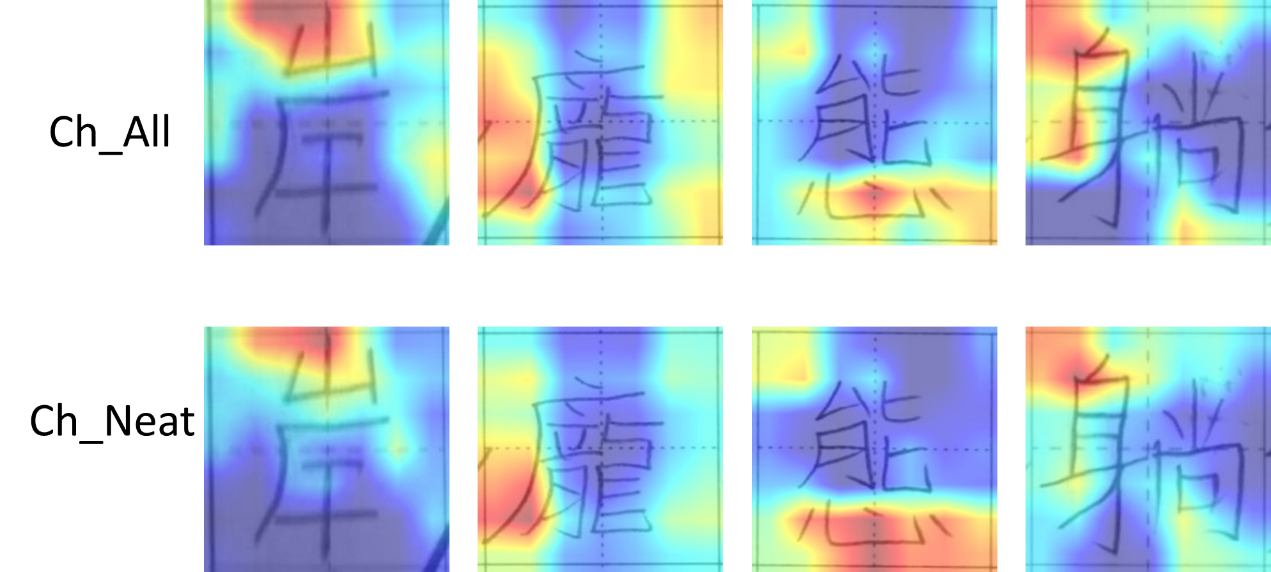


Figure 13: Examples of TD not-centered images

5.3.2 Handwriting characteristics of not-peripheral images

We define a *not-peripheral image* as one where the surrounding 12 blocks were marked 0 and the middle four blocks were marked 1 or 0, as illustrated in Figure 14. The number of not-peripheral images was outlined in Table 15. Although there were not many not-peripheral images, a significant proportion of these images came from the ASD children. Examples of ASD not-peripheral images were provided in Figure 15. This result further confirms that the features the model focuses on may differ from what we have expected.

一張含有 螢幕擷取畫面, 正方形, Rectangle, 行 的圖片

自動產生的描述

Figure 14: Illustration of a not-peripheral image

Table 15: Number of not-peripheral images

|  |  |  |  |
| --- | --- | --- | --- |
| Training set | Total | TD | ASD |
| Ch\_All | 44 | 4 | 40 |
| Ch\_Neat | 55 | 3 | 52 |

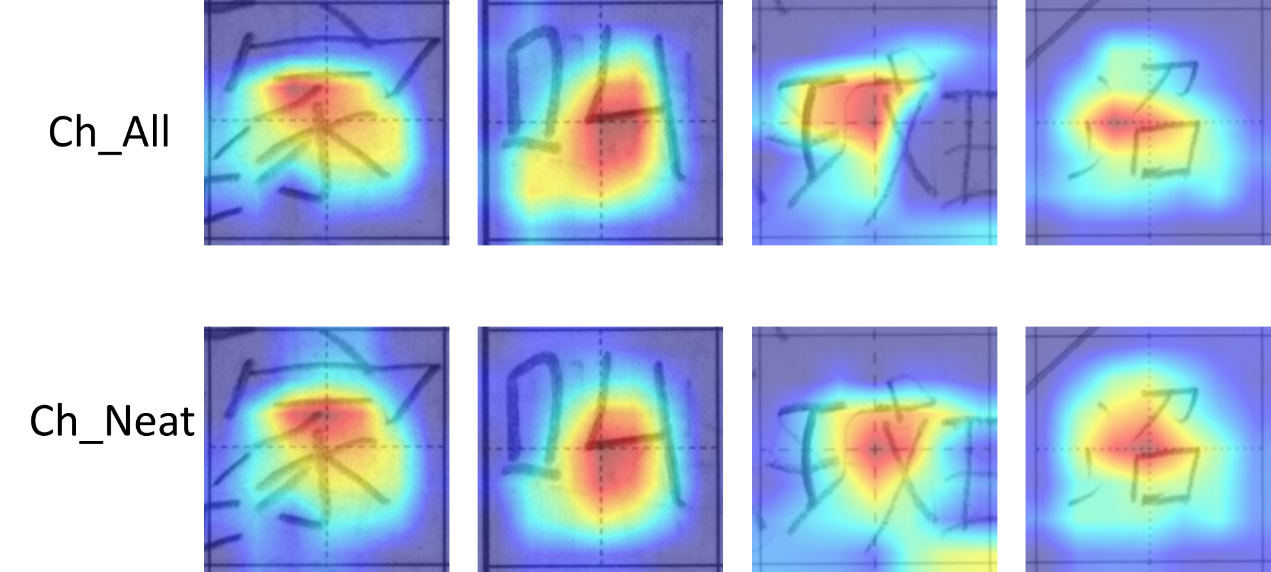


Figure 15: Examples of ASD not-peripheral images

5.3.3 Handwriting characteristics of corner images

We further analyzed the focus of the CAM by dividing an image into four corners, as illustrated in Figure 16. We define a *corner image* as one where the upper left (UL) corner or upper right (UR) corner or lower left corner (LL) or lower right (LR) corner is marked 1 or 0 while the other corners marked 0. The number of corner images was outlined in Table 16.

一張含有 正方形, 鮮豔, Rectangle, 螢幕擷取畫面 的圖片

自動產生的描述

Figure 16: Illustration of four corners

Table 16: Number of corner images

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Training set | UL (TD/ASD) | UR (TD/ASD) | LL (TD/ASD) | LR (TD/ASD) |
| Ch\_All | 524/7 | 78/5 | 61/12 | 79/66 |
| Ch\_Neat | 393/4 | 39/11 | 192/12 | 87/70 |

Observing the outcomes, using both training datasets exhibited similar tendencies. Specifically, the focus of the CAM on the images from the TD children was on UL while for ASD children, it tended to be on LR. Given the traditional habit of writing Chinese characters from top to bottom and from left to right, this pattern suggested that if a Chinese character was written by a TD child, the CAM focused on the starting stroke in the UL corner. Conversely, if the Chinese characters were written by an ASD child, the CAM tended to focus on the end stroke in the LR corner. This also resonated with the challenge faced by autistic children in initiating actions that result in subsequent movements or ultimate goals [9]. Examples of corner images were provided in Figures 17 and 18 for TD and ASD children respectively.

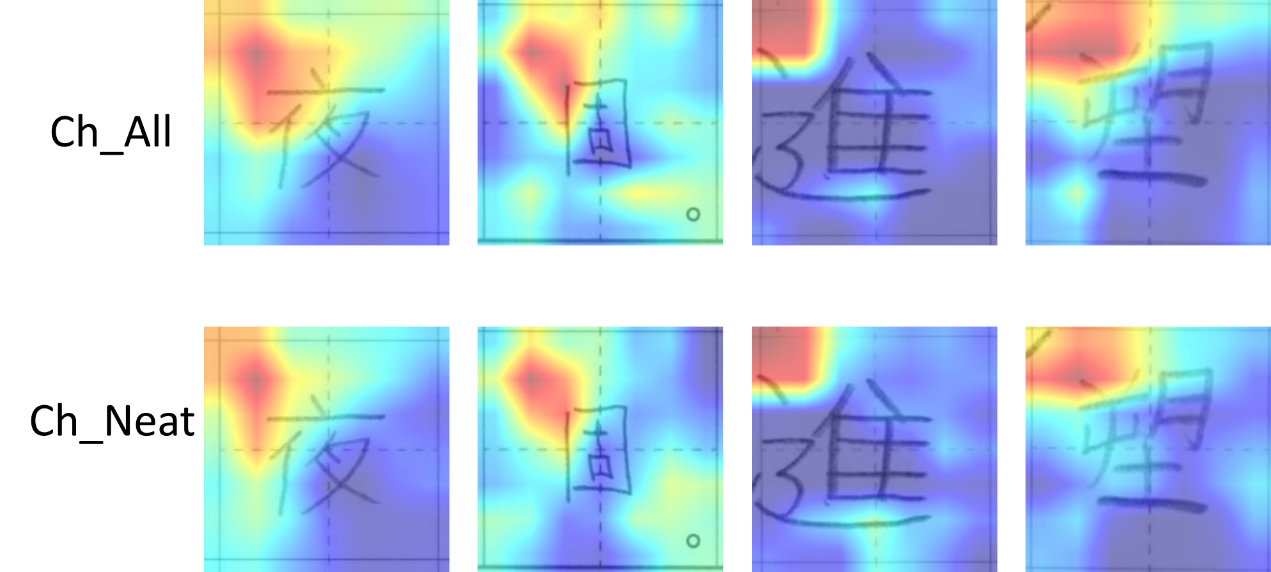


Figure 17: Examples of TD upper left corner images

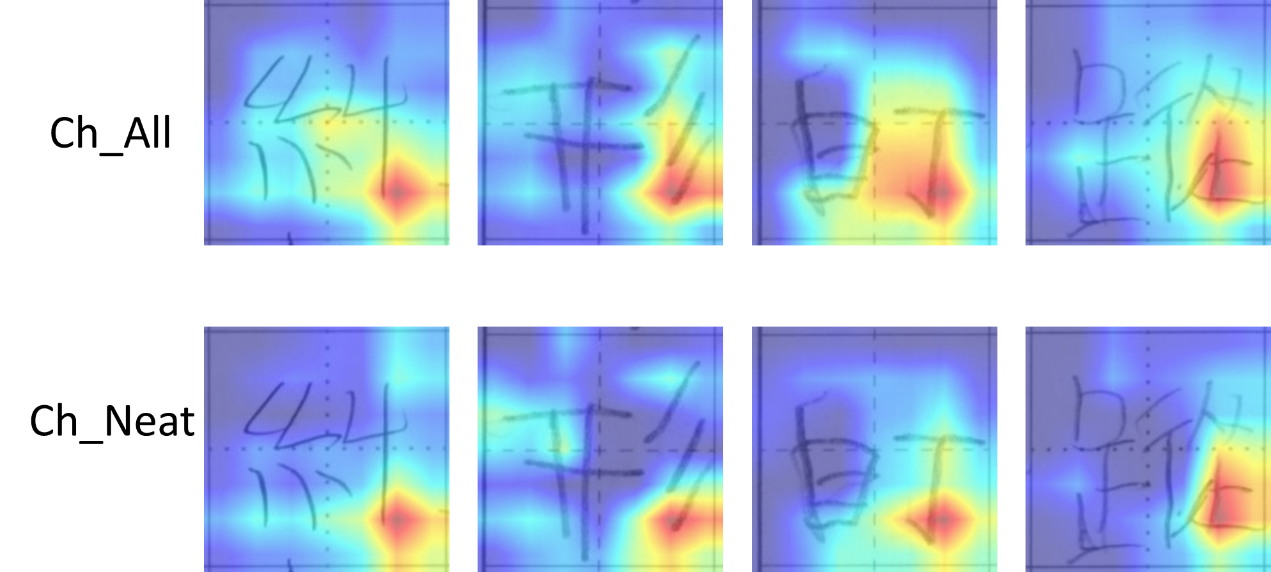


Figure 18: Examples of ASD lower right corner images

# Conclusion and Future Work

In this paper, we introduced the concept of handwriting neatness and defined its criteria for studying the handwriting characteristics of ASD children. By employing oversampling technique for data balancing, we surpassed the performance of the previous study on ASD/TD classification to achieve an F1-score of 0.9720 using the Ch-All training dataset. When using only neatly written Chinese characters, the F1-score was 0.9658. This demonstrates the model's capability to classify whether the Chinese characters were handwritten by ASD or TD children under neatly writing conditions. In subsequent experiments, we trained the model using the Phonetic notation-only dataset and Chinese character + Phonetic notation dataset. Based on the results, we concluded that adding phonetic notations did not enhance the model's performance. Finally, we encoded the CAM results to address the shortcomings associated with the manual observation which is time-consuming and subjective. It reveals that the prediction results and the CAM perspectives of the two training sets, Ch\_All and Ch\_Neat, are very similar. Moreover, it highlights differences between TD and ASD in the CAM results.

Despite the promising results, our study has some limitations. The data size, especially in the ASD category, may affect the generalization of the model. Future work should involve larger and more diverse datasets to ensure robustness and generalization. Our findings contribute to the understanding of the handwriting characteristics of ASD children, shedding light on potential differences that can aid in early detection and intervention of the ASD children. This knowledge can be valuable for teachers and parents in recognizing the unique challenges faced by ASD children in handwriting. The model, when refined and validated, could serve as a supportive tool for educators in assessing and guiding the development of ASD children's writing abilities.

# Reference

1. American Psychiatric Association, D. and A.P. Association, *Diagnostic and statistical manual of mental disorders: DSM-5*. Vol. 5. 2013: American psychiatric association Washington, DC.

2. Lord, C., et al., *Autism from 2 to 9 years of age.* Archives of general psychiatry, 2006. **63**(6): p. 694-701.

3. Hyman, S.L., et al., *Identification, evaluation, and management of children with autism spectrum disorder.* Pediatrics, 2020. **145**(1).

4. Downey, R. and M.J.K. Rapport, *Motor activity in children with autism: a review of current literature.* Pediatric Physical Therapy, 2012. **24**(1): p. 2-20.

5. Green, D., et al., *Impairment in movement skills of children with autistic spectrum disorders.* Developmental Medicine & Child Neurology, 2009. **51**(4): p. 311-316.

6. Fournier, K.A., et al., *Motor coordination in autism spectrum disorders: a synthesis and meta-analysis.* Journal of autism and developmental disorders, 2010. **40**: p. 1227-1240.

7. Fournier, K.A., et al., *Decreased dynamical complexity during quiet stance in children with autism spectrum disorders.* Gait & posture, 2014. **39**(1): p. 420-423.

8. Travers, B.G., et al., *Motor difficulties in autism spectrum disorder: linking symptom severity and postural stability.* Journal of autism and developmental disorders, 2013. **43**: p. 1568-1583.

9. Chen, L.-C., et al., *Postural control and interceptive skills in children with autism spectrum disorder.* Physical Therapy, 2019. **99**(9): p. 1231-1241.

10. Whyatt, C.P. and C.M. Craig, *Motor skills in children aged 7–10 years, diagnosed with autism spectrum disorder.* Journal of autism and developmental disorders, 2012. **42**: p. 1799-1809.

11. Mari, M., et al., *The reach–to–grasp movement in children with autism spectrum disorder.* Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 2003. **358**(1430): p. 393-403.

12. Sacrey, L.-A.R., et al., *Reaching and grasping in autism spectrum disorder: a review of recent literature.* Frontiers in neurology, 2014. **5**: p. 6.

13. David, F.J., et al., *A pilot study: coordination of precision grip in children and adolescents with high functioning autism.* Pediatric Physical Therapy, 2009. **21**(2): p. 205.

14. Fabbri-Destro, M., et al., *Planning actions in autism.* Experimental brain research, 2009. **192**: p. 521-525.

15. Mayes, S.D. and S.L. Calhoun, *Learning, attention, writing, and processing speed in typical children and children with ADHD, autism, anxiety, depression, and oppositional-defiant disorder.* Child Neuropsychology, 2007. **13**(6): p. 469-493.

16. Fuentes, C.T., S.H. Mostofsky, and A.J. Bastian, *Children with autism show specific handwriting impairments.* Neurology, 2009. **73**(19): p. 1532-1537.

17. Beversdorf, D.Q., et al., *Brief report: macrographia in high-functioning adults with autism spectrum disorder.* Journal of Autism and developmental disorders, 2001. **31**: p. 97-101.

18. Johnson, B.P., et al., *A quantitative comparison of handwriting in children with high-functioning autism and attention deficit hyperactivity disorder.* Research in autism spectrum disorders, 2013. **7**(12): p. 1638-1646.

19. Peebles, D.G., *Scml: A structural representation for Chinese characters.* 2007.

20. Yen, L., J. Wong, and A.L.P. Chen, *Identifying Chinese Handwriting Characteristics for Detecting Children with Autism.* The ACM/SIGAPP Symposium on Applied Computing, 2024.

21. Department of Lifelong Education, M.o.E., *The Manual of the Phonetic Symbols of Mandarin Chinese (Digital Version)*. First Edition ed. 2017: Pan, Wen-chung.

22. Taele, P. and T.A. Hammond. *A Geometric-based Sketch Recognition Approach for Handwritten Mandarin Phonetic Symbols I*. in *DMS*. 2008.

23. Selvaraju, R.R., et al. *Grad-cam: Visual explanations from deep networks via gradient-based localization*. in *Proceedings of the IEEE international conference on computer vision*. 2017.

24. Ahmed, I.A., et al., *Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques.* Electronics, 2022. **11**(4): p. 530.

25. Cilia, F., et al., *Computer-aided screening of autism spectrum disorder: Eye-tracking study using data visualization and deep learning.* JMIR human factors, 2021. **8**(4): p. e27706.

26. Lakshmi Praveena, T. and N. Muthu Lakshmi. *A methodology for detecting ASD from facial images efficiently using artificial neural networks*. in *Advances in Computational and Bio-Engineering: Proceeding of the International Conference on Computational and Bio Engineering, 2019, Volume 1*. 2020. Springer.

27. Heinsfeld, A.S., et al., *Identification of autism spectrum disorder using deep learning and the ABIDE dataset.* NeuroImage: Clinical, 2018. **17**: p. 16-23.

28. Sewani, H. and R. Kashef, *An autoencoder-based deep learning classifier for efficient diagnosis of autism.* Children, 2020. **7**(10): p. 182.

29. Kong, Y., et al., *Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier.* Neurocomputing, 2019. **324**: p. 63-68.

30. Mujeeb Rahman, K. and M.M. Subashini, *Identification of autism in children using static facial features and deep neural networks.* Brain Sciences, 2022. **12**(1): p. 94.

31. Zhou, T., et al. *An automated assessment framework for speech abnormalities related to autism spectrum disorder*. in *3rd International Workshop on Affective Social Multimedia Computing (ASMMC)*. 2017.

32. Chojnicka, I. and A. Wawer, *Social language in autism spectrum disorder: A computational analysis of sentiment and linguistic abstraction.* PLoS One, 2020. **15**(3): p. e0229985.

33. Hendr, A., U. Ozgunalp, and M. Erbilek Kaya, *Diagnosis of Autism Spectrum Disorder Using Convolutional Neural Networks.* Electronics, 2023. **12**(3): p. 612.

34. 李艺, 姜杰, and 邓红静, *硬笔汉字书写特征的理解, 描述, 计算实现和应用介绍.* 电化教育研究, 2015. **36**(4): p. 62-69.

35. 庄子明, *基于深度学习的手写汉字识别与美感评分*. 2019, 北京邮电大学.

36. Fleiss, J.L., *Measuring nominal scale agreement among many raters.* Psychological bulletin, 1971. **76**(5): p. 378.

37. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* biometrics, 1977: p. 159-174.

38. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

39. Cortes, C. and V. Vapnik, *Support-vector networks.* Machine learning, 1995. **20**: p. 273-297.

40. Von Winterfeldt, D. and W. Edwards, *Decision analysis and behavioral research.* (No Title), 1986.

41. Altman, N.S., *An introduction to kernel and nearest-neighbor nonparametric regression.* The American Statistician, 1992. **46**(3): p. 175-185.

42. Hosmer Jr, D.W., S. Lemeshow, and R.X. Sturdivant, *Applied logistic regression*. Vol. 398. 2013: John Wiley & Sons.

43. Deng, J., et al. *Imagenet: A large-scale hierarchical image database*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2009.

44. Zhou, B., et al. *Learning deep features for discriminative localization*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.