Automated Detection of Children at Risk of Chinese Handwriting Difficulties Using Handwriting Process Information: An Exploratory Study

Zhiming WU†, Student Member, Tao LIN†a, and Ming LI†, Nonmembers

SUMMARY Handwriting difficulties (HWDs) in children have adverse effects on their confidence and academic progress. Detecting HWDs is the first crucial step toward clinical or teaching intervention for children with HWDs. To date, how to automatically detect HWDs is still a challenge, although digitizing tablets have provided an opportunity to automatically collect handwriting process information. Especially, to our best knowledge, there is no exploration into the potential of combining machine learning algorithms and the handwriting process information to automatically detect Chinese HWDs in children. To bridge the gap, we first conducted an experiment to collect sample data and then compared the performance of five commonly used classification algorithms (Decision tree, Support Vector Machine (SVM), Artificial Neural Network, Naïve Bayesian and k-Nearest Neighbor) in detecting HWDs. The results showed that: (1) only a small proportion (13%) of children had Chinese HWDs and each classification model on the imbalanced dataset (39 children at risk of HWDs versus 261 typical children) produced the results that were better than random guesses, indicating the possibility of using classification algorithms to detect Chinese HWDs; (2) the SVM model had the best performance in detecting Chinese HWDs among the five classification models; and (3) the performance of the SVM model, especially its sensitivity, could be significantly improved by employing the Synthetic Minority Oversampling Technique to handle the class-imbalanced data. This study gains new insights into which handwriting features are predictive of Chinese HWDs in children and proposes a method that can help the clinical and educational professionals to automatically detect children at risk of Chinese HWDs.

key words: handwriting difficulty, automated detection, machine learning, data imbalance

1. Introduction

Handwriting is a survival skill for school-aged children [1], [2]. Chinese children generally begin their journey of writing at preschool age. When they are promoted to primary school, over 50% of their time spent in school is involved in handwriting tasks such as copying and calculating [3]. Unfortunately, approximately 11% to 12% of female and 21% to 32% of male school-aged children have handwriting difficulties (HWDs) [4].

HWDs (also called dysgraphia) are defined as a disturbance or difficulty in the production of written language that is related to the mechanics of writing [5], [6]. Studies (e.g. [7], [8]) have found that HWDs have adverse effects on students’ academic success and failure to attain handwriting competency during school-age years damages students’ self-esteem. Therefore, it is necessary for clinical and educational professionals to offer appropriate intervention to children with HWDs [9]. Especially, early identification and effective intervention can prevent children with HWDs from entering into a cycle of feeling inadequacy and discouragement [10]–[14]. Detecting HWDs is the first crucial step toward intervention for children with HWDs.

Current methods of evaluating HWDs include both subjective and objective techniques (see Sect. 2 for a review). The most common methods are subjective scales and they fall into two main categories [6]: (1) global-holistic evaluations of legibility and (2) analytic evaluations that assess readability in relation to predetermined criteria. Such scales are usually based on handwriting product legibility, performance time or some specific features (e.g., letter form, size, slant and spacing, and line straightness). Subjective evaluation techniques are readily available and technically simple to implement; however, they are usually performed by trained occupational therapists and require considerable amounts of time and effort, and therefore tend to only be performed once for diagnosis [15], which may result in limited accuracy and reliability due to their inherent subjectivity [10], [16]. Another limitation of subjective evaluation techniques is their inability to use handwriting process information that gains insights into what causes HWDs and how to intervene in order to improve handwriting [15].

Recently, the wide use of digitizing tablets (where a user writes on a surface using a special cordless pen) has made it easier to record handwriting process objectively. There has been an increasing interest in using digitizing tablets to evaluate handwriting due the following reasons: (1) tablets can provide objective data on handwriting dynamics, and (2) these data are high-bandwidth (beyond what is observable to the human eye) and can also be collected rapidly and automatically [13]. Studies (e.g. [2], [16], [17]) have found some significant differences in tablet-based handwriting features between typical children (i.e. children without HWDs) and dysgraphic children, indicating the possibility of using tablet-based handwriting information to detect HWDs. For example, Rosenblum et al. [16] found that non-proficient writers required significantly more “in air” time than proficient writers. In addition, several tablet-based measuring tools (e.g. CHAS [2],
POET [18] and MEDDRAW [19]) have been developed with the aim of helping evaluators to measure and analyze the handwriting process.

Note that tablet-based handwriting evaluations have not yet been widely disseminated to clinical and educational professionals. This may be attributable to the fact that handwriting is highly complex and sometimes unstable, and evaluators may not be sufficiently adept at interpreting it [17], even with the help of the tablet-based measuring tools. There has been a need for an automated method for detecting HWDs that can lower the cost of evaluation, provide support to inexperienced therapists and enable re-evaluation throughout therapy to test for improvement [15]. To meet the need, a few studies ([15], [20]) have attempted to use classification algorithms to detect HWDs automatically. The studies have shown the possibility of using classification algorithms to automatically detect HWDs, but they are still at an early stage and have the following limitations. First, they built classifiers only on a small number of features and a wide range of features have not been systematically examined. Second, only a small proportion of children have HWDs in real life, and thus the samples for classification are class-imbalanced data; however, the performance of classification algorithms on such a real (imbalanced) data set has not been examined. Third, they identified HWDs from classification models, which are class-imbalanced data (39 children at risk of HWDs VS. 261 children without HWDs). Then five classifiers (i.e., Decision tree (DT), Support Vector Machine (SVM), Artificial Neural Network (ANN), Naïve Bayesian (NB), and k-Nearest Neighbor (KNN)) were built on the sample data to detect HWDs in children. The sensitivity, specificity, accuracy and AUC (i.e., area under receiver operating characteristic curves) of each classifier were examined and the results showed that the SVM had the best performance among the five classification models; furthermore, the performance of the SVM could be further improved by applying the Synthetic Minority Over-sampling Technique (SMOTE) [24] to addressing the class-imbalanced problem. The exploratory study showed that the SVM had good potential for the automated detection of Chinese HWDs in children and had taken the first step toward developing a computer-aid diagnostic tool for detecting Chinese HWDs.

2. Evaluating Handwriting Performance

2.1 Subjective Evaluation

Handwriting is a complex skill that incorporates various performance components such as fine motor control, precision, kinesthesia, and visual perceptual skills [1], [25]–[27]. The most common method of evaluating handwriting performance is the scale-based evaluation. The scale is subjective and its objective is to measure handwriting quality such as readability or legibility. Two techniques are often employed to measure the handwriting quality: global and analytic scales. The global scale aims to form a general judgment of a written product [28]–[30]. For example, how readable a written passage is compared with standard writing samples that have been previously scored from readable to unreadable. Global scales are readily available, inexpensive, and technically simple to implement; however, they tend to have the following drawbacks: (1) scoring requires an exhaustive and time-consuming comparison between written products and numerous pre-graded samples; and (2) the inherent subjectivity often brings their overall reliability into question.

In contrast, analytic scales use predefined specific criteria relating to legibility or readability to evaluate handwriting quality. These criteria usually include handwriting product legibility, performance time or some specific features (e.g., letter form, size, slant and spacing, and line straightness) [10], [31], [32]. A writing sample is assessed by first grading the sample according to each criterion individually and then calculating an overall score based on these grades. For example, Rosenblum [10] developed the Handwriting Proficiency Screen Questionnaire (HPSQ) to detect HWDs among school-aged children according to three major handwriting characteristics (i.e., legibility, time and speed of performance, and physical and emotional well-being). Analytic scales are less subjective than global scales, but they still require prolonged processing time, and lack the holistic perspective found in global scales. In addition, it is difficult to directly compare the results from different analytic scales. Note that both the global and analytic scales relate to the written output and not the process of handwriting performance, although the analytic scale can yield valuable information about characteristics of writers’ handwriting. However, studies have emphasized that getting insights into handwriting process is necessary when evaluating handwriting performance because it can provide substantive information on motor control mechanisms and the underlying pathology HWDs [9], [13], [33], [34].

2.2 Tablet-Based Evaluation

The advances in tablet technologies have made it easier to record handwriting process objectively. There has been a growing interest in exploring the possibility of using digitizing tablets to detect HWDs because tablet-based
Machine learning is a fast-growing discipline which provides such a method of automatically learning to recognize complex patterns from data [41]. A few efforts [15], [20] have been made to develop automated methods for detecting HWDs using classification algorithms and showed promising results. For example, a significant study by Richardson et al. [15] presented COACH (a Cumulative Online Algorithm for Classification of Handwriting deficiencies) based on the writing features such as pen’s tilt, pressure and in air time and obtained overall classification accuracy of 75%; a recent study [20] also applied SVMs to detecting HWDs in children based on the triangle drawing strategy and obtained the accuracy and sensitivity of 63.48% and 68.57%, respectively. The studies [15], [20] have showed the potential of classification algorithms in detecting HWDs, but they are still at an early stage and have the following limitations. First, they built their classifiers only on a few handwriting features, while we argue that a wide range of features should be investigated to obtain better performance because even when a few features are dominant for detecting HWDs there is information hidden in other features [15]. Second, a small proportion of children have HWDs in real life and such an imbalanced data distribution may jeopardize the classification algorithms [8]. However, previous studies built and evaluated their classifiers on balanced data sets rather than on real (imbalanced) data sets; therefore, the capability of these classifiers in real environments may still be questioned. Moreover, previous automated detection of HWDs extracted handwriting features only from alphabetic language writing [15] or simple shapes drawing [20] rather than Chinese character writing. Each written language has its unique characteristics and formats. Alphabetic languages such as English emphasize smoothness and continuity in their written forms [16], whereas Chinese characters contain sharp turns of stroke, demand frequent pen lifts [21], and involve complex geometric figuration and stroke arrangement within a squared-area [22]. As such, writing Chinese characters may have more complex handwriting patterns than English, which may cause more difficulties for HWD detection [42]. Studies [2], [6], [9] have suggested that the methods assessing English handwriting may not be applicable to Chinese due to the differences between English and Chinese languages. To our best knowledge, the efficacy of classification algorithms in detecting Chinese HWDs in children has not been explored.

3. Experiment

3.1 Participants and Tasks

A total of 300 children (aged 6-7) were recruited from several primary schools. The children have no other medical and physical disabilities that might interfere with the handwriting performance. Informed consent of parents was obtained for all children.

The HPSQ [10] was chosen to identify HWDs in children. [33] The HPSQ is a 10-item questionnaire that was developed to detect school-aged children with HWDs. The ten items are carefully worded so that they can be clear and understandable to teachers and can be directly answerable from their observations of children. These items were scored on a 5-point Likert scale, ranging from 0 to 4 (higher scores indicate poorer handwriting performance). The final questionnaire score was calculated by summing the scores of all the 10 items. The children with a score of 14 and up were considered to be at risk of HWDs, while children with a total score of 0 to 13 were considered typical children (i.e. children without HWDs).

The HPSQ was completed by teachers who received training in identifying HWDs by the HPSQ. A total of 18 teachers were invited to evaluate the handwriting performance of their students. All the teachers have taught the participants (i.e., their students) for at least half a years and were responsible for the subjects involving handwriting (e.g., language, mathematics). The teachers were required to observe the handwriting performance of students in their class for two weeks after receiving the training.

In the experiment, a template consisting of 49 simplified Chinese characters was displayed on the computer screen in seven columns of seven characters (see Fig. 1). The 49 characters were selected from the first-grade textbooks of primary schools in China. They covered all
Fig. 1 The template consisting of 49 simplified Chinese characters.

The six main basic structures of Chinese characters (i.e., independent, left-right, above-bottom, above-middle-bottom, left-middle-right and inside-outside) and most basic stroke units [43] to ensure the representativeness of characteristics of the selected Chinese characters. The display sequences of the columns were randomized.

A sheet of paper with a 7 × 7 grid was affixed to the surface of a digital tablet (WACOM DTZ-1200W tablet) and each grid had a size of 10mm × 10mm. Children were requested to write each character inside the grid accordingly as quickly as possible while keeping the characters readable and legible. Children were not allowed to correct their writing if they wrongly wrote the character. The writing process of the children was also recorded by cameras for later observation.

The children’s teachers were asked to complete the HPSQ for their students. The teachers have taught the subjects involving handwriting (e.g., language, mathematics) for at least half a year, and they have also received training in identifying HWDs by the HPSQ. Before the experiment, the teachers were first required to consciously observe handwriting performance of their students in their class for two weeks; then, they were asked to complete the HPSQ based on the careful observation of the recorded writing process in the experiment, along with the children’s daily handwriting performance. To minimize the subjectivity of the teachers, three teachers were asked to evaluate the same child and complete the HPSQ questionnaires, and the children were included in the HWD group only if they obtained a total score of 14 and up from two or more teachers. We denote the handwriting performance of the students as a label (i.e., HWD or typical). Table 1 summarizes the age, gender, and mean HPSQ scores for the children in the HWD and typical groups, respectively.

### Table 1: Age, gender and HPSQ score of participating children.

|                       | HWD Group | Typical Group |
|-----------------------|-----------|---------------|
| Number of children    | 79        | 261           |
| Mean Age (months)     | 74.6      | 75.1          |
| Gender (girls vs. boys)| 17 vs. 22 | 130 vs. 131   |
| Mean HPSQ score       | 15.92     | 8.25          |

#### 3.2 Procedure

The experimental tasks were performed in a quiet room at school. The experiment was divided into three phases: a welcome phase, a practice phase and a task phase. During the welcome phase, the children’s parents were requested to sign a consent form with a detailed description of the experiment, its duration, and its research purpose and fill out a background questionnaire about their children. At the outset of the practice phase, instructions were read to each child describing the task rules, and each participant was given a brief tutorial on how to complete the task. Children were then allowed to practice for about 5 minutes to ensure that they were familiar with the operations of the handwriting device. During the task phase, each child was required to copy the 49 Chinese characters from the template. After all the children finished the copying task, the teachers completed the HPSQ for their students (i.e. the children).

### 4. Method for Detecting HWDs

Classification is the most common type of supervised learning method and classification models (classifiers) were employed to detect HWDs in our study. A classifier is a mapping of a sample to one of the predefined classes or groups (e.g., HWD or typical in the case). To establish the classifier, the following main steps were completed: (1) for each sample, 34 features (see Table 1) were extracted from handwriting process data as input of the classifiers and a classification label (i.e., HWD or typical) was obtained based on the HPSQ as output of the classifier; and (2) the algorithms of classification learning and feature subset selection were developed to establish the mapping of the features to the labels. A total of 300 samples were collected in the experiment.

#### 4.1 Feature Extraction

Handwriting process was recorded using a WACOM DTZ-1200W tablet at 142 Hz and was represented by a series of sample points along the pen trace with time-stamps. Each sample point included the coordinate positions (x, y), pressure (0 ˜ 1024 levels) of the pen-tip (p), and the altitude and azimuth of the pen (θ, φ), as shown in Fig. 2. These collected sample data were used to extract handwriting features.

For each writing task, a total of 34 handwriting features were extracted from the data recorded during the experiment, which can be generally divided into five categories:
spatial, temporal, velocity, acceleration and pen tilt. Table 2 details the full feature sets. Each feature was linearly scaled to the range [0, 1] by the max-min scaling technique before constructing classifiers [44].

4.2 Feature Subset Selection

Feature subset selection is important for a classifier because the procedure can greatly affect its classification accuracy, generalization performance, and computational efficiency [45]. A wrapper-based feature selection approach was employed to select features because it can be “wrapped around” any existing learning algorithms [45]. The wrapper-based approach is typically computationally expensive, but tends to achieve good results. In this study, a feature selection wrapper employing sequential forward selection (SFS) [44] was used with learning algorithms to select feature subsets. SFS is a bottom-up search procedure where one feature is added at a time to the current feature set. The feature to be added is selected from the subset of the remaining features so that the new feature set generates the maximum value of the performance function over all candidate features for addition.

4.3 Classification Algorithms

Five machine learning algorithms (i.e., Decision Tree (DT), Support Vector Machine (SVM), Artificial Neural Network (ANN), Naïve Bayesian (NB) and k-Nearest Neighbor (KNN)) were used to build classifiers and detect children at risk of HWDs. These techniques were selected primarily because they have been successfully used across a variety of applications and have shown good performance even on nonlinear problems. We adopted the most commonly used parameter settings when training and testing classifiers. The parameter settings were chosen based on experience and knowledge about the problem at hand rather than through exhaustive testing of all possible combinations of parameter

| Categories | Descriptions                                                                 | Abbreviation              |
|------------|------------------------------------------------------------------------------|---------------------------|
| Temporal   | Total task time                                                              | TOTAL-TIME                |
|            | Time when pen is moving on the tablet (i.e., “on paper” time)                | ON-PAPER-TIME             |
|            | Time when pen is moving above the tablet (i.e., “in air” time)               | IN-AIR-TIME               |
|            | Ratio of IN-AIR-TIME to ON-PAPER-TIME                                        | AIR/PAPER-TIME            |
| Velocity   | Average writing velocity                                                     | AVG-VEL                   |
|            | Standard deviation of writing velocity                                        | STD-VEL                   |
|            | Maximum writing velocity                                                     | MAX-VEL                   |
|            | Average writing velocity measured on the x-axis                             | AVG-X-V                   |
|            | Standard deviation of writing velocity measured on the x-axis                | STD-X-V                   |
|            | Maximal writing velocity measured on the x-axis                             | MAX-X-V                   |
|            | Average writing velocity measured on the y-axis                             | AVG-Y-V                   |
|            | Standard deviation of writing velocity measured on the y-axis                | STD-Y-V                   |
|            | Maximal writing velocity measured on the y-axis                             | MAX-Y-V                   |
|            | Average pen angular velocity                                                | AVG-A-V                   |
|            | Standard deviation of pen angular velocity                                   | STD-A-V                   |
| Acceleration| Average writing acceleration                                                | AVG-ACC                   |
|            | Standard deviation of writing acceleration                                  | STD-ACC                   |
|            | Maximal writing acceleration                                                | MAX-ACC                   |
|            | Average writing acceleration measured on the x-axis                         | AVG-X-A                   |
|            | Standard deviation of writing acceleration measured on the x-axis           | STD-X-A                   |
|            | Maximal writing acceleration measured on the x-axis                         | MAX-X-A                   |
|            | Average writing acceleration measured on the y-axis                         | AVG-Y-A                   |
|            | Standard deviation of writing acceleration measured on the y-axis           | STD-Y-A                   |
|            | Maximal writing acceleration measured on the y-axis                         | MAX-Y-A                   |
| Pressure   | Average pen pressure against the tablet                                     | AVG-PRES                  |
|            | Standard deviation of pen pressure against the tablet                       | STD-PRES                  |
|            | Maximal pressure of the pen on the tablet                                   | MAX-PRES                  |
| Pen tilt   | Average pen altitude                                                         | AVG-ALT                   |
|            | Standard deviation of pen altitude                                           | STD-ALT                   |
|            | Average pen azimuth                                                         | AVG-AZI                   |
|            | Standard deviation of pen azimuth                                            | STD-AZI                   |
| Spatial    | Total distance traveled by the pen on the tablet                            | PEN-DIS                   |
|            | Pen travel distance measured on the x-axis                                  | PEN-X-D                   |
|            | Pen travel distance measured on the y-axis                                  | PEN-Y-D                   |

Fig. 2 Handwriting information from tablet and pen.
settings.

The Radial Basis Function was used as kernel functions for SVMs. For ANN, a three-layer back-propagation neural network (BPNN) was adopted using the combination of sigmoid function as the activation function and back-propagation as the learning rule. The ANN was trained using the learning rate of 0.3. In addition, the initial connection weights took values from −1 to 1 and all input values were normalized into [0, 1] before they entered into the networks. The number of nodes in the hidden layer was set to half of the total number of nodes in the input and output layers. For KNN, we set k to be three and used the Euclidean distance as the distance metric.

4.4 Evaluation Method

Performance of the classifiers was evaluated with stratified three-fold cross-validation [42]. Three-fold cross-validation was chosen to strike a balance between the number of folds and the number of data points in each fold. Specifically, accuracy, sensitivity, and specificity measures were adopted to evaluate each classifier (see Eqs. (1)–(3)).

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (1)
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \quad (2)
\]

\[
\text{Specificity} = \frac{TN}{FP + TN} \times 100\% \quad (3)
\]

In Eqs. (1)–(3), the true positives (TP) and true negatives (TN) are the number of the samples that are correctly classified as positive and negative, respectively. The false negatives (FN) refer to the number of the positive samples that were mislabeled as negative, and the false positives (FP) refer to the number of the negative samples that were incorrectly labeled as positive [44]. In the study, the positive samples (P) are the children at risk of HWDs while the negative samples (N) are the children without HWD. Sensitivity measures the proportion of children with HWDs that are correctly identified as having HWDs; specificity measures the proportion of children without HWDs that are correctly identified as typical; accuracy measures the proportion of children is correctly classified. The long-term goal of our study is to develop an auxiliary diagnostic tool for HWDs that can initially detect Chinese children at risk of HWDs as far as possible before the doctor’s official diagnosis is made. Therefore, the sensitivity is the biggest concern, since the low sensitivity means that the children with HWDs who is incorrectly identified as the ones without HWDs may lose the opportunity to receive doctors’ diagnosis. That it, the sensitivity is the most important index for our classifiers.

The Receiver Operating Characteristic (ROC) curve displays the tradeoff between true positive rate (i.e. sensitivity) and false positive rate (i.e. 1 - specificity) of a binary classifier [44]. We adopted the area under the ROC curves (AUC) to evaluate which classifier is better on average in our study. The AUC is a measure of classifier performance on average. If all the test samples are ranked in the descending order of their predicted probability values (larger values for children with HWDs), AUC represents the probability that a randomly chosen instance from the HWD group ranks above a randomly chosen instance from the typical group.

5. Results and Discussion

Five classifiers were first built on the raw (imbalanced) data to detect Chinese HWDs in children using the five machine learning algorithms (i.e., DT, SVM, ANN, NB and KNN). The SVM classifier showed the best sensitivity and average performance among the five classifiers. To further improve the performance of the SVM classifier, the SMOTE [24] was introduced into the SVM algorithm (SVM$_{SMOTE}$) to address the class-imbalanced problem in raw data.

5.1 Performance Comparisons between Common Classifiers

Table 3 summaries the accuracy, sensitivity and specificity for each classifier on the raw (imbalanced) data set. The bold entries highlight the best results obtained for each measure. The results in Table 3 showed that: (1) each classifier performed better than random guesses (higher than 50%) on all the three measures; (2) the ANN and SVM classifiers generated the best overall classification accuracy (75%) and their accuracy was better than that of the DT, NB and KNN; and (3) the SVM classifier produced the best sensitivity (69.1%), while the ANN classifier obtained the best specificity (77%). Note that the SVM and ANN classifiers had equal accuracy, but sensitivity of the SVM classifier was 5% higher than that of the ANN classifier; further, a paired t-test showed that the sensitivity difference between the SVM and ANN classifiers was significant ($t_2 = 6.047$, $p = 0.026$), suggesting that the SVM classifier could recognize the children at risk of HWDs more accurately in comparison with the ANN classifier. These results showed that children at risk of HWDs had different handwriting patterns from typical children, which could be differentiated by classification algorithms.

Figure 3 showed the AUCs of all classifiers. The results in Fig. 3 showed that all the five classifiers produced the results that were better than random guesses (AUC > 0.50), and that the SVM classifier had the best average performance (0.836) among the five classifiers, followed by ANN (0.812), DT (0.597), NB (0.559) and KNN (0.538).

| Classifier | Accuracy | Sensitivity | Specificity |
|------------|----------|-------------|-------------|
| DT         | 0.627    | 0.615       | 0.628       |
| SVM        | 0.750    | 0.691       | 0.751       |
| ANN        | 0.750    | 0.641       | 0.770       |
| NB         | 0.573    | 0.561       | 0.583       |
| KNN        | 0.566    | 0.559       | 0.563       |
| Mean       | 0.652    | 0.609       | 0.659       |
5.2 Performance of SVM Combined with SMOTE

The SVM\_SMOTE model could detect Chinese HWDs with accuracy of 77.0% and sensitivity of 76.9%. The accuracy and sensitivity of the SVM\_SMOTE model were better than that of the study [20] that developed SVM models on drawing strategy features (accuracy = 63.48%; sensitivity = 68.57%). In addition, the accuracy of the SVM\_SMOTE model was similar to that of the study [15] focusing on HWDs of alphabetic languages (accuracy = 75%) [32]. The study [15] did not report the sensitivity of its models. The performance of SVM\_SMOTE is considered to be good in this domain, as diagnosing handwriting deficiency is not an exact science and perfect classification is not expected [15]. For example, the labeling of the sample data may be noisy, which in turn affects our ability to classify the sample correctly.

Figure 4 reports the performance differences between the SVM and SVM\_SMOTE classifiers. The SVM\_SMOTE classifier produced better results for all performance measures than the SVM classifier. Specifically, the SVM\_SMOTE classifier brought about an increase of 17.8%, 2.1%, 2% and 2% in sensitivity, AUC, accuracy and specificity, respectively; furthermore, paired t-tests suggested that the improvement in each measure was significant (sensitivity: \( t_2 = 7.011, p < 0.01 \); AUC: \( t_2 = 6.321, p < 0.05 \); accuracy: \( t_2 = 5.071, p < 0.05 \); specificity: \( t_2 = 5.355, p < 0.05 \)). The results showed that balancing data was helpful for improving the performance of SVMs and this benefit might be more pronounced for the sensitivity measure.

SVM\_SMOTE obtained the best performance in our study and we were also interested in the relative strength of each feature in constructing the SVM\_SMOTE model. An analysis of information gain [44] was used to rank individual feature according to its “information content”, which can represent an objective estimation of how valuable the feature may be in constructing the classifier. Table 4 lists the selected features for SVM\_SMOTE and they are ranked in descending order of importance obtained by information gain analyses.

As Table 4 shows, the selected features were from multiple sources of handwriting information, including temporal, velocity, acceleration, pressure and pen tilt, suggesting that handwriting information of different types could make unique contributions to detecting HWDs. Thus, it is necessary to extract a wide variety of handwriting features to detect Chinese HWDs in children. Note that no distance-related features were included in the SVM\_SMOTE model. The result showed that distance traveled by pen-tip on the tablet might provide less or no information when SVM\_SMOTE was used to detect HWDs. More interestingly, the previous study [15] showed that pressure was the main attribute that contributes to detecting HWDs when an alphabetic language was written, while the AIR/PAPER-TIME feature was ranked first in our study (see Table 4) and was the best discriminator of detecting Chinese HWDs. This may be attributed to the fact that Chinese handwriters engage in greater “in air” activity than do children who write in other languages [16].

### Table 4

| Rank | Features               |
|------|------------------------|
| 1    | AIR/PAPER-TIME         |
| 2    | STD-PRES               |
| 3    | PAUSE-100              |
| 4    | STD-VEL                |
| 5    | AVG-Y-A                |
| 6    | STD-AZI                |
| 7    | TOTAL-TIME             |

6. Conclusions and Future Work

HWDs have been considered to be a severe hindrance to academic learning for school-aged children. Detecting HWDs is the first crucial step toward clinical or teaching intervention for the children with HWDs. However, how to automatically detect HWDs in children remains a great challenge. To our knowledge, the study is the first exploration into the potential of developing classification models on the handwriting process data to detect Chinese HWDs in children. The current study shows that each classifier can obtain the results that are better than random guesses, suggesting the possibility of developing classifiers on handwriting process data to detect HWDs automatically. In addition, SVMs seem have the greatest potential for differentiating handwriting patterns between children at risk of HWDs and typical children compared with other four classifiers (ANN, DT, KNN, and NB). Considering that only a small proportion...
of children (13% in this case) have HWDs in real life, the SMOTE was used to improve the performance of SVMs by balancing the raw imbalanced data. The SVM developed on the balanced data produced by SMOTE shows significant improvement compared with the SVMs developed on the imbalanced data. More importantly, this benefit is pronounced for the sensitivity measure that tends to be of interest to evaluators. The result suggests that class-imbalanced problems should be addressed to obtain good performance when classifiers are employed to detect HWDs in children. Our study suggests that different types of handwriting features should be extracted when classifiers are developed to detect Chinese HWDs in children, since each type may provide a unique source for indicating HWDs. In addition, the analyses in our study offer valuable information in HWD diagnosis and remediation for Chinese children.

In sum, the initial study provides promising results and this would motivate further research that focuses on developing models with better performance. First, more samples should be collected to establish a more accurate model and to help to determine the implications of these results for occupational therapy intervention. Second, the performance of models could be improved by fine-tuning their parameters (e.g. the penalty parameter and kernel parameter of SVMs). Third, a variety of methods of extracting and selecting handwriting features will be studied to improve classification accuracy. Especially, considering that the low sensitivity of classifiers may cause some children with HWDs to lose the opportunity to receive remediate training, the sensitivity of classifiers should be greatly improved in the next step.

Acknowledgements

This study was supported by Science and Technology Planning Program, Sichuan University and Luzhou [2017CDLZ-G30], and Science and Technology Innovation Seeding Project, Sichuan Province, China [No. 17-YCG053].

References

[1] S.S.T. Lam, R.K.C. Au, H.W.H. Leung, and C.W.P. Li-Tsang, “Chinese handwriting performance of primary school children with dyslexia,” Res. Dev. Disabilities, vol.32, no.5, pp.1745–1756, Sept. 2011.

[2] C.W.P. Li-Tsang, A.S.K. Wong, H.W.H. Leung, J.S. Cheng, B.H.W. Chiu, L.F.L. Tse, and R.C.K. Chung, “Validation of the Chinese handwriting analysis system (chas) for primary school students in hong kong,” Res. Dev. Disabilities, vol.34, no.9, pp.2872–2883, 2013.

[3] M.H. Tseng and S.M.K. Chow, “Perceptual-motor function of school-age children with slow handwriting speed,” Am. J. Occup. Ther. Official Publ. Am. Occup. Ther. Assoc., vol.54, no.1, pp.83–88, 2000.

[4] R. Karlströdt and T. Stefansson, “Problems in developing functional handwriting,” Percept. Mot. Skills, vol.94, no.2, pp.623–662, 2002.

[5] L. Hamstrabeltz and A.W. Blöte, “A longitudinal study on dysgraphic handwriting in primary school,” J. Learn. Disabilities, vol.26, no.10, pp.689–699, 1993.

[6] S. Rosenblum, P. Weiss, and S. Parush, “Product and process evaluation of handwriting difficulties,” Educ. Psychol. Rev., vol.15, no.1, pp.41–81, March 2003. DOI:10.1023/A:1021371425220

[7] N. Reithinger, G. Herzog, and A. Ndahi, “Situated multimodal interaction in smartkorn,” Comput. Graph., vol.27, no.6, pp.899–903, 2003. DOI:10.1016/j.cag.2003.08.004

[8] A.D. Sandler, T.E. Watson, M. Fooo, M.D. Levine, W.L. Coleman, and S.R. Hooper, “Neurodevelopmental study of writing disorders in middle childhood,” J. Dev. Behav. Pediatrics Jdbp, vol.13, no.1, pp.17–23, 1992.

[9] M.J.M. Volman, B.M. van Schendel, and M.J. Jongmans, “Handwriting difficulties in primary school children: A search for underlying mechanisms,” Am. J. Occup. Ther., vol.60, no.4, pp.451–460, 2006.

[10] S. Rosenblum, “Development, reliability, and validity of the handwriting proficiency screening questionnaire (HPSQ),” Am. J. Occup. Ther., vol.62, no.3, pp.298–307, 2008.

[11] S. Rosenblum and L. Gafni-Lachter, “Handwriting proficiency screening questionnaire for children (HPSQ-C): development, reliability, and validity,” Am. J. Occup. Ther., vol.69, no.3, pp.1–9, 2015.

[12] S. Graham, K.R. Harris, and B. Fink, “Is handwriting causally related to learning to write? Treatment of handwriting problems in beginning writers,” J. Educ. Psychol., vol.92, no.4, pp.620–633, 2000.

[13] S. Rosenblum, P.L. Weiss, and S. Parush, “Handwriting evaluation for developmental dysgraphia: Process versus product,” Read. Writ., vol.17, no.5, pp.433–458, July 2004. DOI:10.1023/B:READ.0000044596.91833.55

[14] V.W. Berninger and D. Amtmann, “Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice,” Handbook of Learning Disabilities, ed. K.R. Harris, H.L. Swanson, S. Graham, pp.345–363, Guilford, New York, 2003.

[15] A. Richardson, S. Kraus, P.L. Weiss, and S. Rosenblum, “Coach: Cumulative online algorithm for classification of handwriting deficiencies,” Proc. 20th Natl. Conf. Innov. Appl. Artif. Intell., Chicago, Illinois, pp.F1725–1730, 2008.

[16] S. Rosenblum, S. Parush, and P.L. Weiss, “Computerized temporal handwriting characteristics of proficient and non-proficient hand-writers,” Am. J. Occup. Ther., vol.57, no.2, pp.129–138, 2003.

[17] S. Rosenblum, A.Y. Dvorkin, and P.L. Weiss, “Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting,” Hum. Mov. Sci., vol.25, no.4-5, pp.608–621, Oct. 2006. DOI:10.1016/j.humov.2006.07.005

[18] K.W. Poon, C.W.P. Li-Tsang, T.P.L. Weiss, and S. Rosenblum, “The effect of a computerized visual perception and visual-motor integration training program on improving Chinese handwriting of children with handwriting difficulties,” Res. Dev. Disabilities, vol.31, no.6, pp.1552–1560, Nov. 2010. DOI:10.1016/j.ridd.2010.06.001

[19] M.C. Fairhurst, T. Linell, S. Glienâ†œt, R.M. Guest, L. Heutte, and T. Paquet, “Developing a generic approach to online automated analysis of writing and drawing tests in clinical patient profiling,” Behav. Res., vol.40, no.1, pp.290–303, Feb. 2008. DOI:10.3758/BRM.40.1.290

[20] N. Tabatabaey-Mashadi, R. Sudirman, R.M. Guest, and P.I. Khalid, “Analyses of pupils’ polygonal shape drawing strategy with respect to handwriting performance,” Pattern Anal. Appl., vol.18, no.3, pp.571–586, Aug. 2015. DOI:10.1007/s10044-014-0423-5

[21] M.H. Tseng, “Development of pencil grip position in preschool children,” Occup. Ther. J., vol.18, no.4, pp.207–224, 1998.

[22] S.M. Chow, S.W. Choy, and S.K. Mui, “Assessing handwriting speed of children biliterate in English and Chinese,” Percept. Mot. Skills, vol.96, no.2, pp.685–694, 2003.

[23] H.M. Lee, C.C. Sheu, and J.M. Chen, “Handwritten Chinese character recognition based on primitive and fuzzy features via the SEART neural net model,” Appl. Intell., vol.8, no.3, pp.269–285, 1998. DOI:10.1023/a:100827518100

[24] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, “Smote: Synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol.16, no.1, pp.321–357, 2001.
Zhiming Wu is a postgraduate in Sichuan University, graduated from East China Normal University with a degree of bachelor in Software Engineering, and is studying Human-Computer Interaction to complete a doctor degree of Computer Science in Sichuan University.

Tao Lin received the MS degree in computer science from Sichuan University, China, in 2003, and the PhD degree in information science from University of Yamanashi, Japan, in 2007. He also worked at Waseda University, Japan, as a visiting lecturer. Currently, He is a professor at Sichuan University, China.

Ming Li received the bachelor of engineering degree in software engineering from the Southwest University of Science and Technology, and is now studying Human-Computer Interaction to complete a master degree of Computer Science in Sichuan University.