DYPA: A Machine Learning Dyslexia Prescreening Mobile Application for Chinese Children

SHUHAN ZHONG*, The Hong Kong University of Science and Technology, China
SIZHE SONG*, The Hong Kong University of Science and Technology, China
TIANHAO TANG, The Hong Kong University of Science and Technology, China
FEI NIE, The Hong Kong University of Science and Technology, China
XINRUI ZHOU, The Hong Kong University of Science and Technology, China
YANKUN ZHAO, The Hong Kong University of Science and Technology, China
YIZHE ZHAO, The Hong Kong University of Science and Technology, China
KUEN FUNG SIN, The Education University of Hong Kong, China
S.-H. GARY CHAN, The Hong Kong University of Science and Technology, China

Identifying early a person with dyslexia, a learning disorder with reading and writing, is critical for effective treatment. As accredited specialists for clinical diagnosis of dyslexia are costly and undersupplied, we research and develop a computer-assisted approach to efficiently prescreen dyslexic Chinese children so that timely resources can be channelled to those at higher risk. Previous works in this area are mostly for English and other alphabetic languages, tailored narrowly for the reading disorder, or require costly specialized equipment. To overcome that, we present DYPA, a novel **DY**slexia **P**rescreening mobile **A**pplication for Chinese children. DYPA collects multimodal data from children through a set of specially designed interactive reading and writing tests in Chinese, and comprehensively analyzes their cognitive-linguistic skills with machine learning. To better account for the dyslexia-associated features in handwritten characters, DYPA employs a deep learning based multilevel Chinese handwriting analysis framework to extract features across the stroke, radical and character levels. We have implemented and installed DYPA in tablets, and our extensive trials with more than 200 pupils in Hong Kong validate its high predictive accuracy (81.14%), sensitivity (74.27%) and specificity (82.71%).

CCS Concepts: • **Human-centered computing** → *Tablet computers*; • **Social and professional topics** → **People with disabilities**.

Additional Key Words and Phrases: Dyslexia Prescreening, Dyslexia in Chinese, Handwritten Chinese Character Analysis, Learning Disorder, Machine Learning and Deep Learning

*Both authors contributed equally to this research.

Authors' addresses: Shuhan Zhong, The Hong Kong University of Science and Technology, Hong Kong SAR, China, szhongaj@cse.ust.hk; Sizhe Song, The Hong Kong University of Science and Technology, Hong Kong SAR, China, ssongad@cse.ust.hk; Tianhao Tang, The Hong Kong University of Science and Technology, Hong Kong SAR, China, ttangae@cse.ust.hk; Fei Nie, The Hong Kong University of Science and Technology, Hong Kong SAR, China, fnie@connect.ust.hk; Xinrui Zhou, The Hong Kong University of Science and Technology, Hong Kong SAR, China, xzhouaz@connect.ust.hk; Yankun Zhao, The Hong Kong University of Science and Technology, Hong Kong SAR, China, yzhaock@connect.ust.hk; Yizhe Zhao, The Hong Kong University of Science and Technology, Hong Kong SAR, China, yzhaocj@connect.ust.hk; Kuen Fung Sin, The Education University of Hong Kong, Hong Kong SAR, China, kfsin@eduhk.hk; S.-H. Gary Chan, The Hong Kong University of Science and Technology, Hong Kong SAR, China, gchan@cse.ust.hk.

2474-9567/2023/9-ART143 $15.00
https://doi.org/10.1145/3610908

ACM Reference Format:
Shuhan Zhong, Sizhe Song, Tianhao Tang, Fei Nie, Xinrui Zhou, Yankun Zhao, Yizhe Zhao, Kuen Fung Sin, and S.-H. Gary Chan. 2023. DYPA: A Machine Learning Dyslexia Prescreening Mobile Application for Chinese Children. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 143 (September 2023), 21 pages. https://doi.org/10.1145/3610908

1 INTRODUCTION

Dyslexia is a learning disorder in reading and writing. Neurobiological in origin, it is mainly caused by inherent abnormalities in the structure and function of the brain [17, 36]. Children with dyslexia have difficulties in acquiring cognitive-linguistic skills including phonological sensitivity, morphological/semantic awareness, orthographical processing, and fluency [20, 34]. Moreover, the associated symptoms are largely language-based [10, 32, 37]. Similar to other nationalities, dyslexia happens in Chinese as well. The reported prevalence of dyslexia in Chinese school-aged children is 5.4% in the Chinese mainland, 9.7% in Hong Kong, and 7.5% in Taiwan [30], totaling up to more than 10 million children.

Without early intervention, dyslexic children are more likely to suffer from school failure, which may lead to their other social and psychological problems [4]. Early educational therapy can greatly relieve the negative impacts of dyslexia. Timely identification, therefore, is critical for young children with dyslexia [49, 54]. However, this can be challenging for both parents and teachers because the signs and symptoms of dyslexia are not obvious. On the other hand, clinical diagnosis of dyslexia usually involves a battery of paper-based standardized tests, with accredited professionals manually scoring the tests to make the final decision [2, 8, 46]. As in general the tests are lengthy and the services of accredited specialists are costly and undersupplied, having timely and cost-effective assessments is difficult. As a result, a large proportion of dyslexic children would miss the golden period for diagnosis and intervention [6, 52].

In this paper, we research and develop a computer-assisted approach to prescreen dyslexic Chinese children in a cost-effective way, such that professional resources for advising and treatment can be better utilized and channeled to those at higher risk in a timely manner. Our work is challenging because prior arts in this area are mostly designed for English and other alphabetic languages, where the characters reflect the phonetic components of sound [38, 39]. These approaches follow theoretical research of dyslexia in alphabetic languages to focus more on the learning deficits in terms of reading. On the contrary, Chinese is a morphosyllabic language whose characters represent a convergence of both sound and meaning. Moreover, unlike the linear writing of letter strings in alphabetic languages, Chinese characters are hierarchically organized according to certain linguistic regularities by multiple levels of units such as strokes, radicals and subcharacters. It thus requires a combination of multiple cognitive-linguistic skills to learn both the reading and the writing of Chinese (Figure 1). However, existing works on alphabetic languages fail to consider such characteristics and are not applicable to Chinese.

Due to such uniqueness in Chinese, Chinese children with dyslexia have been shown to exhibit learning deficits in both reading and writing, and writing-related features can help identify children with dyslexia [28, 33]. To investigate writing-related features in Chinese for dyslexia, some studies use specialized graphics panels to collect and analyze handwriting data [28]. However, the systems are costly and not so portable, and hence not so accessible, scalable and cost-effective. On the other hand, existing algorithms for Chinese handwriting analysis are mainly proposed for recognition [56]. They rely on holistic features at the character level to map the input to a most likely standard character. Others focus on stroke level errors and penmanship instead for education purposes [22, 45]. These algorithms cannot jointly consider the features across stroke, radical and character levels in children's handwritings, and how to capture the multilevel features from the handwritten characters for dyslexia prediction remains challenging.

Overcoming the challenges, we present DYPA, a **dy**slexia **p**rescreening mobile **a**pplication for Chinese children. We show in Figure 2 the overview of DYPA, which can be installed and run on general consumer-grade tablets equipped with touch screens such that no specialized equipment is required, hence accessible to the public at

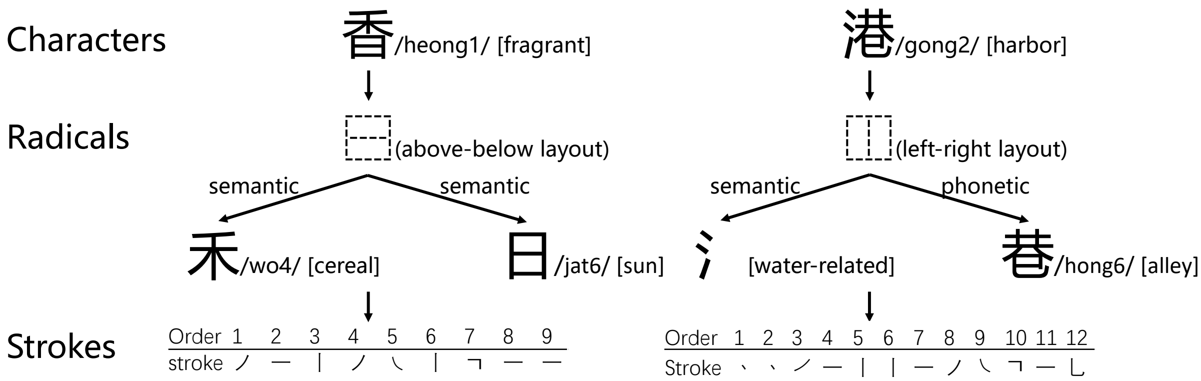


Fig. 1. An example of the Chinese morphosyllabic writing system (香港 is the Chinese word for "Hong Kong"). 香 is composed of two semantic radicals in the above-below layout, with 9 strokes in total. 港 is composed of a semantic radical and a phonetic radical in the left-right layout, with 12 strokes in total.

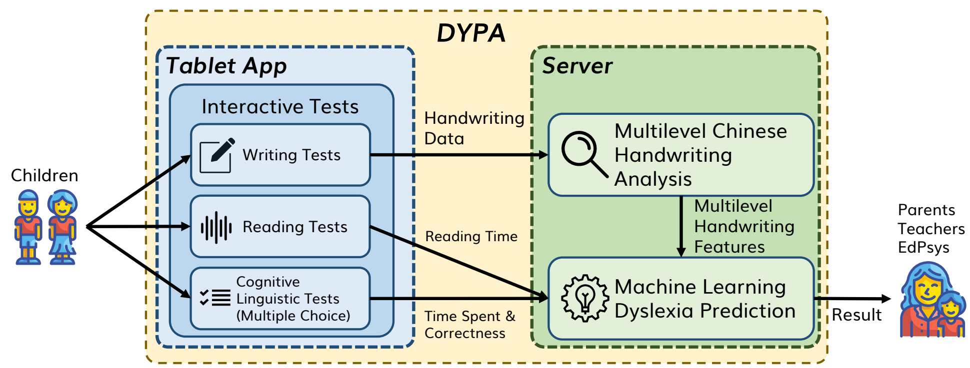


Fig. 2. Overview of DYPA workflow.

low cost. DYPA is a joint effort marrying the expertise of our educational psychologists (EdPsys) and computer scientists. In DYPA, our EdPsys design a series of cartoonized interactive tests that are engaging and friendly for children. The test designs are based on theoretical research of Chinese dyslexia to comprehensively examine associated learning disabilities in terms of both reading and writing. Multimodal data are collected when students interact with DYPA, and then sent to our server to be processed by machine learning and deep learning algorithms designed by our computer scientists to predict the risk of dyslexia. The prescreening process of DYPA is fully automated by the algorithms that no expertise from the specialist is required, and the result of DYPA can be used by parents and teachers for advice on further clinical diagnosis and specialized educational therapy if needed. It can also help relieve the workload of EdPsys for clinical diagnosis.

In DYPA, to examine children's cognitive-linguistic skills through writing, we design different writing tests that require students to write Chinese characters on the touch screen. Furthermore, to better account for the multilevel characteristics in the analysis of children's handwritings, we propose a novel deep learning based multilevel Chinese handwriting analysis framework in DYPA that extracts features across the stroke, radical, and character levels for dyslexia prediction. Specifically, we extract the stroke level features by stroke matching. Considering the tree-structured nature of Chinese characters at the radical level, we propose to extract the radical level features by using a deep image-to-tree model TSDNet [58]. And we extract the character level feature by using a deep convolutional neural network (CNN) model. Fusing the features from the reading, writing, and multiple choice tests, DYPA employs a gradient boosting machine [14] as the machine learning classifier to predict the risk of dyslexia.

We have implemented and installed DYPA on tablets and conducted extensive experiments to evaluate its prescreening performance with more than 200 primary school pupils of six to eight years old with and without dyslexia in Hong Kong. DYPA achieves high predictive accuracy of 81.14%, with sensitivity and specificity 74.27% and 82.71%, respectively. Our results also confirm the efficacy of our multilevel Chinese handwriting analysis framework for dyslexia prediction.

The remainder of this paper is organized as follows. In Section 2, we review the background and related works on dyslexia. We introduce our multilevel Chinese handwriting analysis framework in Section 3, followed by the design and development of DYPA in Section 4. After that, we show our experimental trials and results in Section 5, discuss the advantages, limitations, and potential future direction of our work in Section 6, and conclude in Section 7.

2 BACKGROUND AND RELATED WORK

2.1 Dyslexia in Chinese

Dyslexia is characterized by its diverse causes and heterogeneous manifestations [33]. In research of dyslexia, there has not been a consensus across the world on a coherent definition, since the specific learning difficulties associated with dyslexia are largely language-based [10, 32, 37]. Similar to research on dyslexia in alphabetic languages, researches on Chinese dyslexia have discussed cognitive-linguistic skills, including phonological sensitivity, morphological/semantic awareness, orthographic processing, and fluency from the aspect of reading [34]. Nevertheless, recent studies demonstrate the importance of both reading and writing to identify dyslexia [7], and this issue is more significant in the Chinese morphosyllabic writing system, since most Chinese characters are semantic-phonetic compounds. Evidence suggests that Chinese children with dyslexia exhibit deficits in cognitive-linguistic skills in both reading and character writing, and writing-related features can also help distinguish children with dyslexia [28, 33].

Despite the theoretical advances, current diagnosis of dyslexia in Chinese still focuses mainly on reading deficits. For example, the diagnosis of dyslexia among Hong Kong children is based on a local standardized test, known as the Hong Kong Test of Specific Learning Difficulties in Reading and Writing for Primary School Students (HKT-P) [8, 9]. It consists of character recognition, word reading and dictation, and other cognitive-linguistic tests. Composite scores are evaluated from the test, and children are formally diagnosed with dyslexia when they score significantly below the age norm. While HKT-P contains word dictation tests that require writing, the scores are only computed as binary correctness, which neglects in-depth analysis of handwritten characters. Moreover, these diagnostic tests are usually paper-based and lengthy, which are painful for children to take. Meanwhile, they require accredited specialists to manually score the results and make the final decision on the basis of the marks [2, 8, 46]. The services of accredited specialists are expensive in general and undersupplied, which makes it hard to be timely accessible to the children in need. As a result, a large proportion of dyslexic children would miss the best period for therapy [6, 52].

2.2 Computer-Assisted Dyslexia Prescreening

Since the last decade, machine learning and deep learning have been extensively studied for data analysis and prediction of dyslexia. Literature review of recent advances in this area are made [1, 24, 25, 27]. Among these works, some use paper-based psycho-educational tests to examine children's cognitive-linguistic skills (e.g. [26]). These approaches still require specialists (teachers, doctors, EdPsys) to manually score each test before the results are fed into machine learning models for dyslexia prediction. Other approaches make use of magnetic resonance imaging(e.g. [16]), electroencephalogram (e.g. [44]), or eye movement tracking (e.g. [53]) for data collection and apply machine learning and deep learning to extract biological features from the data for dyslexia prediction. However, these approaches require costly specialized equipment, which prevents them from being convenient and cost-effective.

The prevalence of personal computing devices provides a great opportunity to automate the dyslexia screening process. Web and mobile applications and games are developed to screen children with dyslexia in an engaging, convenient, and cost-efficient way [38, 39]. Most of them are designed for Children speaking specific languages. Rello et al. [42, 43] proposed a web-based game with stages to identify dyslexia in German, English, and Spanish. Their game design considered multiple cognitive-linguistic skills, memory, and attention. And their approach achieves an accuracy of 83% in English. Francese et al. [18] proposed an Android mobile application to identify reading disabilities in Italian children. They propose specific metrics with thresholds for the detection. Tenemaza et al. [50] proposed a mobile application in Spanish with augmented reality (AR) technology to improve the identification of dyslexia in school-aged children. Their result validates the effectiveness of software with an AR interface for early dyslexia detection. However, most of these approaches are designed for English and other alphabetic languages. They focus more on cognitive-linguistic skills in terms of reading, especially phonological sensitivity, and fail to consider Chinese-specific features like handwriting for dyslexia prediction, thereby not readily applicable for dyslexia in Chinese. Some recent studies propose language-independent games for dyslexia prescreening. Rauschenberger et al. [41] presented a web-based game built with language-independent musical and visual elements for dyslexia detection and demonstrated its efficacy on children speaking different languages. Rauschenberger et al. [40] further combined this system with different machine learning models and achieved an accuracy of 74% in German and 69% in Spanish. While impressive, the performances of these approaches are only evaluated among children from western countries. As dyslexia is acknowledged to depend largely on language and culture [10, 32, 37], the efficacy and performance of these approaches remain to be further validated among Chinese children.

To investigate Chinese handwriting performance of primary school children with dyslexia, Lam et al. [28] collected and analyzed students' handwriting data using the Chinese Handwriting Assessment System (CHAS) [29]. They found the writing speed and accuracy as satisfactory discriminators for dyslexia with an accuracy over 70%. However, CHAS collects students' handwriting data with specialized graphics tablets, which are not accessible in common schools and families. Meanwhile, since CHAS is not designed for dyslexia prescreening, it mainly assesses students' handwriting in terms of speed, writing pressure, and stroke-level accuracy, while ignoring the multilevel features across the stroke, radical, and character levels, and associated cognitive-linguistic skills, which are critical for dyslexia prescreening. Yan et al. [55] designed a tablet application for Chinese dyslexia prescreening that bases on handwriting in stroke and radical levels. However, their method still relies on human expertise to manually assess and score students' writing. Man Kit Lee et al. [33] proposed to use machine learning with multilevel writing features from character dictation and demonstrated an accuracy of 78%. However, their method is still paper-based and not automated which requires manual scoring of the data. Compared with these approaches, DYPA is specially designed for dyslexia that it comprehensively analyzes multilevel features in writing as well as other multidimensional cognitive-linguistic skills. At the same time, DYPA runs on consumer-grade touch screen tablets which have been available in many schools and families such that no costly specialized equipment is needed compared with CHAS. Moreover, the DYPA screening process is automated by machine learning and deep learning that does not rely on any manual work by expensive service of specialists to score or label the data compared with Yan et al. [55] and Man Kit Lee et al. [33]. These advantages make DYPA more comprehensive and more cost-effective.

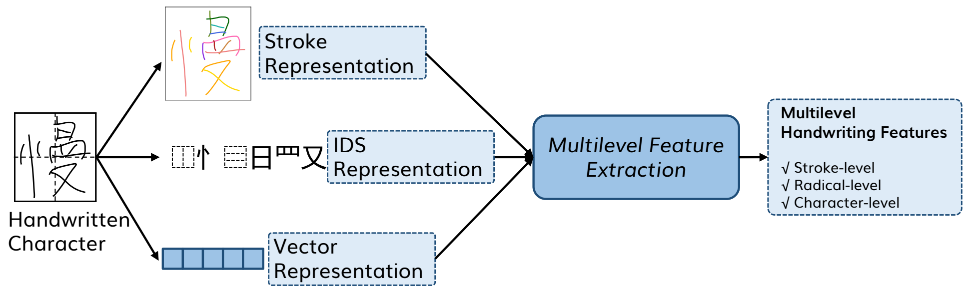


Fig. 3. Multilevel Chinese handwriting analysis framework.

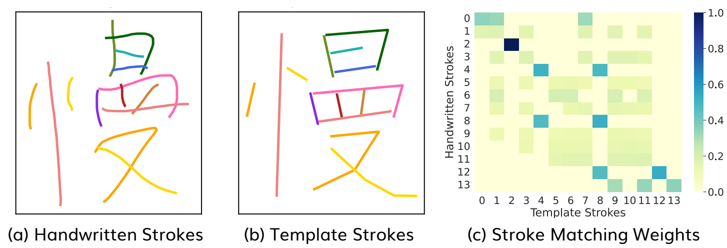


Fig. 4. An example of stroke matching of a character containing 14 strokes in total. (a) and (b) use the same color to represent corresponding strokes in the handwritten and template character. (c) shows the matrix of matching weights between each stroke pair.

3 MULTILEVEL CHINESE HANDWRITING ANALYSIS FRAMEWORK

Chinese characters are hierarchically composed of orthographic units such as subcharacters, radicals, and strokes. These orthographic units typically map onto phonology or semantics by certain regularities. Different compositions, proportions and orientations of them can form different characters and carry totally different meanings and pronunciations [12]. The learning deficits caused by dyslexia can manifest in the acquisition of these regularities across the stroke, radical, and character levels. Thereby, jointly analyzing the multilevel features of writing can help better examine these cognitive-linguistic skills for dyslexia identification. To this end, we propose a novel multilevel Chinese handwriting analysis framework, as shown in Figure 3. In the writing tests in DYPA, students are supposed to write Chinese characters in a square area specified on the touch screen of their tablets. Each stroke of their writing is recorded as a sequence of points in the area, and for each handwritten character, we record the list of strokes in the order of their writing. For multilevel handwriting analysis, we first compute three different representations of each handwritten character, i.e., stroke, Ideographic Description Sequence (IDS), and vector representations. Then, by jointly analyzing these representations, we extract features in the stroke, radical, and character levels to be used in dyslexia prescreening. In this section, we introduce the computation of the representations (Section 3.1-3.3), and present multilevel features extracted (Section 3.4).

3.1 Stroke Representation

Strokes are the basic components that form Chinese characters. There are 8 basic types of strokes that comprise most Chinese characters. Within each character, the strokes are supposed to be written in a certain order. During writing, a single stroke can include a sequence of motions, thus having abrupt changes in direction within the line. Considering the characteristics of stroke writing, we generate the stroke representation of children's handwritten

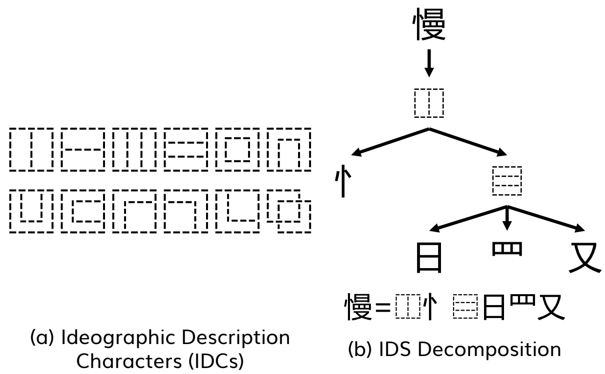


Fig. 5. An example of IDS.

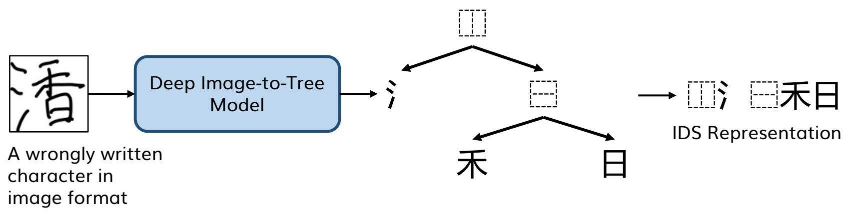


Fig. 6. The generation of IDS representation from a wrongly written character.

characters by stroke matching, to find out the optimal one-to-one correspondence between handwritten strokes and the template [3]. As the example shown in Figure 4, we first segment each stroke by the turning points. Then, we compute pairwise matching weights between each handwritten stroke and the template strokes. For each stroke pair, the matching weight is a sum of normalized (i) centroid distance, differences in (ii) length, (iii) aspect ratio, (iv) direction, and (v) count of segments contained. Finally, the optimal matching correspondence is selected following the condition that it has the largest count of matching strokes with the maximum matching weights. By doing this, we represent each handwritten character by its strokes together with their best-matched strokes in the template.

3.2 IDS Representation

The Ideographic Description Sequence is a syntax to describe the hierarchical structure of ideographic characters including Chinese [15, 35]. It decomposes each character into a combination of its subcharacters and uses special Ideographic Description Characters (IDCs) to specify the structural relationship between them. Since the subcharacters can be recursively decomposed using IDS, the decomposition is actually tree-structured and the IDS representation is obtained by depth-first traversal (DFS) of the decomposition tree (Figure 5). Moreover, IDS is flexible that it can describe both characters and noncharacters, thus very useful to describe children's handwritten characters, which can be wrongly written and not directly recognizable as normal characters.

Research has been done to leverage deep image-to-sequence models to convert Chinese characters in images into their corresponding IDS [57]. Considering the inherent tree structure of Chinese characters at the radical level as well as the tree structure of the IDS, we propose to use a deep image-to-tree model instead for better translation from handwritten characters to IDS. Here we employ TSDNet [58], a deep learning model well-designed for learning to extract tree-structural semantics from images. TSDNet works with a CNN as the encoder to extract visual features from the input image, and introduces a tree-based structure-aware Transformer decoder to decode the visual features into the target tree sequentially. TSDNet makes full use of the structural information in both the input image and the output tree during the decoding process, thus being good for us to understand the radicals and their structural relationships in the characters.

As shown in Figure 6, we first train our TSDNet model using images of Chinese characters as input and their corresponding IDS decomposition tree as output. The model thereby learns to recognize the subcharacters and their structural relationship with the character in the input image. Then we generate an image for each handwritten character from the strokes. Next, we feed the images of children's handwritten characters to our trained TSDNet model to generate the corresponding IDS decomposition tree. After obtaining the decomposition trees of children's handwritten characters, we use DFS to convert the trees into IDS and use them for the IDS representation of those handwritten characters.

3.3 Vector Representation

The strokes and IDS representation provide fine-grained insights into the handwritten characters. Nevertheless, we also need to obtain holistic views of the characters. To this end, we leverage a deep convolutional neural network to generate the vector representation of each handwritten character. Here we employ DenseNet [23] as our backbone network for feature extraction from characters in handwritten images. We first train our deep CNN model on handwritten Chinese character recognition datasets. Then, we convert the children's handwritten characters into images and fed the images to our trained CNN model. We extract the output feature vector of the CNN's last hidden layer and use it as the vector representation of that handwritten character.

3.4 Multilevel Feature Extraction

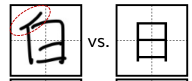
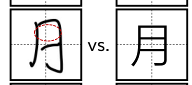
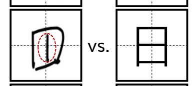
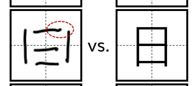
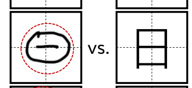
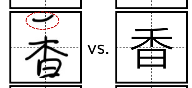
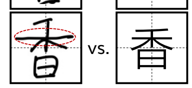
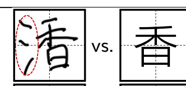
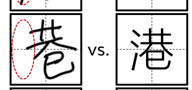
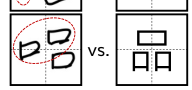
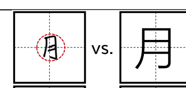
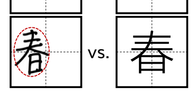
With the three representations computed from children's handwritten characters, we extract multilevel features for dyslexia prescreening. Table 1 summarizes the multilevel features we extracted from the representations.

At the stroke level, we first identify the additional, omitted, and wrongly written strokes from the unmatched strokes in the stroke representation. Then, we recognize broken stroke errors where one stroke in the template is split into several strokes in the handwritten character, as well as joined-up stroke errors where several strokes in the template are written in one stroke in the handwritten character. It is done by comparing the stroke segments in the stroke representations. Next, we detect disproportionate spacing and size among matched strokes. To detect disproportionate spacings, we first compute pairwise centroid distances of each pair of handwritten strokes and compare them with that of their matched stroke pairs in the template. We consider it as too compact/loose spacing when the distance is significantly smaller/larger than the template, respectively. Similarly, we detect disproportionate stroke sizes by computing the pairwise length ratios of each pair of handwritten strokes and comparing them with that in their matched template. We consider it as undersizing/oversizing when the length ratio is significantly smaller/larger than the template strokes. Last, we compute the stroke level similarity as the matched number of strokes divided by the correct number of strokes.

At the radical level, we first identify the additional, omitted, and wrongly written radicals by comparing the IDS representation with the ground truth. Then, we detect wrong spatial layout of radicals by comparing the IDCs in the IDS with the ground truth. Last, we compute the radical level similarity as $1 - d$ where d is the edit distance between the IDS and the ground truth normalized by the sum of their length.

At the character level, we first compute the coordinates of the whole handwritten character's bounding box from the stroke representations. Then, we identify disproportionate character size by comparing the aspect ratio of the bounding box with that of the template. Next, we detect transposition of the character by computing the distance between the bounding box centroid and the center of the writing area. Last, we compute the character

Table 1. Multilevel features and examples, with mistakes in the examples circled out.

Level	Description	Examples vs. Templates
Stroke	Additional strokes	
Stroke	Omitted strokes	
Stroke	Wrong strokes	
Stroke	Broken strokes	
Stroke	Joined-up strokes	
Stroke	Disproportionate stroke spacings	
Stroke	Disproportionate stroke sizes	
Stroke	Stroke level similarity	-
Radical	Additional radicals	
Radical	Omitted radicals	
Radical	Wrong spatial layouts	
Radical	Radical level similarity	-
Character	Disproportionate character sizes	
Character	Transposition	
Character	Character level similarity	-

level similarity by the cosine similarity between the vector representation of the handwritten character and that of the template.

4 DYPA DESIGN AND IMPLEMENTATION

Figure 2 shows the overall workflow of our proposed DYPA system. It is composed of a tablet application that is responsible for interaction with users, along with a server program to complete computation-intensive tasks including the multilevel Chinese handwriting analysis and the machine learning dyslexia prediction. In this section, we introduce the design of the DYPA tablet application in Section 4.1, and how dyslexia is predicted by machine learning in Section 4.2.

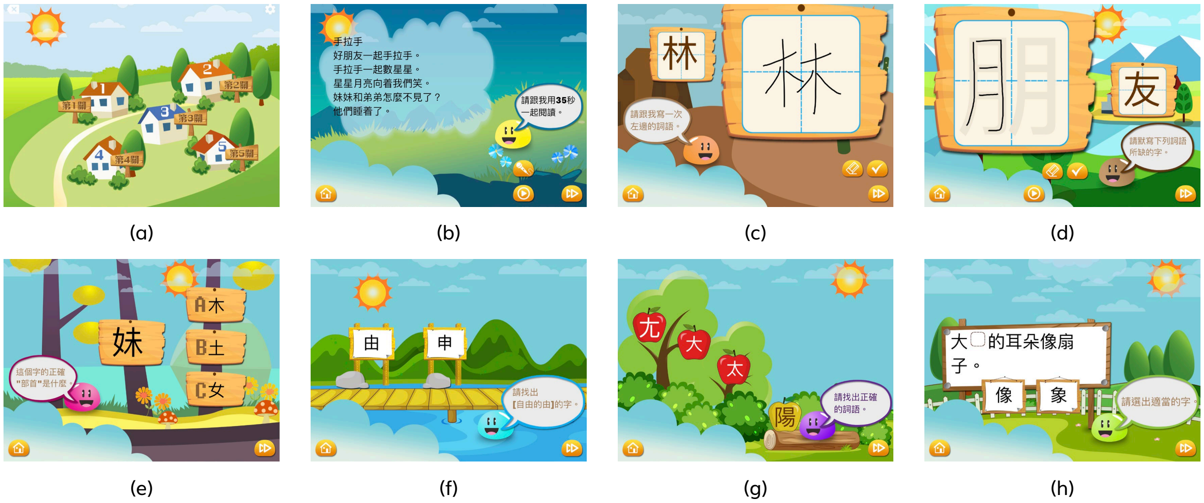


Fig. 7. Screenshots of DYPA Tablet App.

4.1 DYPA Tablet App

DYPA comprehensively examines children's literacy and cognitive-linguistic skills for dyslexia prediction through interactive tests in the DYPA mobile application. As shown in Figure 7, these tests are designed to be cartoonized games so as to be engaging and friendly for children to play with. There are three forms of tests in DYPA, i.e. the reading tests, the writing tests, and the multiple choice tests. In a *Reading Test* (Figure 7(b)), children are supposed to read a sentence, and the time spent will be recorded. In a *Writing Test* (Figure 7(c-d)), children are supposed to write Chinese characters in a square area specified on the touch screen of their tablets, as described in Section 3, and their writing trajectories and the time spent will be recorded. In a *Multiple Choice Test* (Figure 7(e-h)), children are supposed to choose the correct answer by touching the corresponding answer on the touch screen. Their choice together with their time spent will be recorded.

The test contents are specially designed by our EdPsys considering the characteristics of Chinese dyslexia to assess children's cognitive-linguistic skills in terms of reading, writing, phonological, orthographical, and morphological dimensions [34]. Among them, tests for phonological, orthographical, and morphological skills are in the form of multiple choice tests. As shown in Table 2, we further categorize the tests into 12 types by the corresponding skills they examine. Test 1 examines children's reading fluency. Students are required to read the sentence displayed on the screen. Tests 2 and 3 are writing tests. In Test 2, students are asked to copy a given character once, and in Test 3, students are asked to write down the missing character in the 2-character word they hear. Tests 4-6 examine the phonological sensitivity that students are required to choose the character with the same onset (Test 4), rime (Test 5), and pronunciation (Test 6) of the character in the given word. Tests 7-10 are orthographical tests. In Test 7, students are asked to distinguish the correct orientation of the given character from their leftside-right and upside-down reversals as well as rotations. Test 8 requires students to select the radicals that appear in the given character, and Test 9 requires students to choose the correct character that matches the given spatial layout. Tests 10-12 examine morphological/semantic awareness. Students are asked to choose the correct character to be filled into the given word (Test 10) and sentence (Test 11), and choose the correct word to be filled into the given sentence (Test 12). The choices provided contain similar characters and words that require children's semantic awareness to distinguish. Each of these 12 test types contains different questions based on different characters and/or words. Children are supposed to complete the tests several times (Figure 7(a)), during which questions from different test types appear in a random manner.

4.2 Machine Learning Dyslexia Prediction

After the tests, the test data, including time used in each test, correctness of their choices, and the handwriting data are forwarded to the server for processing. The handwriting data is processed by our multilevel Chinese handwriting analysis framework for feature extraction and fed into a machine learning classifier together with other test data, to predict whether a child has a risk of dyslexia or not. In DYPA, we employ XGBoost [14] as the final machine learning classifier in DYPA. XGBoost is a machine learning algorithm of the gradient boosting paradigm [19]. It is fast and accurate to solve many classification tasks. Through XGBoost as the machine learning classifier, we obtain the dyslexia prediction, together with the probability of that prediction that indicates the risk of dyslexia. These results, together with the children's test records and features we have extracted, will be sent to children's parents, teachers, or corresponding EdPsys. The results can provide advice and reference further clinical diagnosis, special education needs, and specific therapy if needed.

5 EXPERIMENTAL TRIALS AND ILLUSTRATIVE RESULTS

To evaluate the effectiveness of our proposed DYPA, we conducted experimental trials in Hong Kong, where students learn to speak Cantonese, a branch of Chinese widely spoken in southern China, and write traditional Chinese characters in school. In this section, we introduce implementation details of DYPA in Section 5.1, and our experiment settings in Section 5.2. Then, we discuss the prescreening performance of DYPA in Section 5.3, and analyze the importance of features extracted in Section 5.4. We further evaluate different machine learning algorithms to be used in DYPA, as well as the contribution of our multilevel Chinese handwriting analysis framework for dyslexia prescreening by ablation studies in Section 5.5.

5.1 Implementation Details

In the experimental trials, we designed the test contents in DYPA based on commonly used words and characters which are supposed to have been taught in kindergarten and early primary school stages in Hong Kong. We implemented DYPA tablet app on Apple iPad with the iPadOS system.

In the multilevel Chinese handwriting analysis framework in DYPA, we build the TSDNet model for IDS representation generation following the original configuration [58]. To train this model, we construct a dataset by selecting 7072 most frequently used traditional Chinese characters[1], and randomly split them into train (80%), valid (5%), and test (15%) subsets. We then generate images of these characters from fonts of Source Han Sans[2] and Serif[3], and align the images with their corresponding IDS decomposition trees. We train our TSDNet model on this dataset using the images as input and the IDS decomposition trees as output. The results on the test set show that our model achieves 90.8% accuracy to precisely translate the given input image into its corresponding IDS decomposition tree. Meanwhile, we build the deep CNN model for vector representation generation following the DenseNet-121 configuration [23]. We train the model on the CASIA-HWDB1.0-1.2 dataset for handwritten Chinese character recognition, using the original dataset split [31]. The results on the test set show that our model achieves 94.9% accuracy in recognizing handwritten characters in the given images.

[1] https://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/faq.php
[2] https://github.com/adobe-fonts/source-han-sans
[3] https://github.com/adobe-fonts/source-han-serif

Table 2. Test types in DYPA.

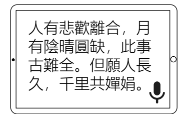
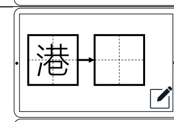
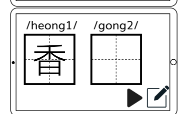
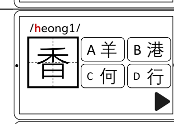
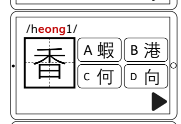
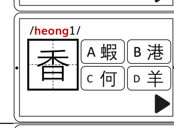
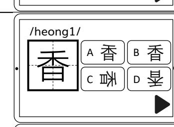
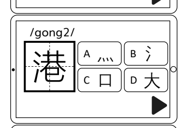
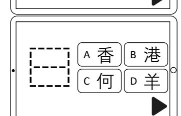
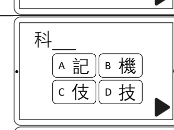
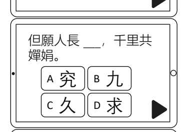
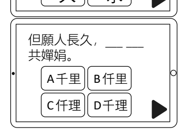
No.	Category	Test Type	Example
1	Reading	Sentence reading	
2	Writing	Character copying	
3	Writing	Character in word dictation	
4	Phonological	Same onset	
5	Phonological	Same rime	
6	Phonological	Homophone	
7	Orthographical	Flip and rotations	
8	Orthographical	Radical in character	
9	Orthographical	Spatial layout	
10	Morphological	Character in word	
11	Morphological	Character in sentence	
12	Morphological	Word in sentence	

Table 3. The demographic information of participants.

Grade	Gender (Male/Female)	Dyslexia (Yes/No)
1	48 / 28	6 / 70
2	42 / 37	16 / 63
3	31 / 21	17 / 35

5.2 Experiment Settings

We conducted our study on Hong Kong primary school students. It was conducted with the oral consent of each student, oral and written consent of each student's parents or legal guardians, each student's teacher, and the school authorities. We randomly invited students between grade 1 to 3 and eventually recruited 207 from three public primary schools in Hong Kong as our participants. None of them reported any intellectual or other uncorrected sensory impairments. And all students have taken the local standardized test for dyslexia by accredited specialists, where 39 of the students were formally diagnosed with dyslexia. In this study, we assume the diagnosis by accredited specialists was accurate and there were no undiagnosed or misdiagnosed cases in our data. Detailed distribution of our participants is shown in Table 3. For the writing tests in DYPA, students wrote with their fingertips on the touch screen. During the experiment period of up to six months, the participants were required to play DYPA three times under proper guidance to avoid invalid experiment data, and we assume the associated conditions don't change over the experiment period. More specifically, each student finished five reading tests, forty-five multiple choice tests and fifteen writing tests each time. We have introduced the details of these three test types and the corresponding data we collect for each of them in Section 4.1. Eventually, we average the relevant scores of all three times as the final input for our classifier.

To quantify the evaluation of our algorithm which solves a binary classification problem, we adopt the commonly used metrics including accuracy, sensitivity, specificity defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{2}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{3}$$

where *TP*, *TN*, *FP*, and *FN* stand for the number of true positive, true negative, false positive, and false negative cases in the result respectively, as well as area under curve (AUC). We define dyslexic as the positive label and non-dyslexic as the negative label in our experiments. In this way, sensitivity and specificity reflect the performance of our method to correctly identify dyslexic children and non-dyslexic children. In addition to the first three metrics, the receiver operating characteristic (ROC) curve further reveals the prescreening performance under different discrimination thresholds. We train and evaluate our machine learning classifier in DYPA in a 5-fold cross-validation manner. Data stratification is applied during sampling to balance the ratio of dyslexia and healthy children in each of the 5 folds.

5.3 Prescreening Performance

Table 4 shows the performance of our proposed DYPA. We also present results of the following screening or diagnostic systems reported in previous papers for comparison.

- *Dytective [43]* is a web-based game to identify dyslexia in English, and the results are validated on English-speaking children.

Table 4. Comparison of DYPA performance with other dyslexia prescreening approaches in terms of accuracy (Acc), sensitivity (Sen), specificity (Spe) and area under curve (AUC).

Method	Language & Features	# of Participants	Acc (%)	Sen (%)	Spe (%)	AUC
Dytective [43]	English	267	84.62	80.24	85.83	-
MusVis [40]	Language-independent	313	74	77	-	-
Man Kit Lee et al. [33]	Chinese dictation	1015	75.1	68.7	80.2	0.83
DYPA (ours)	Chinese reading & writing	207	81.14	74.27	82.71	0.79

Table 5. The number of false positives, false negatives and their corresponding proportions over age and gender.

	Grade 1	Grade 2	Grade 3	Male	Female
False Positive (Ratio %)	11 (37.93%)	10 (34.48%)	8 (27.59%)	18 (62.07%)	11 (37.93%)
False Negative (Ratio %)	1 (10%)	5 (50%)	4 (40%)	7 (70%)	3 (30%)

- *MusVis [40]* is a web-based game built with language-independent musical and visual elements for dyslexia detection. The results are validated on a group of English, German, and Spanish speaking children.
- Man Kit Lee et al. [33] uses manually labeled features from children's handwritten characters with machine learning to predict dyslexia. The results are validated on Chinese children.

From the results, we can observe that DYPA reaches a high predictive accuracy of 81.14%, together with good sensitivity and specificity of 74.27% and 82.71%, respectively. DYPA also achieves a good AUC of 0.79. These results also show that DYPA does not over-fit to the dataset. DYPA achieves comparable prescreening accuracy to the system for alphabetic languages by Rello et al. [43], and outperforms the language-independent system by Rauschenberger et al. [40], which indicates that DYPA can capture Chinese-specific features for accurate dyslexia prescreening in Chinese. On the other hand, DYPA outperforms the approaches by Man Kit Lee et al. [33] that using Chinese handwriting features alone, which shows that by DYPA is able to comprehensively consider the learning deficits in terms of reading and writing to identify dyslexia in Children.

To analyze the prediction error, we list the distribution of mispredicted students over age and gender in Table 5. The numbers of false positives from grade 1 to grade 3 are comparable to each other. Since there are only six dyslexia students in grade 1, it is hard to draw any conclusions from the false negatives in grade 1, and the number in grade 2 and grade 3 are close to the ground truth dyslexia ratio 1 : 1.06. Hence, we believe that age is not a significant factor of prediction error. When it comes to gender, false positives have a gender ratio of 1.64 : 1 which is close to the ratio 1.41 : 1 for all students. The gender ratio of false negatives, 2.33 : 1, is higher than the overall ratio, but given the fact that the gender ratio in dyslexia population is around 2 : 1 [11], we believe that our experiment results do not have a special distribution over gender.

5.4 Feature Importance

To understand how features from different test categories contribute to the identification of dyslexia, we compute the feature importance for our XGBoost classifier. We choose the impurity-based feature importance [47], which counts the number of times a specific feature is used for a split. A higher importance indicates that the feature is more frequently used and thus has more contribution to the classifier. The results are normalized and have a sum of one. As shown in Figure 8, features from all categories contribute to the prediction, which proves the effectiveness of the tests and features we designed in DYPA. Among those features, three writing-based features contribute the most, which reflect the complicated manifestations of Chinese dyslexia in writing. More specifically, stroke-level and radical-level features have visibly higher importance than character-level features, demonstrating

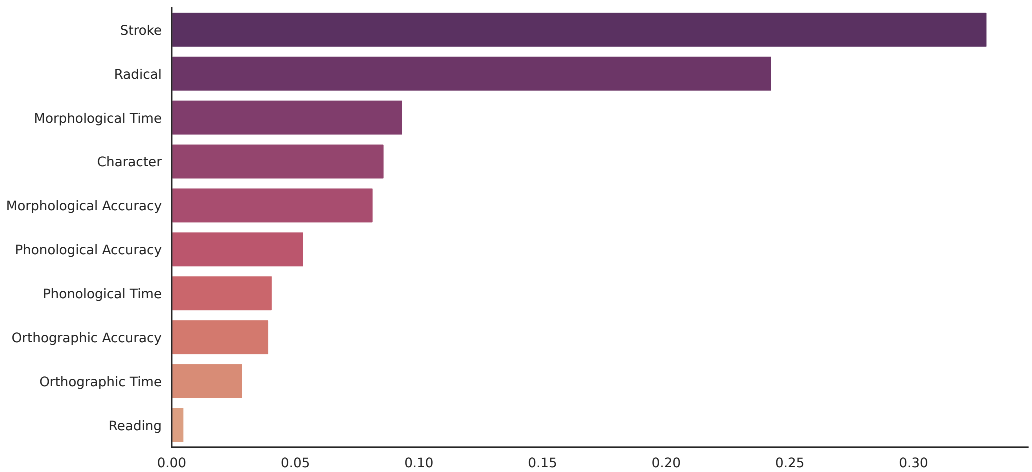


Fig. 8. The importance of different features in DYPA.

that the two subcharacter level features are more discriminative between dyslexic and non-dyslexic students. Among the other features, morphological features are more predictive than phonological features which are acknowledged to be the most predictive indicator of dyslexia in alphabetic languages. This is consistent with previous studies on Chinese dyslexia that the importance of phonological sensitivity as an indicator of dyslexia in Chinese is less than that of alphabetic languages [34]. It also confirms previous studies that reveal the importance of morphological awareness in Chinese dyslexia [48].

5.5 Ablation Study

We study different choices of machine learning algorithms in DYPA and their influence on prescreening performance. The results are shown in Table 6. KNN and MLP suffer from low sensitivity, resulting in a high possibility of false negatives. Their slightly better accuracy comes at a cost that around 30% of the potential patients could miss the best period for therapy. Meanwhile, logistic regression and naive Bayes are not satisfying in terms of specificity, meaning that they are more likely to predict a healthy user as dyslexic. The SVM and XGBoost algorithms demonstrate good prescreening accuracy as well as a more balanced prescreening ability on both dyslexic and healthy users rather than in favor of only one group. Thus, we eventually choose XGBoost as the final classifier in our DYPA system for its slightly higher accuracy over SVM.

To further evaluate the efficacy of the character writing tests and our multilevel Chinese handwriting analysis framework in a quantitative way, we study the ablations of DYPA by only using a subset of our three-level handwriting features, namely stroke, radical and character. Their performances and some of the corresponding ROC curves are shown in Table 4 and Figure 9. First, by removing all three handwriting features from DYPA, the model suffers a significant accuracy drop by 17.38%. When we consider one of the three features into DYPA, the prescreening accuracy enjoys at least a 7% rise. As the most important feature type, the stroke level features alone can bring up the accuracy by more than 10%. This result, together with the feature importance analysis above, demonstrates the significance of handwriting in Chinese dyslexia prescreening. Among the three variants that

Table 6. Comparison of different classifiers in DYPA.

Name	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Naive Bayes	64.74	97.14	57.09	0.77
Logistic Regression	72.92	97.14	67.20	0.82
KNN	85.01	33.21	96.99	0.65
SVM	77.27	79.64	76.77	0.78
MLP	83.08	48.57	91.05	0.70
XGBoost	81.14	74.27	82.71	0.79

Table 7. Comparison of DYPA performances using different feature sets.

Stroke	Radical	Character	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
			63.76	51.07	66.67	0.59
✓			74.89	58.93	78.57	0.69
	✓		70.50	48.93	75.51	0.62
		✓	72.95	58.57	76.17	0.67
	✓	✓	74.36	59.29	77.93	0.69
✓		✓	76.32	68.93	77.97	0.73
✓	✓		80.17	68.93	82.71	0.76
✓	✓	✓	**81.14**	**74.27**	**82.71**	**0.79**

include two of the three feature levels, DYPA *w/o radical* and DYPA *w/o stroke* perform significantly worse than DYPA *w/o character*, which also echoes with our discussion in Section 5.4 on the importance of subcharacter level processing in Chinese character writing. Moreover, the combination of stroke features and radical features boosts the accuracy more significantly than any of these two alone, while this is not observed on all other combinations, which indicates the strong correlation between these two feature levels.

6 DISCUSSION

Our experimental results demonstrate the satisfying performance of our proposed method to prescreen dyslexia in Chinese children. Our findings underscore the potential for using consumer-grade tablets to help average Chinese families and schools to prescreen dyslexia in their children and students in a timely and cost-efficient manner. However, we should note that DYPA is a prescreening approach instead of a clinical diagnostic approach. The result of DYPA is not robust enough compared to clinical diagnosis regarding accuracy, sensitivity, and specificity. And it should never be treated as a substitution for the clinical diagnosis. Instead, we design DYPA to be an assistive tool for the non-professionals such as parents, teachers, and schools for preliminary and cost-effective dyslexia prescreening, such that limited professional resources for advising and treatment can be better utilized and channeled to those at higher risk in a timely manner. While prescreening tools like DYPA can help identify dyslexic children in a preliminary way, they may also make incorrect predictions that can cause problems to the children if not handled properly. How to combine prescreening tools like DYPA and clinical diagnosis services to better identify dyslexic children and avoid potential harm caused need to be further studied considering the realistic situation and available resources of the children, families, schools, and society. Meanwhile, in our experiments, we assume no undiagnosed case by accredited specialists in our data. However, due to the diverse causes and heterogeneous manifestations of dyslexia, as well as different clinical criteria adopted for dyslexia

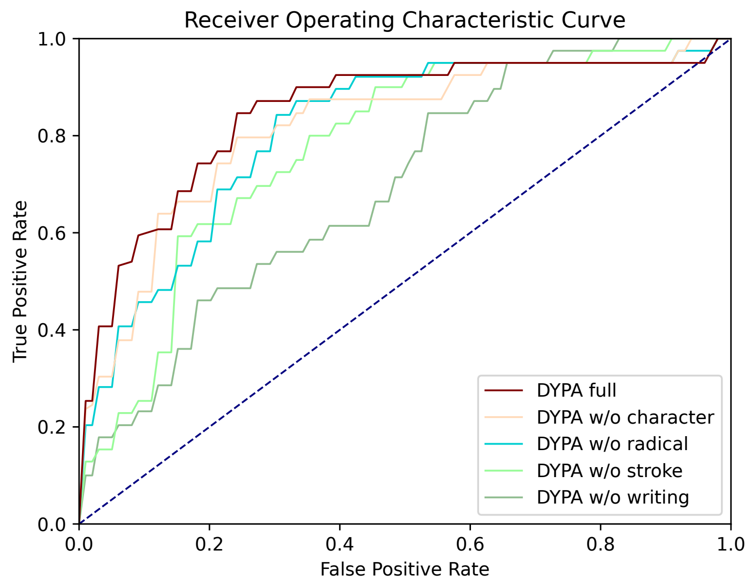


Fig. 9. The ROC curve of DYPA using different feature sets.

diagnosis, there can be a considerable amount of dyslexic children to be undiagnosed [5]. It goes beyond the scope of this study to consider those cases, and we also hope DYPA can assist EdPsys by providing more evidence in their clinical diagnosis to help avoid those cases.

Furthermore, from the feature importance analysis, we find our writing-related features to be significantly discriminative for identifying dyslexia in Children, especially at the stroke and the radical level. Unlike alphabetic languages that emphasize smoothness and continuity in their written forms, strokes in Chinese characters contain frequent sharp turns, diverse directions, and complex geometric configurations [51]. Writing Chinese characters demands visual discrimination of the subtle differences in the form, position, and proportion of strokes, which causes difficulties for children with dyslexia. Meanwhile, most Chinese compound characters are composed of phonetic and semantic radicals, which map onto the rime of a syllable and the meaning of a character by certain regularities. However, these regularities are less predictable than the pronunciation of letters and words in alphabetic languages, since different compositions, proportions and orientations of the radicals in Chinese characters can form different characters and carry totally different meanings and pronunciations [12]. These unique logographic and morphosyllabic features of Chinese necessitate visual, phonetic, and orthographic skills, which are associated with Chinese character acquisition among children with and without dyslexia. On this premise, it is worthwhile to explore applying our multilevel character analysis framework to other writing systems such as Japanese and Korean that employ similar logographic features. In summary, our result is consistent with theoretical researches on dyslexia in Chinese and provides new evidence and observations for the diagnosis and screening of Children with dyslexia.

In DYPA reading tests, we only consider the time spent by children to read a sentence. Although recent development in deep learning and speech recognition makes it possible for us to recognize and analyze speech signals, it is still unclear how to effectively incorporate the speech signals for dyslexia prescreening. It is worth investigating more speech-related features using deep learning based speech recognition technologies to help dyslexia prescreening.

In our proposed multilevel character analysis framework, we use stroke matching, pretrained TSDNet, and pretrained CNN for representation and feature extraction. Regarding the sequential stroke-by-stroke writing process of Chinese characters, it would also be good to consider a combination of CNN and sequence models like recurrent neural networks (RNNs) and Transformers to directly extract multilevel features. However, training a deep learning model of this kind for dyslexia related analysis may require much more dyslexia related data than we are able to collect in this study. Therefore, we adopt the current design that the TSDNet and CNN can be pretrained on available public datasets for feature extraction. Meanwhile, our experimental results demonstrate that our design also achieves a satisfying performance. With more data available, it would be promising to experiment the combination of CNNs and sequence models in the future. Meanwhile, despite the fact that we combine machine learning and deep learning to automate the data processing and dyslexia prediction so as to avoid human expertise involved, one limitation of our proposed DYPA is that the machine learning dyslexia classifier needs to be trained on a manually labeled dataset. The collection of this dataset may be considerably labor-intensive, and the labeling may require the expertise of educational psychologists. After the construction of the labeled dataset and the training of machine learning dyslexia classifier, DYPA can work in an automated way for dyslexia prescreening. It is worthwhile to explore combining crowdsourcing or unsupervised learning technologies to further lessen this effort. Meanwhile, since DYPA is used by young children, it is totally possible that they do not take the tests properly and thus produce invalid data such as unfinished tests or messy handwritings which invalidate the screening results. Therefore, it is important to have them accompanied by their parents or teachers when playing with DYPA to provide proper guidance. It is also worthwhile to leverage human-computer interaction and human-AI collaboration technologies to improve the game design and game process, so as to ensure a more smooth, friendly, and engaging test experience for children, and also avoid invalid input data.

We note that some high performance tablets are already able to handle locally computation-intensive tasks in DYPA, including data processing, feature extraction, and classification based on machine learning and deep learning. And keeping the test data locally is a better practice to protect user privacy. However, our current DYPA system still relies on the server to complete these tasks. This is to relax the requirements of computing power of the tablets that DYPA can run on, such that DYPA can be more accessible and cost-effective to people in need. We leave it as our future work to improve the computation efficiency of algorithms (e.g. [13, 21]) in DYPA so as to make it compatible with more mobile devices.

7 CONCLUSION

In this paper, we focus on the computer-assisted dyslexia prescreening in Chinese children. We propose DYPA, a novel **dy**slexia **p**rescreening mobile **a**pplication for Chinese children. DYPA collects multimodal data from children through a set of specially designed interactive reading and writing tests in Chinese, and comprehensively analyzes their cognitive-linguistic skills with machine learning. To better account for the dyslexia-associated features in handwritten characters, DYPA employs a deep learning based multilevel Chinese handwriting analysis framework to extract features across the stroke, radical and character levels. Experimental results on students with and without dyslexia from Hong Kong demonstrate good predictive accuracy of DYPA to be 81.14%, with sensitivity and specificity to be 74.27% and 82.71%. Results also prove the efficacy of our multilevel Chinese handwriting analysis framework. Future directions of this work may include the extension of the proposed approach to other ideographic languages such as the Japanese and Korean languages, which may share the same characteristics with Chinese in terms of dyslexia, as well as exploring speech-related features with deep learning based speech recognition technologies to help dyslexia prescreening.

ACKNOWLEDGMENTS

This work was supported, in part, by Hong Kong General Research Fund (under grant number 16200120) and Innovation and Technology Fund for Better Living (ITB/FBL/B051/20/P).

REFERENCES

[1] Norah Dhafer Alqahtani, Bander Alzahrani, and Muhammad Sher Ramzan. 2023. Deep Learning Applications for Dyslexia Prediction. *Applied Sciences* 13, 5 (2023), 2804.

[2] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders* (fifth edition ed.). American Psychiatric Association, Washington, DC, USA. https://doi.org/10.1176/appi.books.9780890425596

[3] Weihua An and Chao Li. 2011. Automatic matching of character strokes for computer-aided Chinese handwriting education. In *Proceeding of the international conference on e-Education, entertainment and e-management*. IEEE, Bali, Indonesia, 283–288.

[4] Jason L Anthony and Christopher J Lonigan. 2004. The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of educational psychology* 96, 1 (2004), 43.

[5] Chiara Barbiero, Isabella Lonciari, Marcella Montico, Lorenzo Monasta, Roberta Penge, Claudio Vio, Patrizio Emanuele Tressoldi, Valentina Ferluga, Anna Bigoni, Alessia Tullio, et al. 2012. The submerged dyslexia iceberg: how many school children are not diagnosed? Results from an Italian study. *PloS one* 7, 10 (2012), e48082.

[6] Chiara Barbiero, Marcella Montico, Isabella Lonciari, Lorenzo Monasta, Roberta Penge, Claudio Vio, Patrizio Emanuele Tressoldi, Marco Carrozzi, Anna De Petris, Anna Giulia De Cagno, et al. 2019. The lost children: The underdiagnosis of dyslexia in Italy. A cross-sectional national study. *PLoS One* 14, 1 (2019), e0210448.

[7] Virginia W Berninger, Kathleen H Nielsen, Robert D Abbott, Ellen Wijsman, and Wendy Raskind. 2008. Writing problems in developmental dyslexia: Under-recognized and under-treated. *Journal of school psychology* 46, 1 (2008), 1–21.

[8] C. S. H. Ho, D. Chan, K. Chung, S. M. Tsang, S. H. Lee, and C. Y. C. Fong. 2015. *The Hong Kong test of specific learning difficulties in reading and writing for primary school students - third edition [HKT-P (III)].* Hong Kong Specific Learning Difficulties Research Team, Hong Kong SAR, China.

[9] C. S. H. Ho, D. Chan, K. Chung, S. M. Tsang, S. H. Lee, and R. W. Y. Cheng. 2007. *The Hong Kong test of specific learning difficulties in reading and writing for primary school students - second edition [HKT-P (II)].* Hong Kong Specific Learning Difficulties Research Team, Hong Kong SAR, China.

[10] David W Chan. 2002. Developmental dyslexia in Hong Kong: An overview and lessons from an international perspective. *Educational Journal* 30, 2 (2002), 1–20.

[11] David W Chan, Connie Suk-han Ho, Suk-man Tsang, Suk-han Lee, and Kevin KH Chung. 2007. Prevalence, gender ratio and gender differences in reading-related cognitive abilities among Chinese children with dyslexia in Hong Kong. *Educational Studies* 33, 2 (2007), 249–265.

[12] Lily Chan and Terezinha Nunes. 1998. Children's understanding of the formal and functional characteristics of written Chinese. *Applied psycholinguistics* 19, 1 (1998), 115–131.

[13] Jierun Chen, Tianlang He, Weipeng Zhuo, Li Ma, Sangtae Ha, and S.-H. Gary Chan. 2022. TVConv: Efficient Translation Variant Convolution for Layout-Aware Visual Processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE/CVF, New Orleans, LA, USA, 12548–12558.

[14] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

[15] cjkvi. 2022. IDS data. https://github.com/cjkvi/cjkvi-ids original-date: 2011-05-23T16:04:00Z.

[16] Laura Tomaz Da Silva, Nathalia Bianchini Esper, Duncan D Ruiz, Felipe Meneguzzi, and Augusto Buchweitz. 2020. Visual explanation for identification of the brain bases for dyslexia on fMRI data. *arXiv preprint arXiv:2007.09260* (2020).

[17] Jean-François Démonet, Margot J Taylor, and Yves Chaix. 2004. Developmental dyslexia. *The Lancet* 363, 9419 (2004), 1451–1460.

[18] Rita Francese, Clara Monaco, and Claudia Nicoletti. 2018. An Android Application for Helping in the Identification of Children with Reading Difficulties. In *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good* (Bologna, Italy) *(Goodtechs '18)*. Association for Computing Machinery, New York, NY, USA, 226–231. https://doi.org/10.1145/3284869.3284915

[19] Jerome H Friedman. 2002. Stochastic gradient boosting. *Computational statistics & data analysis* 38, 4 (2002), 367–378.

[20] Sheryl M. Handler and Walter M. Fierson. 2011. Learning Disabilities, Dyslexia, and Vision. *Pediatrics* 127 (2011), e818 – e856.

[21] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. 2019. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*. IEEE/CVF, Long Beach, CA, USA, 1314–1324.

[22] Zhi-Hui Hu, Yun Xu, Liu-Sheng Huang, and Howard Leung. 2009. A Chinese Handwriting Education System with Automatic Error Detection. *Journal of Software* 4, 2 (April 2009), 101–107. https://doi.org/10.4304/jsw.4.2.101-107

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. 2017. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 2261–2269.

[24] A Jothi Prabha and R Bhargavi. 2019. Prediction of dyslexia using machine learning — a research travelogue. In *Proceedings of the Third International Conference on Microelectronics, Computing and Communication Systems: MCCS 2018*. Springer, Singapore, 23–34.

[25] Shahriar Kaisar. 2020. Developmental dyslexia detection using machine learning techniques: A survey. *ICT Express* 6, 3 (2020), 181–184.

[26] Rehman Ullah Khan, Julia Lee Ai Cheng, and Oon Yin Bee. 2018. Machine learning and Dyslexia: Diagnostic and classification system (DCS) for kids with learning disabilities. *International Journal of Engineering & Technology* 7, 3.18 (2018), 97–100.

[27] Pavan Kumar Varma Kothapalli, V. Rathikarani, and Gopala Krishna Murthy Nookala. 2022. A Comprehensive Survey on Predicting Dyslexia and ADHD Using Machine Learning Approaches. In *Inventive Systems and Control (Lecture Notes in Networks and Systems)*. Springer Nature, Singapore, 105–121. https://doi.org/10.1007/978-981-19-1012-8_8

[28] Sutie ST Lam, Ricky KC Au, Howard WH Leung, and Cecilia WP Li-Tsang. 2011. Chinese handwriting performance of primary school children with dyslexia. *Research in developmental disabilities* 32, 5 (2011), 1745–1756.

[29] Cecilia W.P. Li-Tsang, Agnes S.K. Wong, Howard W.H. Leung, Joyce S. Cheng, Billy H.W. Chiu, Linda F.L. Tse, and Raymond C.K. Chung. 2013. Validation of the Chinese Handwriting Analysis System (CHAS) for primary school students in Hong Kong. *Research in Developmental Disabilities* 34, 9 (2013), 2872–2883. https://doi.org/10.1016/j.ridd.2013.05.048

[30] Yuhang Lin, Xuanzhi Zhang, Qingjun Huang, Laiwen Lv, Anyan Huang, Ai Li, Kusheng Wu, and Yanhong Huang. 2020. The Prevalence of Dyslexia in Primary School Children and Their Chinese Literacy Assessment in Shantou, China. *International Journal of Environmental Research and Public Health* 17, 19 (2020), 7140. https://doi.org/10.3390/ijerph17197140

[31] Cheng-Lin Liu, Fei Yin, Da-Han Wang, and Qiu-Feng Wang. 2011. CASIA online and offline Chinese handwriting databases. In *2011 international conference on document analysis and recognition*. IEEE Computer Society, Los Alamitos, CA, USA, 37–41.

[32] Li Liu, Ran Tao, Wenjing Wang, Wenping You, Danling Peng, and James R Booth. 2013. Chinese dyslexics show neural differences in morphological processing. *Developmental cognitive neuroscience* 6 (2013), 40–50.

[33] Stephen Man Kit Lee, Hey Wing Liu, and Shelley Xiuli Tong. 2022. Identifying Chinese Children with Dyslexia Using Machine Learning with Character Dictation. *Scientific Studies of Reading* 0, 0 (June 2022), 1–19. https://doi.org/10.1080/10888438.2022.2088373

[34] Catherine McBride, Ying Wang, and Leo Man-Lit Cheang. 2018. Dyslexia in Chinese. *Current Developmental Disorders Reports* 5, 4 (2018), 217–225.

[35] Eric Muller. 2012. L2/12-081 Extensions to Ideographic Description Sequences, take 2. http://unicode.org/L2/L2012/12081-ids.html

[36] Roderick I Nicolson and Angela J Fawcett. 1999. Developmental dyslexia: the role of the cerebellum 1. *Dyslexia* 5, 3 (1999), 155–177.

[37] Eraldo Paulesu, J-F Démonet, Ferruccio Fazio, Eamon McCrory, Valerie Chanoine, Nicky Brunswick, Stefano F Cappa, Giuseppe Cossu, Michel Habib, Chris D Frith, et al. 2001. Dyslexia: Cultural diversity and biological unity. *Science* 291, 5511 (2001), 2165–2167.

[38] Stratigoula Politi-Georgousi and Athanasios Drigas. 2020. Mobile Applications, An Emerging Powerful Tool for Dyslexia Screening and Intervention: A Systematic Literature Review. *International Journal of Interactive Mobile Technologies (iJIM)* 14, 18 (Nov. 2020), 4–17. https://doi.org/10.3991/ijim.v14i18.15315 Number: 18.

[39] Maria Rauschenberger, Ricardo Baeza-Yates, and Luz Rello. 2019. Technologies for dyslexia. In *Web Accessibility*. Springer, London, 603–627. https://doi.org/10.1007/978-1-4471-7440-0_31

[40] Maria Rauschenberger, Ricardo Baeza-Yates, and Luz Rello. 2020. Screening Risk of Dyslexia through a Web-Game Using Language-Independent Content and Machine Learning. In *Proceedings of the 17th International Web for All Conference* (Taipei, Taiwan) *(W4A '20)*. Association for Computing Machinery, New York, NY, USA, Article 13, 12 pages. https://doi.org/10.1145/3371300.3383342

[41] Maria Rauschenberger, Luz Rello, Ricardo Baeza-Yates, and Jeffrey P. Bigham. 2018. Towards Language Independent Detection of Dyslexia with a Web-Based Game. In *Proceedings of the 15th International Web for All Conference* (Lyon, France) *(W4A '18)*. Association for Computing Machinery, New York, NY, USA, Article 17, 10 pages. https://doi.org/10.1145/3192714.3192816

[42] Luz Rello, Miguel Ballesteros, Abdullah X Ali, Miquel Serra, Daniela Alarcón Sánchez, and Jeffrey P Bigham. 2016. Dytective: diagnosing risk of dyslexia with a game. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '16)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 89–96.

[43] Luz Rello, Enrique Romero, Maria Rauschenberger, Abdullah Ali, Kristin Williams, Jeffrey P Bigham, and Nancy Cushen White. 2018. Screening dyslexia for English using HCI measures and machine learning. In *Proceedings of the 2018 international conference on digital health (DH '18)*. Association for Computing Machinery, New York, NY, USA, 80–84. https://doi.org/10.1145/3194658.3194675

[44] Z Rezvani, M Zare, G Žarić, M Bonte, J Tijms, MW Van der Molen, and G Fraga González. 2019. Machine learning classification of dyslexic children based on EEG local network features. *BioRxiv* (2019), 569996.

[45] Sara Rosenblum, Shula Parush, and Patrice L Weiss. 2003. Computerized temporal handwriting characteristics of proficient and non-proficient handwriters. *The American Journal of Occupational Therapy* 57, 2 (2003), 129–138.

[46] Gerd Schulte-Körne. 2010. The prevention, diagnosis, and treatment of dyslexia. *Deutsches Arzteblatt international* 107 41 (2010), 718–26.

[47] scikit learn. 2023. Feature importance for gradient boosting classifier, scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html#sklearn.ensemble.GradientBoostingClassifier.feature_importances_

[48] Hua Shu, Catherine McBride-Chang, Sina Wu, and Hongyun Liu. 2006. Understanding Chinese developmental dyslexia: morphological awareness as a core cognitive construct. *Journal of educational psychology* 98, 1 (2006), 122.

[49] Margaret J Snowling. 2013. Early identification and interventions for dyslexia: a contemporary view. *Journal of Research in Special Educational Needs* 13, 1 (2013), 7–14.

[50] Maritzol Tenemaza, Rosa Navarrete, Erika Jaramillo, and Andrés Rodriguez. 2019. Specific Dyslexia Exploratory Test (TEDE): Two Tasks Using Augmented Reality. In *Advances in Usability, User Experience and Assistive Technology*, Tareq Z. Ahram and Christianne Falcão (Eds.). Springer International Publishing, Cham, 925–933.

[51] Mei Hui Tseng. 1998. Development of pencil grip position in preschool children. *The Occupational Therapy Journal of Research* 18, 4 (1998), 207–224.

[52] Chi-man Tsui, Cecilia W. P. Li-Tsang, and Pui Yee Grace Lung. 2012. *Dyslexia in Hong Kong: Challenges and Opportunities*. IntechOpen, Hong Kong SAR, China. https://doi.org/10.5772/32929 Publication Title: Learning Disabilities.

[53] Ivan Vajs, Vanja Ković, Tamara Papić, Andrej M Savić, and Milica M Janković. 2022. Dyslexia detection in children using eye tracking data based on VGG16 network. In *2022 30th European Signal Processing Conference (EUSIPCO)*. IEEE, 1601–1605.

[54] Aryan van der Leij. 2013. Dyslexia and early intervention: what did we learn from the Dutch Dyslexia Programme? *Dyslexia* 19, 4 (2013), 241–255.

[55] FungKa Yan, Fung Kwong Chiu, Chan Aidan, and Yu Yi Ching. 2021. A Digital Tool to Provide Pre-Screening to Dyslexia in Hong Kong. In *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*. IEEE, Los Alamitos, CA, USA, 755–761.

[56] Chunxia Zhang, Longxue Li, and Xudong Li. 2021. A Survey of Chinese Character Recognition Research Based on Deep Learning. In *2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC)*. IEEE, USA, 926–931.

[57] Jianshu Zhang, Jun Du, and Lirong Dai. 2020. Radical analysis network for learning hierarchies of Chinese characters. *Pattern Recognition* 103 (2020), 107305.

[58] Shuhan Zhong, Sizhe Song, Guanyao Li, and S.-H. Gary Chan. 2022. A Tree-Based Structure-Aware Transformer Decoder for Image-To-Markup Generation. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), Oct. 10–14, 2022, Lisboa, Portugal*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3503161.3548424