Design an Intelligence System for Early Identification on Developmental Dyslexia of Chinese Language

Man-Ching Yuen[1][a], Ka-Fai Ng[2], Ka-Ming Lau[1], Chun-Wing Lam[1] and Ka-Yin Ng[1]
[1]*iFREE GROUP Innovation and Research Centre, Department of Applied Data Science, Hong Kong Shue Yan University, China*
[2]*Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, China*

Keywords: Dyslexia, Traditional Chinese Character, App, Early Identification, Cloud.

Abstract: People with dyslexia have difficulties in fluently reading and writing characters which highly affect their learning progress. It is very important to identify dyslexic students that need intervention and extra support during their childhood. However, the waiting time for dyslexia assessment services is often long. To address the above problem, we propose a cloud-based early identification system of dyslexia. We design and develop a mobile app with AWS cloud platform as server. We have identified 27 representative traditional Chinese characters for handwriting data collection. After the first round of the data collection, 66 children aged 5-7 were recruited in Hong Kong. We carry out K-means clustering algorithm to investigate the characteristics of data points on a feature map for each character. We find out that some Chinese words contain more distinguishable characteristics for identifying children with dyslexia. Since children aged 5-7 are still learning how to write traditional Chinese characters properly, children with no risk of dyslexia still have certain possibilities of writing characters with characteristics of handwriting by children with dyslexia. It increases the difficulties of the early identification on developmental dyslexia of Chinese language. Finally, we present our findings and future work.

1 INTRODUCTION

Dyslexia is a kind of life-span learning disability disorder (Wang, J. and Perez, L., 2017). People with dyslexia have difficulties in fluently reading and writing characters which highly affect their learning progress. Tsui et al. showed that Hong Kong primary school students, aged 6-12, with dyslexia issue write significantly slower and inaccurate (Tsui, C.M., Li-Tsang, W.P.C. and Lung, P.Y., 2012). However, Dyslexia is not an uncommon symptom. Based on an earlier study (Chan, D.W., Ho, C.S.H., Tsang, S.M., Lee, S.H. and Chung, K.K., 2007), the prevalence rate of dyslexia in Hong Kong was 9.7% (6.2% mild severity, 2.2% moderate and 1.3% severe). In 2013, Sprenger-Charolles et al. suggested that around 17% of the world's population experience dyslexia (Sprenger-Charolles, L., Colé, P. and Serniclaes, W., 2013).

It is very important to identify dyslexic students that need intervention and extra support during their childhood. However, the waiting time for dyslexia assessment services is often long. Besides, traditional psychological diagnosis consumes much time and resources. For example, the assessment provided by the British Dyslexia Association takes up to 3 hours (Asvestopoulou, T., Manousaki, V., Psistakis, A., Smyrnakis, I., Andreadakis, V., Aslanides, I.M. and Papadopouli, M., 2019). It is necessary to have an early identification system of dyslexia for teachers and parents, so that intervention can be provided as soon as possible.

As machine learning has emerged into daily life, researchers have applied different machine learning models to analyse any specific patterns from behaviors of dyslexic children especially for their handwriting images. Various approaches have been conducted to detect dyslexia using machine learning, including eyeball movement tracking (Biswas, A. and Islam, M.S., 2021), handwriting motion and pressure (Isa, I.S., Rahimi, W.N.S., Ramlan, S.A. and Sulaiman, S.N., 2019; Košak-Babuder, M., Kormos,

[a] https://orcid.org/0000-0003-2551-7746

Yuen, M., Ng, K., Lau, K., Lam, C. and Ng, K.
Design an Intelligence System for Early Identification on Developmental Dyslexia of Chinese Language.
DOI: 10.5220/0011281500003286
In *Proceedings of the 19th International Conference on Wireless Networks and Mobile Systems (WINSYS 2022)*, pages 46-52
ISBN: 978-989-758-592-0; ISSN: 2184-948X

J., Ratajczak, M. and Pižorn, K., 2019), neuroimages for biomarkers (Lam, S.S., Au, R.K., Leung, H.W. and Li-Tsang, C.W., 2011), and brain images (Usman, O.L. and Muniyandi, R.C., 2020). Although the biomarkers are trackable, the assessment equipment is expensive and the assessment process is hard to scale up. These solutions are not suitable for common usage at home or school.

There were studies conducted on convolution analysis on handwritings of English, Spanish (Drotár, P. and Dobeš, M., 2020) and Indian students (Mahone, E.M. and Schneider, H.E., 2012). However, there are only a few research works related to Chinese characters, especially traditional Chinese characters that are more difficult to analyse compared with simplified Chinese characters. Tseng showed that traditional Chinese characters contain sharp turns and frequent pen lifts in which the symptoms should be more critical (Tseng, M. H., 1998).

To address the above problems, we design and develop a cloud-based system for early identification of dyslexia, where machine learning methodology is adopted to identify dyslexia involving traditional Chinese characters. The main contributions are:

- To design and develop a mobile app with AWS cloud platform as server. Our framework can support real-time performance evaluation on children handwriting wherever the number of concurrent testers increases
- To identify 27 representative traditional Chinese characters which are commonly taught in kindergartens or training centers in Hong Kong for data analysis experiments
- To collect the handwritings of 27 traditional Chinese characters from 66 children, where 25 have dyslexia and 41 do not have dyslexia
- To carry out some preliminary experiments to investigate the characteristics of the handwritten character images

The organization of this paper is as follows. Section 2 presents the related work. Section 3 describes our proposed dyslexia identification system. Section 4 shows the preliminary experimental result analysis. Section 5 draws out the conclusion and the future work.

2 RELATED WORKS

Various approaches have been conducted to detect dyslexia using machine learning. Thomais et al used an eye-ball tracker to analyze the eyeball movement (Biswas, A. and Islam, M.S., 2021) with the highest accuracy of 89.39%. Several groups learnt features from handwriting motion and pressure (Isa, I.S., Rahimi, W.N.S., Ramlan, S.A. and Sulaiman, S.N., 2019; Košak-Babuder, M., Kormos, J., Ratajczak, M. and Pižorn, K., 2019) showed promising results. Some others applied CNN on neuroimages for biomarkers and achieved accuracies of 73.2% (Lam, S.S., Au, R.K., Leung, H.W. and Li-Tsang, C.W., 2011). Another similar research analyzed brain images while students were reading and resulted in an accuracy of 72.73% (Usman, O.L. and Muniyandi, R.C., 2020). A research in Malaysia tried to increase the performance by applying the result OCR with a 73.77% accuracy. However, these solutions are not suitable for common usage at home or school.

The above methods showed a very promising result. However, these methods take too much time and resources to sample a candidate. To solve this, researchers studied detecting dyslexia via handwriting. Xing et.al. showed it is reliable to distinguish writers with handwritings using convolutional neural networks (CNN), and the proposed work, DeepWriter, achieved 99.01% on 301 writers and 97.03% on 657 writers (Xing, L. and Qiao, Y., 2016). Later, several researchers studied whether writers have dyslexia with a similar approach. Spoon et.al. gathered students' English and Spanish exercise books and applied CNN with Keras to detect dyslexia. They achieved an accuracy of 77.6% with 1200 samples from K-6 students (Spoon, K., Crandall, D. and Siek, K., 2019). Yogarajah et.al. conducted a similar research for Hindi characters achieving 86.14% (Yogarajah, P. and Bhushan, B., 2020). Optical character recognition (OCR) with Artificial neural network (ANN) focused on analysing 8 characters and achieved a test accuracy of 57.5% (Wei, P., Li, H. and Hu, P., 2019).

3 DYSLEXIA IDENTIFICATION SYSTEM

3.1 Overview of System Design

In this paper, we present an early identification system for detecting dyslexia with traditional Chinese characters, which is a cloud-based AI system. For children doing the assessment test, parents have to print out the worksheets and ask their children to write the traditional Chinese characters on the worksheets. After completion, parents have to take pictures of the worksheets, crop the images, and upload the cropped images to the cloud system for data analysis. Figure 1 shows the system architecture of the early identification system of dyslexia. We develop our machine learning by using AWS SageMaker for data analysis.

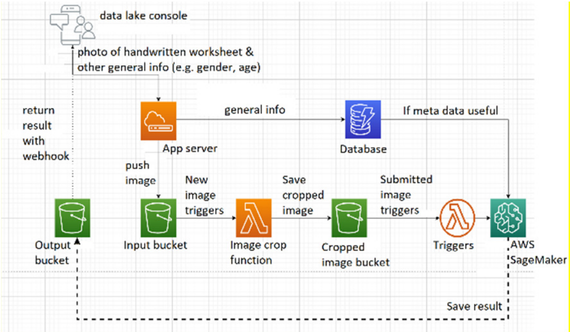


Figure 1: System architecture of the early identification system of dyslexia.

3.2 Mobile App Design

Figure 2 shows the web app interface in the web browser. Since the user interface is in traditional Chinese, we add some description in English. The functions are providing user guide, downloading worksheet, uploading worksheet, setting and previewing previous testing results. Since the duration of children handwriting is also a performance indicator of dyslexia, the mobile app allows parents to collect the time spent on handwriting of their children by either using a system timer or inputting the value manually (as shown in Figure 3 and Figure 4). Before parents take pictures and upload each of the 3 worksheets, parents have to scan the QR code in the left corner of the worksheet to make sure they are uploading the correct worksheets (as shown in Figure 5). When the image taken is shown on the mobile app, parents can crop

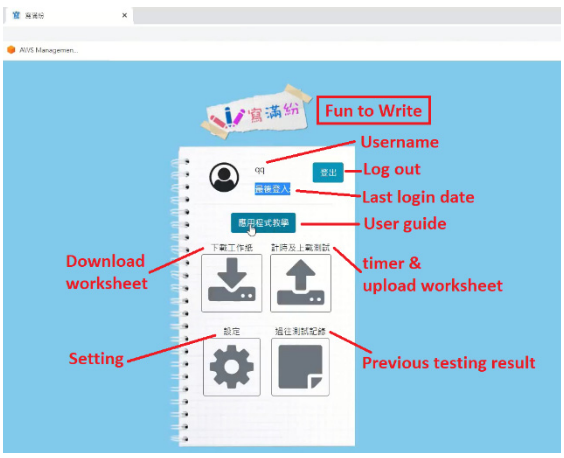


Figure 2: The web app interface in the web browser.

the image by adjusting the little yellow boxes at each corner (as shown in Figure 6). The analysis carried out by the cloud AI system indicates the risk of having dyslexia for testers (as shown in Figure 7).

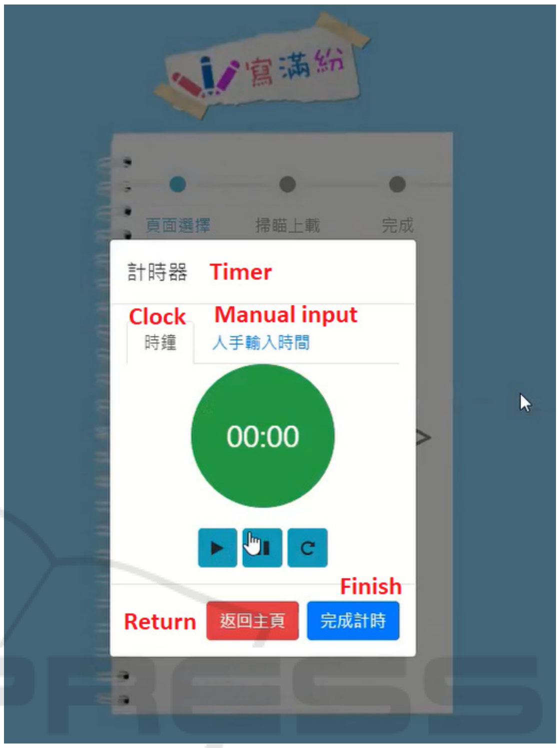


Figure 3: The mobile app can collect the time spent on handwriting by using a system timer.

Figure 4: The mobile app can collect the time spent on handwriting by inputting the value manually.

Figure 5: Scan the QR code in the left corner of the worksheet to make sure parents are uploading the correct worksheets.

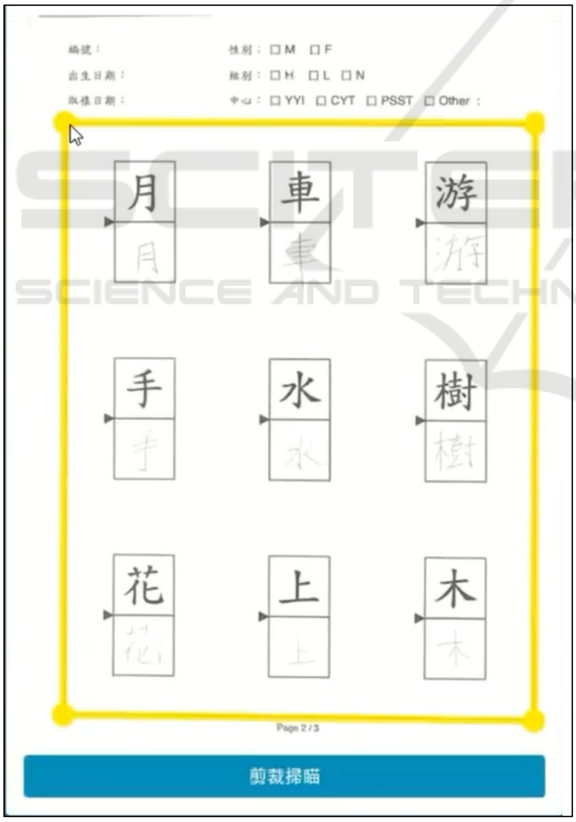


Figure 6: Parents can crop the image by adjusting the little yellow boxes at each corner.

Figure 7: The analysis result indicates the risk of having dyslexia for the tester.

Figure 8: Project promotion website.

4 EXPERIMENTS

4.1 Participants and Tasks

To collect handwriting for model training, we plan to invite about 30 children with dyslexia and 200 children without dyslexia from kindergartens. Handwriting data is planned to be collected periodically every 6-9 months. After the first round of the data collection, a total of 66 children (aged 5-7) were recruited from several kindergartens or training centers of Yan Chai Hospital Social Services Department (YCH) in Hong Kong. The children do not have physical or mental disabilities which might affect the handwriting performance and lead to bias in the data collected for model training. Informed consent of parents was obtained for all children. In order to raise the awareness of dyslexia among teachers and parents, we develop a website for project promotion as shown in Figure 8.

We use the following 4 steps to carry out pre-screening on the participating children in order to distinguish whether they are dyslexia or not.

1. Teachers will screen all participating children's performance on writing and reading traditional Chinese words during class. Teachers will screen out the students who are not suitable to participate in the experiment.
2. Occupational Therapists (OTs) will carry out Visual Perceptual (VP) assessment (only for suspected cases of dyslexia).
3. Educational Psychologists (EPs) / Clinical Psychologists (CPs) will carry out formal assessments of dyslexia (only for suspected cases of dyslexia in order to determine the level of dyslexia).
4. The children are categorized into 3 groups:
(1) H – High risk of having dyslexia;
(2) L – Low risk of having dyslexia;
(3) N – No risk of having dyslexia / Normal.

Table 1 summarizes the age and gender for children in the 3 groups, which are normal, low and high risk of having dyslexia respectively.

Table 1: Age and gender of participating children.

	High	**Low**	**No**
Number of children	19	6	41
Mean age (months)	67.2	69.0	69.9
Gender (girls vs. boys)	4 vs. 15	1 vs. 5	8 vs. 32

Figure 9 shows the 27 representative traditional Chinese characters in the handwriting collection. The 27 characters are commonly taught in kindergartens or training centers in Hong Kong. The words cover six 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 to ensure that the selected Chinese characters are sufficiently representative of the characteristics of traditional Chinese. The idea on how to select the words for data collection is also inspired by Wu et al. presented in 2019 (Wu, Z., Lin, T., and Li, M., 2019). The display sequences of the words across 3 worksheets were randomized regardless of the level of writing difficulties and the structure of words.

We design a special set of worksheets for easier and better quality sampling. A set of 3-page worksheets contain the 27 representative traditional Chinese characters as shown in Figure 10. On the worksheet, each character is featured with a printed character alongside and great margin. The printed characters provide direct reference to the students to maintain quality of the written text, especially for normal students. Besides, the margin is designed to capture all strokes and details in case students write out of the box. The dimension of the box is 2.5cm x 2.5cm which is the same as the children learning handwriting at the childhood at age 5-7 in Hong Kong.

丁	草	狗	月	車	游	男	魚	山
媽	王	田	手	水	樹	日	家	和
學	士	好	花	上	木	女	火	牙

Figure 9: The 27 traditional Chinese characters used in the experiment.

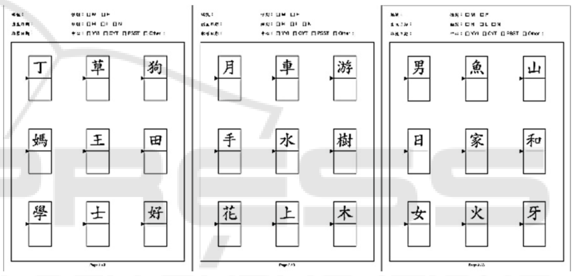


Figure 10: A set of 3-page worksheets contain the 27 representative traditional Chinese characters.

4.2 Observation from Screening

Based on our observation, for each traditional Chinese character, some handwriting characters have characteristics of handwriting by children with dyslexia but these characters are written by children without dyslexia. Moreover, children of high risk of dyslexia have higher possibility of writing characters with characteristics of handwriting by children with dyslexia; while children of no risk of dyslexia still have certain possibilities. The main reason behind this is children aged 5-7 are still learning how to write traditional Chinese characters properly.

For another finding, some handwriting characters do not have characteristics of handwriting by children with dyslexia but these characters are written by children with dyslexia. It is because writing abilities in children with dyslexia can be improved by training. It demonstrates the difficulties and importance of this project.

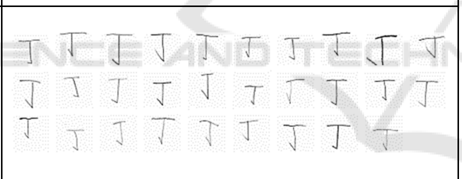
4.3 Preliminary Experimental Results

We carry out K-means clustering algorithm for each traditional Chinese character by setting k as 3, plot all data points on a 2D feature map. We summarize our observations from the feature maps as follows.

First, we find that data points representing images not having characteristics of handwriting by children with dyslexia are usually in the same cluster. Use character "丁" (Ding) as an example. The result of character "丁" (Ding) are shown in Table 2 and Figure 11. In Table 2, all data points in the cluster 0 do not have characteristics of handwriting by children with dyslexia. Figure 11 shows the cluster distribution of character "丁" (Ding) in the feature map.

Second, some words contain more distinguishable characteristics for identifying children with dyslexia, such as "學" (Learn), "游" (Swim), "和" (And), while some are not such as "樹" (Tree), "上" (Up). Figure 12 and Figure 13 show the cluster distribution of character "和" (And) and "上" (Up) in the feature maps respectively.

Table 2: Image of character "丁" (Ding) in 3 clusters.

Cluster	Sample
0	
1	
2	

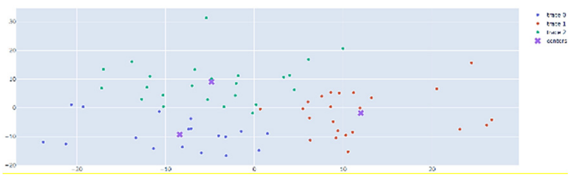


Figure 11: Cluster distribution of character "丁" (Ding) in the feature map.

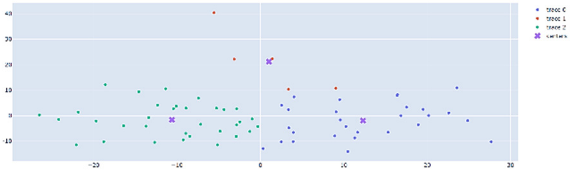


Figure 12: Cluster distribution of character "和" (And) in the feature map.

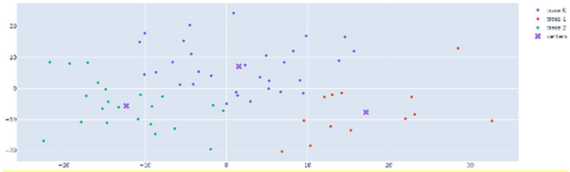


Figure 13: Cluster distribution of character "上" (Up) in the feature map.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a cloud-based early identification system of dyslexia. We have designed and developed a mobile app with AWS cloud platform as server. For data collecting for model training, we plan to invite about 30 children with dyslexia and 200 children without dyslexia from kindergartens. After the first round of the data collection, 66 children aged 5-7 were recruited in Hong Kong. We have identified 27 representative traditional Chinese characters for handwriting data collection. We have carried out K-means clustering algorithm to investigate the characteristics of data points on a feature map for each character.

In the future, we will design the machine learning model of the identification system. Besides, we will continue recruiting children to collect handwriting for model training. Moreover, we consider oversampling by applying data augmentation to enlarge the data size and also look into different techniques to generate more reliable samples for model training.

ACKNOWLEDGEMENTS

This research was in part supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. UGC/FDS15/E02/20).

Besides, the authors would like to thank the research team from Rehabilitation Services (Early Education & Training) from Yan Chai Hospital

Social Services Department in Hong Kong for their professional suggestions and their time for carrying out experiments. Besides, the authors would like to thank the participating children and their parents who volunteered their time and valuable feedback on this study.

REFERENCES

Asvestopoulou, T., Manousaki, V., Psistakis, A., Smyrnakis, I., Andreadakis, V., Aslanides, I.M. and Papadopouli, M. (2019). Dyslexml: Screening tool for dyslexia using machine learning. arXiv preprint arXiv:1903.06274.

Biswas, A. and Islam, M.S. (2021). An Efficient CNN Model for Automated Digital Handwritten Digit Classification. Journal of Information Systems Engineering and Business Intelligence, 7(1), pp.42-55.

Chan, D.W., Ho, C.S.H., Tsang, S.M., Lee, S.H. and Chung, K.K. (2007). Prevalence, gender ratio and gender differences in reading-related cognitive abilities among Chinese children with dyslexia in Hong Kong. Educational Studies, 33(2), pp.249-265.

Drotár, P. and Dobeš, M. (2020). Dysgraphia detection through machine learning. Scientific reports, 10(1), pp.1-11.

Isa, I.S., Rahimi, W.N.S., Ramlan, S.A. and Sulaiman, S.N. (2019). Automated Detection of Dyslexia Symptom Based on Handwriting Image for Primary School Children. Procedia Computer Science, 163, pp.440-449.

Košak-Babuder, M., Kormos, J., Ratajczak, M. and Pižorn, K. (2019). The effect of read-aloud assistance on the text comprehension of dyslexic and non-dyslexic English language learners. Language Testing, 36(1), pp.51-75.

Lam, S.S., Au, R.K., Leung, H.W. and Li-Tsang, C.W. (2011). Chinese handwriting performance of primary school children with dyslexia. Research in developmental disabilities, 32(5), pp.1745-1756.

Mahone, E.M. and Schneider, H.E. (2012). Assessment of attention in preschoolers. Neuropsychology review, 22(4), pp.361-383.

Spoon, K., Crandall, D. and Siek, K. (2019). Towards Detecting Dyslexia in children's handwriting using neural networks. In Proceedings of the International Conference on Machine Learning AI for Social Good Workshop, Long Beach, CA, USA (pp. 1-5).

Sprenger-Charolles, L., Colé, P. and Serniclaes, W. (2013). Reading acquisition and developmental dyslexia. Psychology Press.

Tseng, M. H. (1998). Development of pencil grip position in preschool children. Occupational Therapy Journal of Research, 18, 207-224.

Tsui, C.M., Li-Tsang, W.P.C. and Lung, P.Y. (2012). Dyslexia in Hong Kong: challenges and opportunities. InTech.

Usman, O.L. and Muniyandi, R.C. (2020). CryptoDL: Predicting Dyslexia Biomarkers from Encrypted Neuroimaging Dataset Using Energy-Efficient Residue Number System and Deep Convolutional Neural Network. Symmetry, 12(5), p.836.

Wang, J. and Perez, L. (2017). The effectiveness of data augmentation in image classification using deep learning. Convolutional Neural Networks Vis. Recognit, 11, pp.1-8.

Wei, P., Li, H. and Hu, P. (2019). Inverse discriminative networks for handwritten signature verification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Wu, Z., Lin, T., and Li, M. (2019). Automated Detection of Children at Risk of Chinese Handwriting Difficulties Using Handwriting Process Information: An Exploratory Study. IEICE TRANSACTIONS on Information and Systems, 102(1), 147-155.

Xing, L. and Qiao, Y. (2016). October. Deepwriter: A multi-stream deep CNN for text-independent writer identification. In 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR) (pp. 584-589). IEEE.

Yogarajah, P. and Bhushan, B. (2020). Deep Learning Approach to Automated Detection of Dyslexia-Dysgraphia. In The 25th IEEE International Conference on Pattern Recognition.