Deep Learning Approach to Automated Detection of Dyslexia-Dysgraphia

Pratheepan Yogarajah[1] [0000−0002−4586−7228] and Prof. Braj Bhushan[2] [0000−0001−5177−1209]

[1] School of Computing, Engineering and Intelligent Systems, Ulster University, Northern Ireland, UK
p.yogarajah@ulster.ac.uk

[2] Department of Humanities & Social Sciences, Indian Institute of Technology (IIT), Kanpur, India
brajb@iitk.ac.in

Abstract. Owing to the concern regarding lack of reliance on cognitive profile-based diagnosis of dyslexia, we conducted an automated diagnosis exclusively based on the handwriting sample. Handwriting sample of 54 children (36 males, 18 females) studying between classes 1^{st} to 5^{th} identified with 'strong evidence of risk' on Dyslexia Screening Test-Junior (DST-J) was collected. 14 Hindi words (i.e. 5 two letter words, 6 words with Matras (vowel signs), and 3 conjoined consonants (Sanyukt Akshar)) were selected for this study on the basis of graded difficulty level. These Hindi words share features such as matras, killer strokes (halants), and Sanyukt Akshar. A total of 267 images, 164 from children with dyslexia-dysgraphia and 103 from age-matched normal control, were collected for this study. These images were resized to a fixed height of 113 pixels along with different width sizes depending on the image aspect ratio. A random number of patches with size of 113×113 pixels were generated from each image. A Convolutional Neural Network (CNN) using Keras and Tensorflow was successful in automatically identifying powerful features with average accuracy of $(86.14 \pm 1.02)\%$. The findings endorse deep learning approach in automated detection/diagnosis of dyslexia-dysgraphia.

Keywords: Dyslexia · Dysgraphia · Convolutional Neural Network · Hindi Language.

1 Introduction

Learning disabilities, especially dyslexia and dysgraphia, are fairly common in schools worldwide. It is characterized by difficulty in age appropriate reading in the absence of any other biological or socio-cultural deprivation. Such children have problem making patterns of words and letters presented through visual mode. They have significant difficulty with speed and accuracy of decoding word, spelling and text comprehension. Irrespective of the intact phonetics and semantics they also have problem in correctly identifying the written word and this leads to dysgraphia. As a result, they repeatedly use compensatory technique, substitution, mirror writing, reversals, omissions, inversions, unclear fonts, inappropriate formation, etc.

Although the common man's understanding of dyslexia restricts it to a reading disorder, there is enough evidence to endorse it as a problem in writing skills as well. Dysgraphia is described as a handwriting learning disability associated with dyslexia. It is also associated with dyspraxia, a disorder of developmental coordination, and attention deficit. All of this fall under the ambit of neurodevelopmental disorders. Deuel [9] has classified dysgraphia into three sub-types - dyslexic dysgraphia, spatial dysgraphia, and motor dysgraphia. The existing tests of dysgraphia depends on analysis of handwriting and tests such as the handwriting proficiency screening questionnaire [17], the scoring of which relies on human judgement making it highly subjective and dependent on the availability of trained human resource.

Several researchers have reported writing problems in children [4], college students [8], and adults [5]. Berninger et al. [4] found almost equal number of indicators of reading as well as writing problems in children and adults with dyslexia. As reading is theorized as a central component of writing [10], the difficulties recorded in children with dyslexia can be due to reading difficulties to certain extent. Reading and writing, both depend on interrelated processes inasmuch as the difficulty in processing phonological information affects decoding words and encoding of phonological information is needed while writing [11], [12]. This might have a bearing on poor handwriting skills of such children.

The conventional diagnosis of dyslexia involves psychological assessment of cognitive abilities. Based on the administration of some standardized tests, the psychologists quantify the reasoning capacity of the child in order to exclude the possibility of reading disability due to mental deficiency. This is important as studies have found significant gap of 18 months between reading and school levels in children below 9 years of age [6]. Large number of studies have administered WISC-III and WISC-R for assessing the cognitive abilities of children with dyslexia. Bannatyne [2] classified WISC-R subtests into three categories— spatial, conceptual, and sequential abilities to further suggest children with learning disabilities have higher spatial abilities than conceptual abilities, and least sequential abilities. Berk [3] has raised concerns about WISC-R profile to diagnose specific learning disabilities. On the other hand, Grégoire [13] found 15.4% false positive and 64.3% false negative rates of diagnosis for specific learning disabilities, thus indicating severe limitation with Bannatyne's profile. The inconsistency in IQ profile of children with dyslexia is obvious in the recent research and there is a shift away from diagnosis of specific learning disability based on cognitive profile, especially the assessment based on WISC-R and WISC-III. According to Clercq-Quaegebeur [7], "WISC-III failed to clearly identify typical profiles and cognitive deficits in dyslexia".

The recent trend of research shows an inclination towards automated diagnosis of dysgraphia based on the handwriting sample [1], [15], [16]. In their attempt to diagnose dysgraphia, Asselborn et al.[1] extracted 53 handwriting features in the handwriting sample of 56 children with dysgraphia and using the Random Forest classifier they could achieve it with 96.6% sensibility and 99.2% specificity. Spoon et al. [18] developed a system that used computer vision and deep learning to classify handwriting samples as indicative of dyslexia or not. Using five-fold cross validation, they obtained an average accuracy of $55.7 \pm 1.4\%$. This accuracy is very low for an automated diagnosis system.

The review of literature flags two distinct issues pertaining to diagnosis of dyslexia-dysgraphia - lack of reliance on cognitive profile based diagnosis of dyslexia and reliable automated diagnosis exclusively based on the handwriting sample. Further, the studies reporting machine learning outcome for successful automated diagnosis have conducted their study on English and European languages. India has approximately 140 million children studying in primary schools. Roughly 15% of them suffer from dyslexia and a substantial percentage of them have dysgraphia. As 40% of the Indian population speaks Hindi, the overall number of school-going children with dyslexia/dysgraphia is very large. This tells us the magnitude of the problem.

Although Hindi shares certain features with other languages, such as reading-writing from left to right, it has some unusual features. Diacritics (matras), killer strokes (halants), and conjoined consonants (Sanyukt Akshar) are some of them. The earlier work of Meena et al. [14] suggest the nature of errors committed by a child with dyslexia while typing Hindi text using a virtual keyboard interface with visual and auditory feedbacks. During our extensive literature search we did not find any study on automated diagnosis using Hindi alphabet. Considering this into account, the present study aims to explore if dyslexia-dysgraphia can be identified on the basis of handwriting sample of Hindi words or not. We also attempted to examine the complex shape/pattern of letters to identify the geometric shape that distinctively parses out dyslexia-dysgraphia in the Hindi letters. In this paper, Convolutional Neural Network (CNN) using Keras and Tensorflow is applied to identify dyslexia-dysgraphia on the basis of handwriting sample of Hindi words.

The research paper is organized as follows: section 2 provides the details of the data collection process and the implementations methodologies, section 3 shows the results and section 4 discusses the outcomes and the future direction.

2 Our Approach

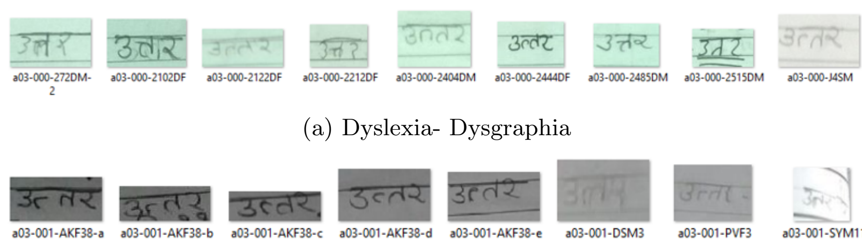
The participants selection and the handwriting images collections are initial steps of our approach. The following sections provide the details of participants selection and the data collection.

2.1 Participants Selection

At the initial level, language teachers were requested to identify supposedly academically poor students in language, both Hindi and English, compared to others in their class. This initial screening was done for children studying from classes 1st to 5th with perceived academic difficulty and poor scholastic record. All the students identified by the language teachers were administered Dyslexia Screening Test-Junior (DST-J) by Angela Fawcett and Rod Nicolson. The score on DST-J ranged between .90 and 3.09 (Mean 1.93, SD .54). On the basis of DST-J scores, 54 children (36 males, 18 females) were finally selected for this study. The mean age of the participants was 8.39 years (SD 1.43). Only those identified with 'strong evidence of risk' were included in the study. Any child with dyscalculia or any other deficit were excluded. These were the respective inclusion and exclusion criteria. The study was approved by the Institutional Ethics Committee of Indian Institute of Technology, Kanpur (protocol number IITK/IEC/2017-18 1/16).

2.2 Handwriting Sample

Owing to the sensitivity of the issue, we did not ask the participants to write any script. Instead, we collected handwriting sample from the Hindi notebooks of these 54 children and searched for common words/frequently appearing words and made an exhaustive list (Table 1).



(a) Dyslexia- Dysgraphia

(b) Normal

Fig. 1: A sample of handwriting images of one of the three sanyukta letters

Thereafter, we attempted identifying specific nature of pattern in these words, if any. Finally, 14 Hindi words were selected for this study - five two letter words, six words with matras (vowel signs), and three conjoined consonants (Sanyukt Akshar) words. These words were chosen as they represent graded level of difficulty. A total of 267 images, 164 of children with dyslexia-dysgraphia and 103 of aged matched normal control, were collected for this study.

2.3 Pre-processing

A sample of handwriting images of one of the three sanyukta letters (उत्तर) from the Dyslexia-Dysgraphia and the normal children is shown in Figure 1. It clearly indicates that there are significant intra class variations such as color, blurring, size, etc. in the collected images. To overcome the color issue, all the images are

Table 1: Database Details

No.	Description	Words	No. of images
1	Two Letters	एक	47
2	Two Letters	कर	15
3	Two Letters	पर	34
4	Two Letters	उस	6
5	Two Letters	यह	6
6	Two Letters & Matra	रहा	27
7	Two Letters & Matra	नहीं	28
8	Two Letters & Matra	कौन	12
9	Two Letters & Matra	मेरे	12
10	Two Letters & Matra	क्या	29
11	Two Letters & Matra	क्यों	18
12	Three sanyukta Letters	उत्तर	17
13	Three sanyukta Letters	बच्चा	9
14	Three sanyukta Letters	वाक्य	7
Total			267

converted to grayscale images. The grayscale images are resized to a fixed height of 113 pixels, along with different width sizes (depends on image aspect ratio). A random number of patches with size of 113×113 pixels are generated from each image. If the width is less than 113 pixels, then the image is resized to 113×113 pixels. In the case of width is greater than 113 pixels, a set of number of patches is generated based on the size of the width. A sample of random patches from a particular image is shown in the Figure 2.

In our method, we used visual features of the Hindi handwriting images from children for detecting Dyslexia-Dysgraphia. Our technique is based on an existing hand writing recognition approach [19],[20]. Using this approach, they tried to recognise different writers based on the handwriting images. We altered the approach to accept that there are only two type of writers - one with Dyslexia-Dysgraphia, and one without. Spoon et. al. [18] used the similar concept, and obtained an average accuracy of $55.7 \pm 1.4\%$ using [19], [20]. Therefore, in this paper, a modified (fine turned) version of [19],[20] is used (as shown in the Figure 3) to achieve a higher accuracy.

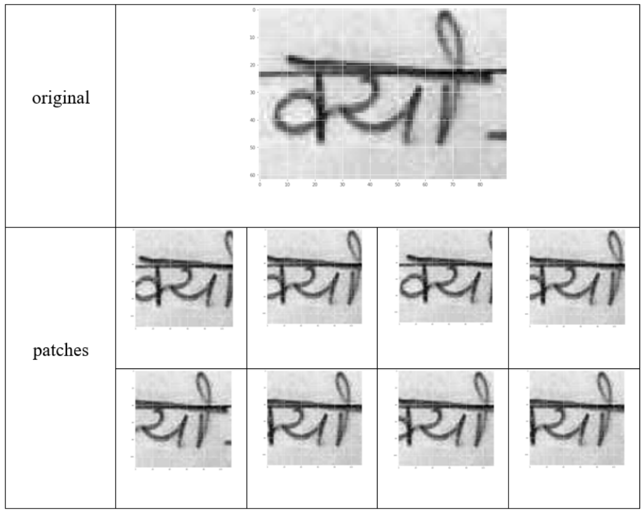


Fig. 2: A sample of random patches.

```
Model: "Our Proposed Model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
zero_padding2d_1 (ZeroPaddin (None, 115, 115, 1)       0
_________________________________________________________________
lambda_1 (Lambda)            (None, 56, 56, 1)         0
_________________________________________________________________
conv1 (Conv2D)               (None, 28, 28, 32)        832
_________________________________________________________________
activation_1 (Activation)    (None, 28, 28, 32)        0
_________________________________________________________________
pool1 (MaxPooling2D)         (None, 14, 14, 32)        0
_________________________________________________________________
conv2 (Conv2D)               (None, 14, 14, 64)        18496
_________________________________________________________________
activation_2 (Activation)    (None, 14, 14, 64)        0
_________________________________________________________________
pool2 (MaxPooling2D)         (None, 7, 7, 64)          0
_________________________________________________________________
conv3 (Conv2D)               (None, 7, 7, 128)         73856
_________________________________________________________________
activation_3 (Activation)    (None, 7, 7, 128)         0
_________________________________________________________________
pool3 (MaxPooling2D)         (None, 3, 3, 128)         0
_________________________________________________________________
flatten_1 (Flatten)          (None, 1152)              0
_________________________________________________________________
dropout_1 (Dropout)          (None, 1152)              0
_________________________________________________________________
dense1 (Dense)               (None, 16)                18448
_________________________________________________________________
activation_4 (Activation)    (None, 16)                0
_________________________________________________________________
dropout_2 (Dropout)          (None, 16)                0
_________________________________________________________________
output (Dense)               (None, 2)                 34
_________________________________________________________________
activation_5 (Activation)    (None, 2)                 0
=================================================================
```

Fig. 3: The proposed CNN model details.

2.4 Convolutional Neural Network

The Convolutional Neural Network (CNN) is applied to this application as it automatically identifies deep powerful features. In this application, the CNN is implemented using Keras and Tensorflow [21]. There are 3 convolutional layers, 3 max-pooling (MP) layers, 2 fully-connected (FC) layers and a output layer in the CNN. Figure 3 shows the details of the components used in our experiment.

It is not advisable to pass the entire dataset into the neural net at once due to the memory constraint. So, the dataset is divided into number of batches of size "batch size", and repeatedly iterating over the entire dataset for a certain number of "epochs" [21].

3 Results

The dataset is usually split into training data and test data. The training set contains a known output and the model learns on this data in order to be generalized to other data later on. The test data will be used in order to test the model's prediction.

However, split into training data and test data does have its dangers - What if one subset of the data has mostly people from Dyslexia-Dysgraphia? This will result in overfitting, even though shuffle operation is applied. Therefore, two approaches such as train/split and cross validation are considered in the experiments.

3.1 Train/Test Split

The dataset (i.e. total image patches) is split into Training, Validation and Testing as shown in Figure 4.

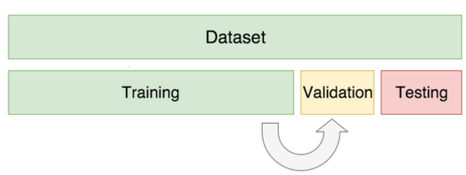


Fig. 4: Dataset partition for Training, Validation and Testing.

The total dataset is split into Training:Validation:Testing in a $16:4:5$ ratio. In the experiment, a variety of "batch size" $(1, 4, 16, 32)$ are applied. The highest accuracy obtained from the experiment with a batch size of 16. The "epochs" value is chosen as 50 for a better result.

The model is evaluated using a fixed trials (i.e. 10 times). Figure 8 shows the accuracy and loss values for each "epochs" value of a sample of 5 trials. For each trial, the dataset is shuffled prior to being split. These plots are valuable for getting an idea of the behaviour during the training and the validation of the model performance.

In each trial, the accuracy values are calculated for testing as 91.62%, 85.65%, 71.00%, 88.62%, 86.18%, 86.08%, 87.94%, 89.30%, 92.77% and 92.59% respectively. Next, the classification accuracy is summarized by calculating the mean and standard deviation. Based on the experiment results, an average accuracy of $(87.18 \pm 5.95)\%$ is achieved. The distribution of scores are showed using box and whisker plot [22] as shown in Figure 5.

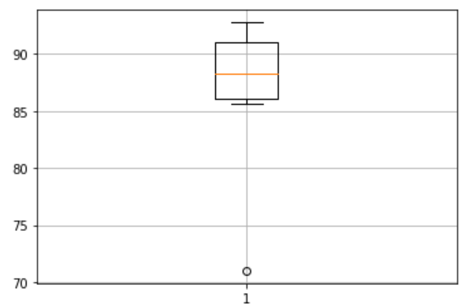


Fig. 5: Accuracy: mean = 87.18 std = 5.95.

3.2 Cross Validation

It is similar to train/test split, but it is applied to more subsets. First split the dataset into k subsets, and train on $k-1$ subsets, and hold the last subset for test. This is called K-Fold cross validation. The dataset (i.e. total image patches) is split into K-Fold cross validation as shown in Figure 6.

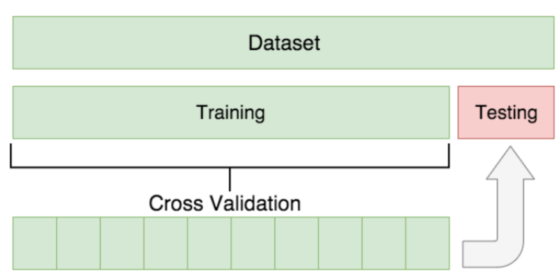


Fig. 6: Dataset partition for K-Fold cross validation.

In this experiment, a 5-Fold cross validation is used, i.e. $K = 5$. In each trial, the accuracy value is calculated for each round and the final accuracy is calculated using the average values of the 5 rounds.

A violin plot [23] is more informative than a plain box plot. In fact while a box plot only shows summary statistics such as mean/median and interquartile ranges, the violin plot shows the full distribution of the data. Therefore, a violin plot is used to show more informative details of each trial. Figure 7 shows the average accuracy values of randomly selected 5 trials.

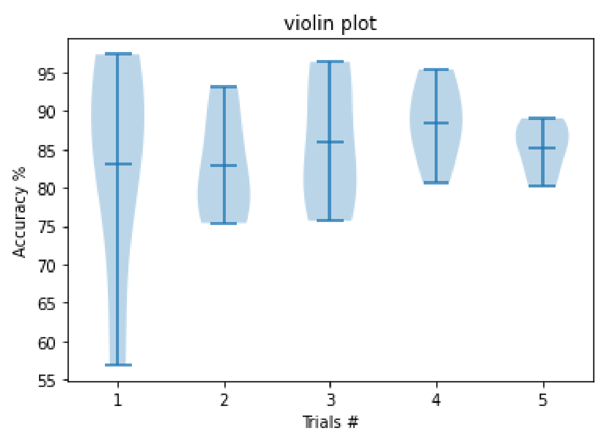
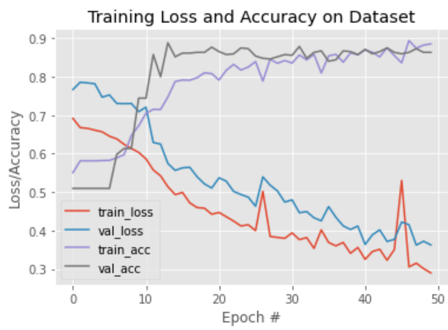


Fig. 7: The average accuracy values are 83.17%, 82.79%, 85.97%, 88.48% and 85.21% respectively.

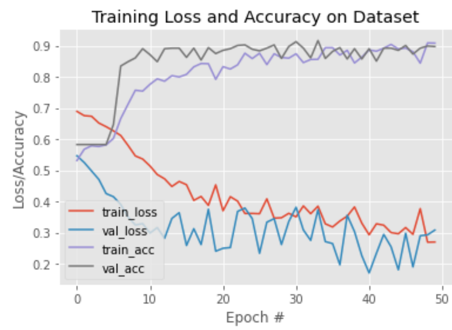
In each trial, the accuracy values are calculated for testing as 83.17%, 82.79%, 85.97%, 88.48% and 85.21% respectively. Next, the classification accuracy of cross validation based approach is summarized by calculating the mean and standard deviation. Based on the experiment results, an average accuracy of $(85.12 \pm 2.0)\%$ is achieved.

The overall accuracy of $(86.14 \pm 1.02)\%$ is computed using the accuracy values of 87.18% and 85.12% from the train/test split and the cross validation experiments. These results show a valuable evidence of detection of dyslexia based on handwriting images using image processing and deep learning technologies. Due to the prevalence of dyslexia the number of cases that go undiagnosed, particularly in elementary school.

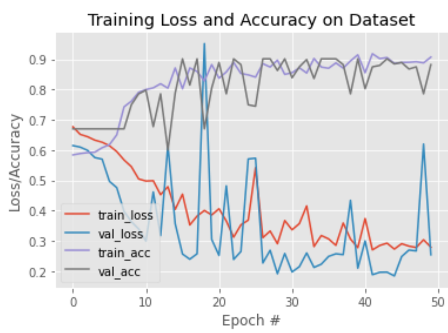
The dataset used in the experiments is small. More data is required to study the results further, especially more data from the students with dyslexia.



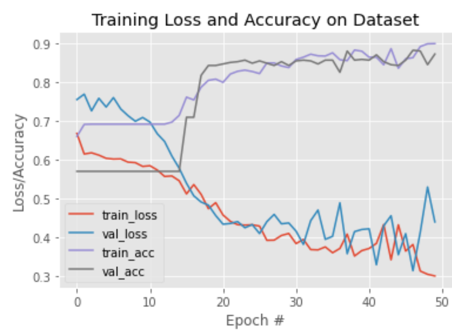
(a) Run #: 1



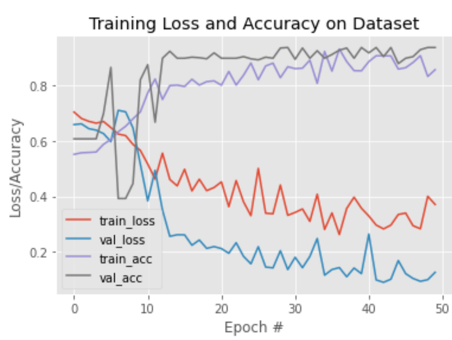
(b) Run #: 2



(c) Run #: 3



(d) Run #: 4



(e) Run #: 5

Fig. 8: Accuracy and Loss details of Training and Validation

4 Conclusion

This work presents an automated detection system to detect the presence of dyslexia symptoms in school children studying between classes 1^{st} to 5^{th} based on their handwriting images collected from their school note books. The automated detection system is developed by using Deep Learning technique (i.e. CNN) had shown significant results. In this work, the letters are cropped manually, for further improvement, handwritten recognition by using Optical Character Recognition (OCR) can be used. However, additional methods such as cursive and skew methods need to be incorporated as OCR is not stable to detect and recognise the handwritten characters due to the shape and style of the handwriting.

From the results of the collected data, it can be concluded that this work is able to detect the symptoms of dyslexia in children using the handwritten images, however for further improvement on detection of dyslexia symptom needed to consider other aspects, for example phonics. Compared to other studies this project has shown significant development of detect dyslexia using Hindi handwritten images.

Acknowledgment

This study was partly supported by Ministry of Human Resource and Development, Government of India under grant number MHRD/HSS/2016142S to the second author.

References

1. Asselborn, Thibault, Thomas Gargot, Łukasz Kidziński, Wafa Johal, David Cohen, Caroline Jolly, and Pierre Dillenbourg. "Automated human-level diagnosis of dysgraphia using a consumer tablet." NPJ digital medicine 1, no. 1 (2018): 1–9.
2. Bannatyne, Alexander. "Diagnosis: A note on recategorization of the WISC scaled scores." Journal of Learning Disabilities 7 (1974): 272–274.
3. Berk, Ronald A. "The value of WISC-R profile analysis for the differential diagnosis of learning disabled children." Journal of Clinical Psychology 39, no. 1 (1983): 133–136.
4. Berninger, Virginia W., Kathleen H. Nielsen, Robert D. Abbott, Ellen Wijsman, and Wendy Raskind. "Writing problems in developmental dyslexia: Under-recognized and under-treated." Journal of school psychology 46, no. 1 (2008): 1–21.
5. Bruck, Maggie. "Component spelling skills of college students with childhood diagnoses of dyslexia." Learning Disability Quarterly 16, no. 3 (1993): 171–184.
6. Cheminal, Renée, and Vincent Brun. Les dyslexies. No. 17. Elsevier Masson, 2002.
7. De Clercq-Quaegebeur, Maryse, Séverine Casalis, Marie-Pierre Lemaitre, Béatrice Bourgois, Marie Getto, and Louis Vallée. "Neuropsychological profile on the WISC-IV of French children with dyslexia." Journal of Learning Disabilities 43, no. 6 (2010): 563–574.

8. Connelly, Vincent, Sonya Campbell, Morag MacLean, and Jim Barnes. "Contribution of lower order skills to the written composition of college students with and without dyslexia." Developmental neuropsychology 29, no. 1 (2006): 175–196.
9. Deuel, Ruthmary K. "Developmental dysgraphia and motor skills disorders." Journal of Child Neurology 10, no. (1995): S6–S8.
10. Graham, Steve. "A revised writer (s)-within-community model of writing." Educational Psychologist 53, no. 4 (2018): 258–279.
11. Graham, Steve, and Michael Hebert. Writing to read: Evidence for how writing can improve reading. Washington, DC: Alliance for Excellent Education, 2010.
12. Graham, Steve, and Michael Hebert. "Writing to read: A meta-analysis of the impact of writing and writing instruction on reading." Harvard Educational Review 81, no. 4 (2011): 710–744.
13. Grégoire, Jacques. L'évaluation clinique de l'intelligence de l'enfant: Théorie et pratique du WISC-III. Vol. 229. Editions Mardaga, 2000.
14. Meena, Yogesh Kumar, Anirban Chowdhury, Ujjwal Sharma, Hubert Cecotti, Braj Bhushan, Ashish Dutta, and Girijesh Prasad. "A hindi virtual keyboard interface with multimodal feedback: a case study with a dyslexic child." In Proceedings of the 32nd International BCS Human Computer Interaction Conference 32, 2018, pp. 1–5.
15. Mekyska, Jiri, Marcos Faundez-Zanuy, Zdenek Mzourek, Zoltan Galaz, Zdenek Smekal, and Sara Rosenblum."Identification and rating of developmental dysgraphia by handwriting analysis." IEEE Transactions on Human-Machine Systems 47, no. 2 (2016): 235–248.
16. Pagliarini, Elena, Lisa Scocchia, Mirta Vernice, Marina Zoppello, Umberto Balottin, Sana Bouamama, Maria Teresa Guasti, and Natale Stucchi. "Children's first handwriting productions show a rhythmic structure." Scientific reports 7, no. 1 (2017): 1–10.
17. Rosenblum, Sara. "Development, reliability, and validity of the Handwriting Proficiency Screening Questionnaire (HPSQ)." American Journal of Occupational Therapy 62, no. 3 (2008): 298–307.
18. K. Spoon, D. Crandall and K. Siek. "Towards Detecting Dyslexia in Children's Handwriting Using Neural Networks." International Conference on Machine Learning AI for Social Good Workshop, 2019.
19. P. Dwivedi. Handwriting Recognition Using TensorFlow and Keras. https://towardsdatascience. com/handwriting-recognition-using-tensorflow-and-keras-819b36148fe5, January, 2018.
20. Xing, Linjie, and Yu Qiao. "Deepwriter: A multi-stream deep CNN for text-independent writer identification." In 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016. pp. 584–589.
21. Train and evaluate with Keras. `https://www.tensorflow.org/guide/keras/train_and_evaluate`. Accessed: 18-Sep-2020.
22. Box plot review (article), Khan Academy. `https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/box-whisker-plots/a/box-plot-review`. Accessed: 18-Sep-2020.
23. Violin plot (article). `https://blog.bioturing.com/2018/05/16/5-reasons-you-should-use-a-violin-graph/`. Accessed: 18-Sep-2020.