 Copyright @ by LDW 2021

Applying a Convolutional Neural Network to Screen for Specific Learning Disorder

Nuriel Shalom Mor
Darca and Bnei Akiva Schools, Israel

Kathryn Dardeck
William James College, USA

Early detection is key to intervening with students diagnosed with a specific learning disorder (SLD), which includes problems with spelling, grammar, punctuation, and clarity and organization of written expression, as a means of preventing potential negative consequences from this disorder. Deep convolutional neural networks (CNNs) perform better than human beings in many visual tasks such as making a medical diagnosis from visual data. The purpose of this study was to evaluate the ability of a deep CNN to detect students with a diagnosis of an SLD based on their handwriting. A so-called MobileNetV2 deep CNN architecture was used by applying transfer learning. The model was trained using a data set of 497 images of handwriting samples from students with a diagnosis of an SLD as well as students without this diagnosis. The detection of an SLD on the validation set yielded a mean area under the receiver operating characteristics curve of 0.89. This novel attempt to detect students with the diagnosis of an SLD using deep learning can potentially provide fast initial screening of students who may meet the criteria for a diagnosis of an SLD, thereby improving their chances of effective intervention.

Keywords: **Specific Learning Disorder, Deep Learning, Deep CNNs, Transfer Learning**

Introduction

A specific learning disorder (SLD) is a neurodevelopmental condition that can be detected only after formal education starts (American Psychiatric Association [APA], 2013). About 10% of school-age children in the United States are diagnosed with this disability (Fortes et al., 2016; Gorker et al., 2017). An SLD can manifest itself in several academic areas, including reading, writing, and mathematics (APA, 2013). When the diagnosis is focused on reading, symptoms may include difficulty with word accuracy, reading fluency, and reading comprehension. A disability in written expression, in turn, shows itself in difficulties with spelling, grammar, punctuation, and organization. Finally, mathematical disabilities may manifest themselves in severe problems with memorizing mathematical facts, with fluent calculation, and with mathematical reasoning (APA, 2013).

The above symptoms are further clarified according to degree of severity, ranging from mild, moderate to severe (APA, 2013). A diagnosis of an SLD is complex and made through a combination of observation, interviews, family history, and school reports (American Psychiatric Association, 2013; McDonough et al., 2017), which can lead to a late diagnosis.

*Please send correspondence to: Nuriel S. Mor, PhD, Licensed Psychologist, Software Engineering Program with an Emphasis on AI, Darca and Bnei Akiva Schools, 7 Ort Israel Street, Bat Yam, Israel, Email: nuriel.mor@gmail.com.

Early identification of SLD is vital. If the challenges remain undetected, detrimental consequences, including high levels of psychological distress, depression, suicidality, and poorer overall mental health, may ensue (APA, 2013). Early diagnosis and intervention, on the other hand, can significantly mitigate the negative impact of an SLD on mental health (APA, 2013) by helping to prevent the frustration and decrease in wellbeing caused by an undiagnosed SLD, as found in a study conducted in Italy (Lombardi et al., 2019).

Deep Learning and Diagnosis

Deep learning algorithms are more accurate than human problem-solving strategies for many visual tasks (such as strategic board games, human and chimpanzee facial recognition, plant disease identification, and object recognition) (Esteva et al., 2017; Ferentinos et al., 2018; Schofield et al., 2019). In addition, they perform better than humans in medical diagnosis based on visual data such as skin cancer classification, breast cancer screening, and pneumonia detection (Esteva et al., 2017; McKinney et al.,2020; Rajpurkar et al., 2017). Indeed, advances in computation, very large data sets, and emerging new techniques enable deep learning algorithms to recognize very complex patterns in data that are beyond human perception (Esteva et al., 2017).

Successful deep learning applications are increasingly helping with the medical diagnostic process (Esteva et al., 2017; Kermany et al., 2018; McKinney et al ., 2020; Rajpurkar et al., 2017). For example, deep learning applications for mental disorder screening have been based mainly on data from neuroimaging (Galatzer-Levy et al., 2014; Vieira et al., 2017). Thus, a range of psychiatric and neurological disorders such as post-traumatic stress disorder (PTSD), depression, schizophrenia and more, can be screened from neuroimaging data using deep learning (Vieira et al., 2017). In addition, neurodevelopmental disorders such as attention deficit hyperactivity disorder and autism spectrum disorder can be screened from neuroimaging data with deep learning (Heinsfeld et al., 2018; Vieira et al., 2017).

With the exception of a few studies (Gurovich et al., 2019; Mor & Dardeck, 2018; Rad et al., 2018; Shukla et al., 2017), most research on using deep learning has employed neuroimaging to flag possible mental disorders. Unfortunately, this impedes implementation of deep learning in the diagnostic screening process of mental disorders because neuroimaging is rarely used in psychology due to its high cost (Galatzer-Levy et al., 2014).

Mor and Dardeck (2018) identified people at risk for PTSD using readily collected ecological risk factors and deep learning. Similarly, Shukla et al. (2017) detected developmental disorders from facial images using deep learning. Further, they built a deep learning model that performed better than humans in recognizing and differentiating among a spectrum of neurodevelopmental disorders, including autism spectrum disorder, fetal alcohol syndrome, Down syndrome, progeria, cerebral palsy, and intellectual disability. In addition, other researchers (Gurovich et al., 2019) have built a deep learning model that identifies facial phenotypes of more than 200 genetic syndromes such as Lubs XL MR, fragile X MR, Prader–Willi, MR XL Bain type, Angelman, Ch1p36 del, fetal alcohol, Potocki–Lupski, Rett, and many more.

Finally, Rad et al. (2018) used deep learning to detect stereotypical motor movement in patients with autism spectrum disorder.

Deep Learning and SLD Diagnosis

An SLD may affect handwriting in a way that can be visually distinguished (Li-Tsang et al., 2018). Symptoms may include severe problems in written expression, including difficulty with spelling, grammar, punctuation, and organization (APA, 2013). Handwriting performance and sensorimotor skills may also be a sign of an SLD (Li-Tsang et al., 2018). Thus, students with this condition write at slower speed and with greater variation in written character size (Lam et al., 2011) and require more time to complete handwriting assignments in class (Engel-Yeger et al., 2009). Engel-Yeger et al. (2009) suggested that their movements were less mature than their nonlabeled peers, and their performance is less accurate in space and time. Students with an SLD were found to erase more and complain about fatigue. The legibility of their handwriting was poor compared to the handwriting of their nondisabled counterparts. These findings apply to students with dysgraphia and other types of an SLD.

Modeling Approach

Deep convolutional neural networks (CNNs) represent state-of-the-art technology in visual tasks (Esteva et al., 2017). For example, MobileNetV2 is a deep CNN that achieves cutting-edge results in visual tasks (Sandler et al., 2018), including object detection, face attributes, fine-grain classification, and landmark recognition (Howard et al., 2017). The great benefit of MobileNet models is that they are designed to be deployed on mobile devices, allowing a rapid inference from a photo taken on a mobile device (Howard et al., 2017; Sandler et al., 2018). MobileNet models were trained on the ImageNet data set, which contains more than 14 million images with 1,000 object categories (Howard et al., 2017; Sandler et al., 2018).

Transfer learning is a technique whereby a model developed for a given task is reused as the starting point for a model to be used on another task. Specifically, it involves removing the last layer of the pre-trained deep neural network, adding new layers suitable for a current specific task, and then training with a new data set (Esteva et al., 2017; Khan et al., 2019). Transfer learning is very useful when researchers wish to utilize pre-trained, state-of-the-art deep neural networks (Khan et al., 2019).

The purpose of the current study was to evaluate the ability of deep learning using transfer learning and the MobileNetV2 model to distinguish between students who have an SLD and those who do not. Outfitted with deep learning, mobile devices can assist with rapid screening of students with an SLD based on their handwriting. This, in turn, may contribute to early detection and intervention after a careful follow-up evaluation.

Method

Sample and Outcome Measure

The target population for the study included high school students between 15 and 18 years old, grades 10 to 12, from Hadash High School, Bat-Yam, Israel. Consistent with the prevalence of SLDs reported in the literature (APA, 2013), the prevalence of an SLD for the students who comprised the sample was about 11%. Seventeen of the 152 students who participated (11%) had a diagnosis of an SLD.

Handwriting samples were collected from 152 students who volunteered to participate in the study. No remuneration was promised or given. Students agreed to provide their old notebooks for use in the study. About 500 pages of handwriting were scanned and saved as images. Two sealed and locked boxes were placed in one classroom for a few hours after the school day for two consecutive days. One box was intended for notebooks of students who had previously been diagnosed as having an SLD, while the other was designed for students without an SLD diagnosis. (The students had previously been diagnosed, unrelated to the present study.) The notebook collection process was voluntarily conducted with complete anonymity. The outcome measure of the study was a dichotomized variable of no diagnosis of an SLD vs. diagnosis of an SLD.

The pre-trained MobileNetV2 (Sandler et al., 2018) was utilized using transfer learning as it is a suitable architecture for transfer learning in visual tasks as needed in this study. The last SoftMax layer of the MobileNetV2 architecture designed for classification of 1,000 different classes of the ImageNet data set was removed, and three hidden layers of Relu neurons were added: Layer 1 of 800 neurons, Layer 2 of 400 neurons, and Layer 3 of 200 neurons. Additionally, the last layer of a single sigmoid neuron for classifying the two desired classes in this study was added: no diagnosis of an SLD vs. diagnosis of an SLD. The four layers that were added on top of the pre-trained MobileNetV2 were trained on the collected training set.

Table 1 presents the deep neural network architecture and model summary using MobileNetV2 and transfer learning. Specifically, it shows the layers in the model and the number of units in each layer. The images were pre-processed for MobileNetV2. The shape of the images was 224X224, and image normalization was applied. In addition, dropout and data augmentation were used to enhance the performance and generalizability of deep neural networks (Perez & Wang, 2017; Srivastava et al., 2014). A more detailed description of the methodology is provided in the link found in the discussion section.

Table 1. ***Model Summary***

Layer	Output Shape	Param #
Keras Layer	(None, 1280)	2257984
Dense	(None, 800)	1024800
Dropout	(None, 800)	0
Dense	(None, 400)	320400
Dropout	(None, 400)	0
Dense	(None, 200)	80200
Dropout	(None, 200)	0

Total params: 3,683,585.
Trainable params: 1,425,601.
Non-trainable params: 2,257,984.

Validation and Accuracy Metrics

The data set of 497 images of handwriting was randomly divided into a training set of 447 images and a validation set of 50 images. Five metrics were used to estimate the performance of the deep learning model: area under the curve (AUC), precision, recall, F-score, and accuracy. All the metric values reported represent the results obtained from the validation set.

Area under the curve refers to the area between the receiver operating characteristic (ROC) curve and the x-axis. The receiver operating characteristic curve is defined by plotting the true positive rate against the false-positive rate at different thresholds (Majnik & Bosnić, 2013). The area between the receiver operating characteristic is an unbiased metric of performance and may be compared to the AUC of different systems (Karstoft et al., 2015). *Precision* is defined by true positives divided by the sum of true positives and false positives (Goutte & Gaussier, 2005). *Recall* is defined by true positives divided by the sum of true positives and false negatives (Goutte & Gaussier, 2005). The *F-score* is a balanced metric, defined by a weighted average of precision and recall (Hand & Christen, 2018). Finally, *accuracy* is defined by all true predictions of the model divided by the total of all predictions (Sim et al., 2019).

Experiments and Results

The model was trained for 25 epochs. The greatest accuracy occurred after 21 epochs and then started to decline from epoch 22, as expected because of overfitting (Cha et al., 2019). Number of epochs refers to the number of times that the learning algorithm works through the entire training data set. Figure 1 shows the working system – that is, the procedure from the input of handwriting image to the output that gives the classification. The model yielding the best accuracy was saved for further analysis of performance metrics. The model yielded the following results: AUC = 0.89, precision = 0.94, recall = 0.89, F-score = 0.91, and accuracy = 0.92. Figure 2 presents the changes in accuracy during training.

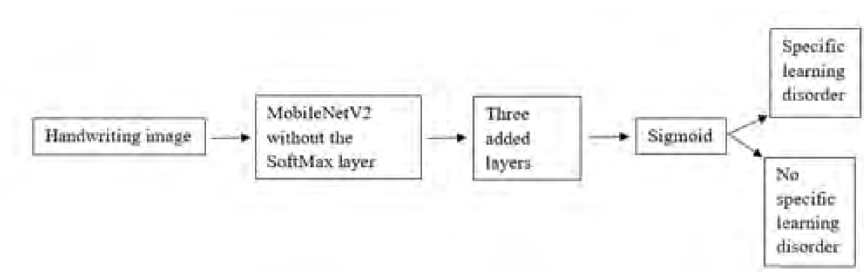

***Figure 1.* The working system.**

Figure 2. **Accuracy during training.**

Discussion

The present study evaluated the ability of deep learning algorithms to screen students with an SLD by using their handwriting. It is the first study to apply deep learning to screening for an SLD classification from handwriting samples, which were easily collected for fast inference and detection.

The study model yielded an AUC of 0.89, indicative of a good predictive model in the domain of mental diagnostics (Galatzer-Levy et al., 2014). Further, a precision of 0.94, recall of 0.89, F-score of 0.91, and accuracy of 0.92 indicate that the model yields very good results with regard to SLD detection compared to other studies that have used deep learning to detect mental disorder (Vieira et al., 2017). That is, values of performance metrics in other studies using deep learning to detect mental disorders from neuroimaging data were between 0.65 and 0.95 (Vieira et al., 2017). Further, the reported accuracy of a model designed to identify facial phenotypes of genetic disorders using deep learning was 0.91 (Gurovich et al., 2019), and the AUC and F-score of a model designed to identify people at risk for PTSD using ecological factors and deep learning were 0.91 and 0.83, respectively (Mor & Dardeck, 2018).

The finding that deep learning applied to handwriting samples provides efficient initial screening of students for SLD is promising. Given that about 10% of school-age children have an SLD (APA, 2013; Fortes et al., 2016), this can make the otherwise complex task of SLD diagnosis faster and simpler.

It is important to mention that we are not suggesting that such a model would replace the essential diagnostic process in which professionals consider a combination of information from observations, interviews, family history, and school reports (APA, 2013). We are suggesting, however, that a model such as the one designed for this study can provide fast initial screening of students for SLD, thereby significantly contributing to early detection and intervention.

Applicability of the Study

Smartphone applications that can help with the initial screening of medical or mental disorders would provide low-cost universal access to essential diagnostic care (Esteva et al., 2017). The deep learning model built in this study is based on MobileNet, which was designed for smartphones (Howard et al., 2017). MobileNet provides fast and accurate performance deployed on mobile devices (Howard et al., 2017). Outfitted with a CNN, mobile devices can aid educators, reading specialists, and other relevant professionals with a means to achieve fast initial screening of SLD. Screening for students with this condition using this system requires no more than taking a photo of a student's handwriting on a smartphone, uploading, and sending it to the model, and receiving the model answer. For further clarification, the system designed in this study may be viewed at https://colab.research.google.com/drive/1SUByhCjS29pR_njEwFKD7v3YFZ9C_i9H

Limitations and Recommendation for Future Work

This study was conducted with students at Hadash High School, Bat Yam, Israel. The results of this initial study, therefore, cannot be generalized beyond this specific Hebrew-speaking population. In addition, the size of the training set was relatively small, which is an important limitation as the size of the training data set is the most important factor for enhancing the generalizability of deep learning models (Perez & Wang, 2017). Therefore, a major recommendation for future work is to significantly increase the size of the handwriting data set.

In addition, it would be important and interesting to assess the handwriting of students using multiple languages to get a picture of how the algorithm holds up across different alphabets and writing systems. For each language, it would be necessary to develop a new system and re-examine its performance. In order to increase the generalizability of our model to different populations, the main recommendations for future work, therefore, include collecting handwriting samples from many different populations in many different languages.

Summary, Conclusion, and Future Directions

This study demonstrated the feasibility of screening students with an SLD based on handwriting samples using a deep learning algorithm. The model designed for the study is easily deployed on smartphones, enabling fast initial screening of students with an SLD simply by taking a photo of their handwriting. Early intervention is essential for children with an SLD, and a system such as this may significantly contribute to early detection and subsequent intervention. The system is far from being a universal, optimal solution because the training data set was limited. It is hoped, however, that the study's findings will serve as an inspiration for the future development of a universal solution for early screening and detection of SLD, which would ideally include many different populations from across the world.

References

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.

Cha, K. H., Petrick, N., Pezeshk, A., Graff, C. G., Sharma, D., Badal, A., ... & Sahiner, B. (2019, March). Reducing overfitting of a deep learning breast mass detection algorithm in mammography using synthetic images. In *Medical Imaging 2019: Computer-Aided Diagnosis* (Vol. 10950, p. 1095004). International Society for Optics and Photonics.

Engel-Yeger, B., Nagauker-Yanuv, L., & Rosenblum, S. (2009). Handwriting performance, self-reports, and perceived self-efficacy among children with dysgraphia. *American Journal of Occupational Therapy*, *63*(2), 182-192.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*(2), 311-318.

Fortes, I. S., Paula, C. S., Oliveira, M. C., Bordin, I. A., de Jesus Mari, J., & Rohde, L. A. (2016). A cross-sectional study to assess the prevalence of DSM-5 specific learning disorders in representative school samples from the second to sixth grade in Brazil. *European Child & Adolescent Psychiatry, 25*(2), 195-207.

Galatzer-Levy, I. R., Karstoft, K. I., Statnikov, A., & Shalev, A. Y. (2014). Quantitative forecasting of PTSD from early trauma responses: A machine learning application. *Journal of Psychiatric Research, 59*(5-6), 68-76.

Gorker, I., Bozatli, L., Korkmazlar, Ü., Karadag, M. Y., ceylan, C., Sogut, C., ... & Turan, N. (2017). The probable prevalence and sociodemographic characteristics of specific learning disorder in primary school children in Edirne. *Archives of Neuropsychiatry, 54*(4), 343.

Goutte, C., & Gaussier, E. (2005, March). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European Conference on Information Retrieval* (pp. 345-359). Springer.

Gurovich, Y., Hanani, Y., Bar, O., Nadav, G., Fleischer, N., Gelbman, D., ... & Bird, L. M. (2019). Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine, 25*(1), 60.

Hand, D., & Christen, P. (2018). A note on using the F-measure for evaluating record linkage algorithms. *Statistics and Computing, 28*(3), 539-547.

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE data set. *NeuroImage: Clinical, 17*, 16-23.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). *Mobilenets: Efficient convolutional neural networks for mobile vision applications*. arXiv preprint arXiv:1704.04861.

Karstoft, K. I., Statnikov, A., Andersen, S. B., Madsen, T., & Galatzer-Levy, I. R. (2015). Early identification of posttraumatic stress following military deployment: Application of machine learning methods to a prospective study of Danish soldiers. *Journal of Affective Disorders, 184,* 170-175.

Khan, S., Islam, N., Jan, Z., Din, I. U., & Rodrigues, J.J.C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters, 125*, 1-6.

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Dong, J. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, *172*(5), 1122-1131.

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

Li-Tsang, C. W., Li, T. M., Lau, M. S., Ho, C. H., & Leung, H. W. (2018). Handwriting assessment to distinguish comorbid learning difficulties from attention deficit hyperactivity disorder in Chinese adolescents: A case-control study. *International Journal of Methods in Psychiatric Research, 27*(4), e1718.

Lombardi, E., Offredi, I., Bettoni, R., Sarti, D., Traficante, D., & Vernice, M. (2019). Well-being in secondary school students: a comparison between specific learning disorder and difficulties. In *Proceedings 1st SRLD Conference* (pp. 114-115). Cleup.

Majnik, M., & Bosnić, Z. (2013). ROC analysis of classifiers in machine learning: A survey. *Intelligent Data Analysis, 17*, 531-558.

McDonough, E. M., Flanagan, D. P., Sy, M., & Alfonso, V. C. (2017). Specific learning disorder. In *Handbook of DSM-5 disorders in children and adolescents* (pp. 77-104). Springer.

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Etemadi, M. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89-94.

Mor, N. S., & Dardeck, K. L. (2018). Quantitative forecasting of risk for PTSD using ecological factors: A deep learning application. *Journal of Social, Behavioral, and Health Sciences*, *12*(1), 4.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness, and correlation. *International Journal of Machine Learning Technology*, *2*(1), 37-63.

Rad, N. M., Kia, S. M., Zarbo, C., van Laarhoven, T., Jurman, G., Venuti, P., ... & Furlanello, C. (2018). Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. *Signal Processing*, *144*, 180-191.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Lungren, M. P. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4510-4520). .IEEE

Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., Matsuzawa, T., Biro, D., & Carvalho, S. (2019). Chimpanzee face recognition from videos in the wild using deep learning. *Science Advances, 5*(9), eaaw0736.

Shukla, P., Gupta, T., Saini, A., Singh, P., & Balasubramanian, R. (2017, March). A deep learning framework for recognizing developmental disorders. In *Proceedings of 2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 705-714). IEEE.

Sim, J. Y., Jang, Y., Kim, W. C., Kim, H. Y., Lee, D. H., & Kim, J. H. (2019). Comparing the accuracy (trueness and precision) of models of fixed dental prostheses fabricated by digital and conventional workflows. *Journal of Prosthodontic Research, 63*(1), 25-30.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*, 1929-1958.

Vieira, S., Pinaya, W. H., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience & Biobehavioral Reviews*, *74*, 58-75.

Authors' Note

We wish to thank teaching assistant Karin Volovik for her assistance in gathering and processing data for this study.