Retinal Image Analysis for Simultaneous Classification and Severity Grading of Attention-Deficit Hyperactivity Disorder and Autism Spectrum Disorder using Deep Learning

Luke Lee[1] and Kevin Ingram[#]

[#]Advisor

ABSTRACT

Over the last 20 years, the number of children born with Autism Spectrum Disorder (ASD) and Attention-Deficit Hyperactivity Disorder (ADHD) has significantly increased. According to the U.S Centers for Disease Control and Prevention, the number of children with ASD in the United States had increased from 1 in 150 in 2000 to 1 in 36 in 2020, which is an increase of 316.7% and the number of children with ADHD has been around 10% of Population. As the number of children with ASD and ADHD is increasing rapidly, an early diagnosis of both mental disorders is essential, which can contribute to the improvement of the condition of children significantly. By employing retinal images, the early detection of ASD and ADHD in the developmental stages of childhood becomes feasible. Consequently, the development of a diagnostic system that identifies the presence of ASD and ADHD is essential for healthcare interventions and the enhancement of children's well-being when potential remediation is attainable. This research aims to develop a system for early diagnosis of ASD and ADHD utilizing retinal images and deep learning through convolutional neural networks. The proposed approach classifies the retinal images into disorder categories and severity levels. Experimental results demonstrate the viability of the proposed approach as a biometric for the early diagnosis of ASD and ADHD.

Introduction

Autism Spectrum Disorder (ASD) and Attention-Deficit Hyperactivity Disorder (ADHD) are neurodevelopmental disorders that present serious challenges in various aspects of life. ASD is a complex disorder that affects social interactions, communication skills, and behavioral patterns. It manifests as a spectrum, meaning that the symptoms can range from mild to severe, and each individual's experience with the disorder is unique. Common challenges include difficulties in understanding social cues, impaired verbal communication, and a tendency for repetitive behaviors or fixations on specific interests. Similarly, ADHD is characterized by persistent issues with attention, hyperactivity, and impulsivity. These symptoms often manifest in daily life as difficulties in maintaining focus, excessive fidgeting or restlessness, and impulsive decisions that may lead to negative consequences. Similar to ASD, ADHD can also vary in severity and impact different areas of life, including academic performance, occupational success, and interpersonal relationships.

One crucial similarity between ASD and ADHD is the immense importance of early diagnosis. Identifying these disorders at an early stage opens the door for prompt intervention, which is often more effective when initiated at a younger age. Early therapeutic interventions can focus on enhancing social and communication skills in children with ASD, while behavioral therapies for ADHD can help in managing symptoms and improving focus. Moreover, an early diagnosis is not just beneficial for the child but also for the parents and caregivers. It provides a much-needed framework for understanding the child's unique behavioral patterns and developmental needs. This understanding enables parents to adopt more effective parenting strategies, tailored to their child's condition, thereby fostering a more supportive and positive parent-child relationship. By emphasizing the importance of early diagnosis and intervention, we can significantly alter the trajectory of these disorders, offering children a better chance at a fulfilling and independent life. Therefore, it is crucial to propose an automated method to detect ASD and ADHD at an early stage of children which can be held easily and quickly, with high accuracy.

In this study, I introduce a novel method for diagnosing ASD and ADHD by utilizing retinal images. Given the close association between retinal images and brain structure, I explored the potential of using retinal images to detect mental disorders, particularly in cases where individuals with ASD and ADHD exhibit distinctive brain structures compared to those without such disorders. The focus of my research is on classifying ASD and ADHD in children through the analysis of retinal images. Employing convolutional neural networks, I trained a machine learning model using retinal images from individuals with ASD, ADHD, and those without any mental disorders, encompassing variations in disorder severity.

The following chapters of this paper are structured as follows: Chapter 2 provides background knowledge to enhance the comprehension of the proposed approach. In Chapter 3, I explain the details of the proposed method, explaining its components and underlying mechanisms. Chapter 4 offers a comprehensive presentation of experimental results and an in-depth analysis. Finally, Chapter 5 summarizes the key findings and implications of this research.

Related Work

Retinal Image

Retinal imaging involves capturing detailed images of the retina, the light-sensitive tissue at the back of the eye responsible for converting light into electrical signals sent to the brain. It consists of several layers of cells, including photoreceptor cells (rods and cones) that capture light and convert it into electrical signals. These signals are then transmitted through the optic nerve to the brain, where they are interpreted as visual information.

Retinal imaging has various applications in medical diagnosis, including detecting and monitoring eye diseases such as diabetic retinopathy, macular degeneration, and glaucoma. Additionally, recent research has explored the potential of retinal imaging in diagnosing and monitoring systemic conditions, including neurological disorders like Alzheimer's disease (Cheung et al. 2022), and Parkinson's disease (Hasan et al. 2023).

Inspired by this previous research, I present a pioneering neurodevelopmental disorder diagnosis system that leverages retinal images. This system processes retinal images as input and generates a probability prediction indicating whether the patient has a neurodevelopmental disorder. Further elaboration on the workings of this system will be offered in Chapter 3.

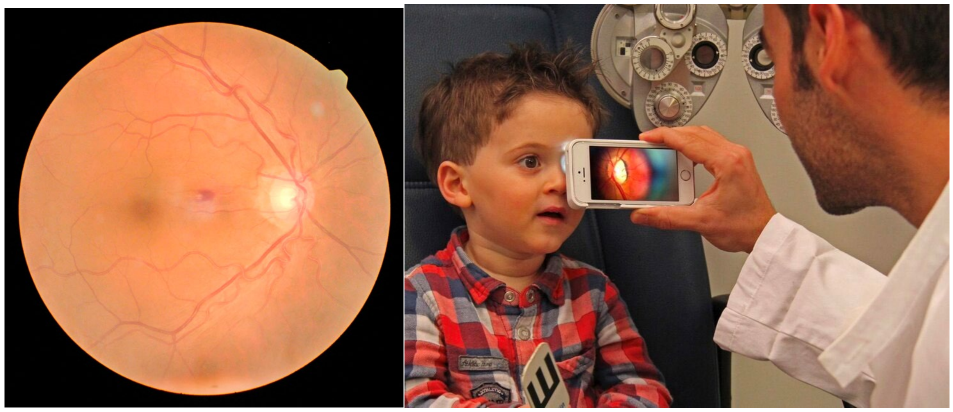


Figure 1. Example of retinal image and retinal screening device (Russo 2020)

Image Classification with Convolutional Neural Network

An image classification system, particularly one based on convolutional neural networks, is a technology that automates the process of identifying and assigning predefined labels or categories to images. This is a type of computer vision task, with applications ranging from medical diagnosis, autonomous vehicles, and facial recognition to the analysis of satellite imagery.

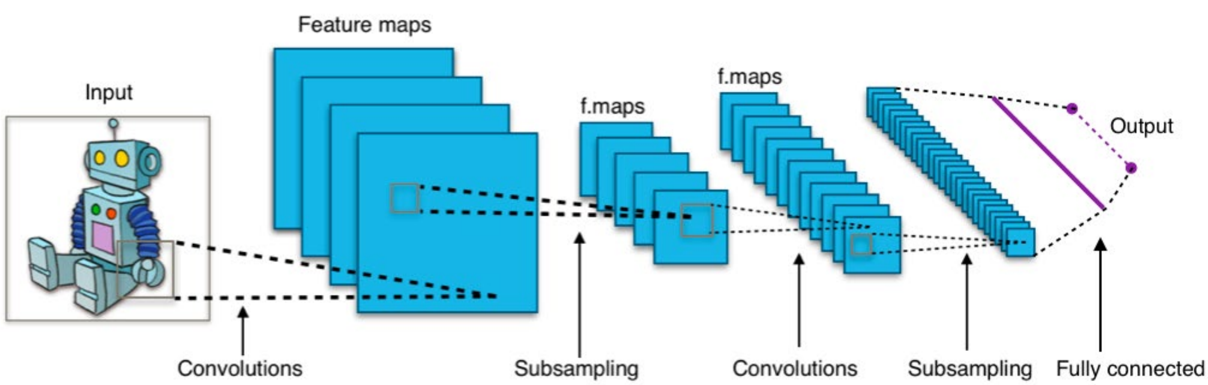


Figure 2. Example of the convolutional neural network-based classification system (Sharma 2023)

Convolutional neural networks are a type of deep neural network designed specifically for image-related tasks. They have proven highly successful in image classification due to their ability to automatically learn hierarchical representations of features. The main operation of convolutional neural networks is convolve operations that apply filters (also known as kernels) to input images, capturing local patterns like edges or textures. Multiple convolutional layers are stacked to capture increasingly complex features. Convolutonal neural networks excel at capturing spatial hierarchies and recognizing patterns which makes them ideal for image-related tasks.

Taking advantage of this benefit, my research employs a convolutional neural network-based classification system to analyze retinal images for diagnosing neurodevelopmental disorders, such as ADHD and ASD. A detailed exposition of the network architecture is provided in Chapter 3

Proposed Method

Figure 3 illustrates the architecture of the proposed neurodevelopmental disorder classification system. This system takes a singular retinal image as input and generates two predictions: the disorder class and its corresponding severity level. The architecture comprises three modules, including convolutional layers, and two classifiers - one for disorder classification and another for severity classification.

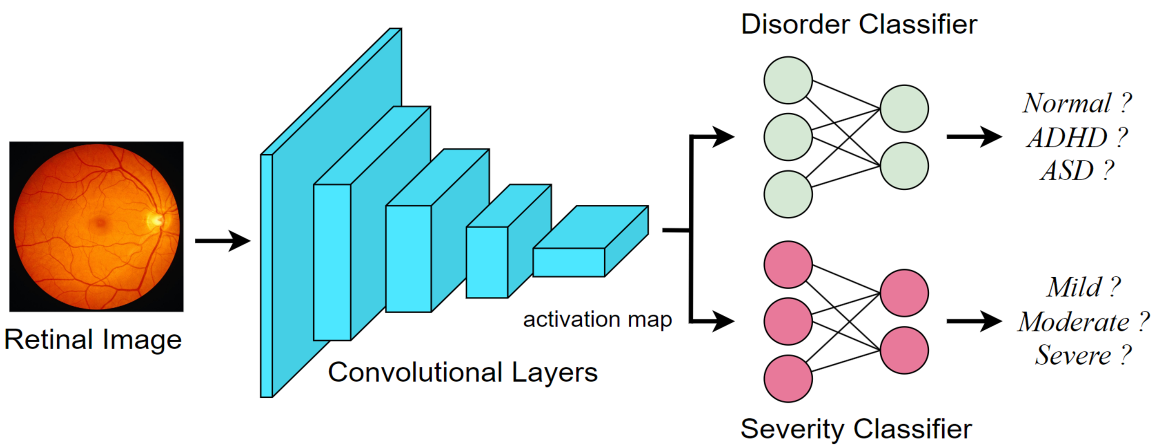


Figure 3. Architecture of the proposed neurodevelopmental disorder classification system

The convolutional layer processes the input image to extract an activation map, mathematically representing meaningful visual patterns present in the provided retinal image. This activation map is directed to both the disorder and severity classifiers to predict the output which encompasses the type of disorder and its corresponding severity level. Equations 1 and 2 summarize the prediction process.

Equation 1: Disorder classifier

$$DCL\left(Z; \phi_{disorder}\right) : Z \rightarrow P_{disorder}$$

Here, Z denotes the activation map extracted from the convolutional layers and $\Phi_{disorder}$ represents the learnable parameters of the disorder classifier. The variable $P_{disorder}$ represents the probability of predicting the disorder type, including normal, ADHD, and ASD.

Equation 2: Severity classifier

$$SCL\left(Z; \phi_{severity}\right) : Z \rightarrow P_{severity}$$

In Equation 2, $\Phi_{severity}$ severity represents the learnable parameters of the severity classifier, and $P_{severity}$ signifies the probability of predicting severity types, including mild, moderate, and severe. For network training, I employ the cross-entropy loss function which quantifies the difference between the prediction probabilities and the ground truth distribution. Equation 3 explains the cross-entropy loss function.

Equation 3: Cross entropy loss function

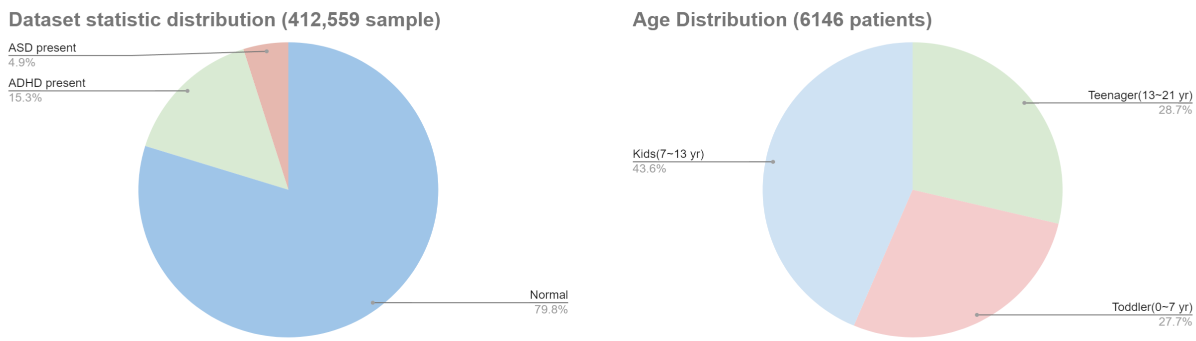
$$L = -log_e P$$

In Equation 3, P signifies the probability of each prediction from both the severity and disorder classifiers. This loss function is commonly employed for training various types of classification networks. The function yields a zero value when the model produces perfectly accurate results and approaches infinity when deviations occur.

For the architectures of the proposed method, I implement three-layer neural networks for both the disorder and severity classifiers. In terms of convolutional layers, I explored several established convolutional neural network architectures. The selection of convolutional neural network architectures will be examined further in Chapter 4.

Experimental Results

In this chapter, I provide a comprehensive overview of the dataset employed for training the proposed method. I utilized the Neurodevelopmental Disorder Retinal Image Dataset (AI-Hub 2023), which includes retinal image samples labeled into three classes: normal, ADHD, and ASD. Additionally, ADHD and ASD image samples are categorized based on severity levels, including mild, moderate, and severe. The distribution of samples in the dataset is illustrated in Figure 4.



(a) (b)

Figure 4. Dataset sample distribution. (a): class distribution and (b): patient age distribution

The dataset comprises a total of 412,559 retinal image samples, collected from children aged between 0 to 21. Within the overall dataset, 79.8% corresponds to normal samples, 15.3% to ADHD, and the remaining 4.9% to ASD. To assess the performance of the proposed method, four commonly employed inference metrics in image classification tasks—accuracy, recall, precision, and F1-score—are utilized, as explained in Figure 5.

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

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\text{-}score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Figure 5. Inference metrics used in this research

In the experiment, I conducted a study by selecting eight distinct convolutional neural network architectures to construct the convolutional layers in the proposed method. Subsequently, I trained each architecture using the dataset with identical training hyperparameters. The four aforementioned inference metrics were measured for each trained model, and the results are summarized in Table 1 and illustrated in Figure 6.

Table 1. Inference metric evaluation and comparison (disorder classification)

	Accuracy	Recall	Precision	F1-Score
VGG19 (Simonyan et al. 2014)	0.8465 (±0.0016)	0.7895 (±0.0012)	0.6766 (±0.0007)	0.7104 (±0.0008)
MobileNetV2 (Sandler et al. 2018)	0.8570 (±0.0005)	0.8091 (±0.0008)	0.6910 (±0.0011)	0.7316 (±0.0009)
ShuffleNet (Zhang et al. 2018)	0.8555 (±0.00016)	0.8082 (±0.0014)	0.6905 (±0.0013)	0.7288 (±0.0011)
Xception (Fran et al. 2017)	0.8637 (±0.0010)	0.8189 (±0.0013)	0.6972 (±0.0015)	0.7375 (±0.0009)
HRNet-w32 (Wang et al. 2020)	0.8656 (±0.0012)	0.8184 (±0.0011)	0.7005 (±0.0009)	0.7384 (±0.0013)
ResNext-50 (Xie et al. 2017)	0.8704 (±0.0009)	0.8231 (±0.0009)	0.7055 (±0.0008)	0.7436 (±0.0011)
Densenet-121 (Huang et al. 2017)	0.8823 (±0.0012)	0.8262 (±0.0014)	**0.7254 (±0.0009)**	0.7561 (±0.0015)
Resnet-50 (He et al. 2016)	**0.8894 (±0.0012)**	**0.8419 (±0.0008)**	0.7241 (±0.0007)	**0.7621 (±0.0009)**

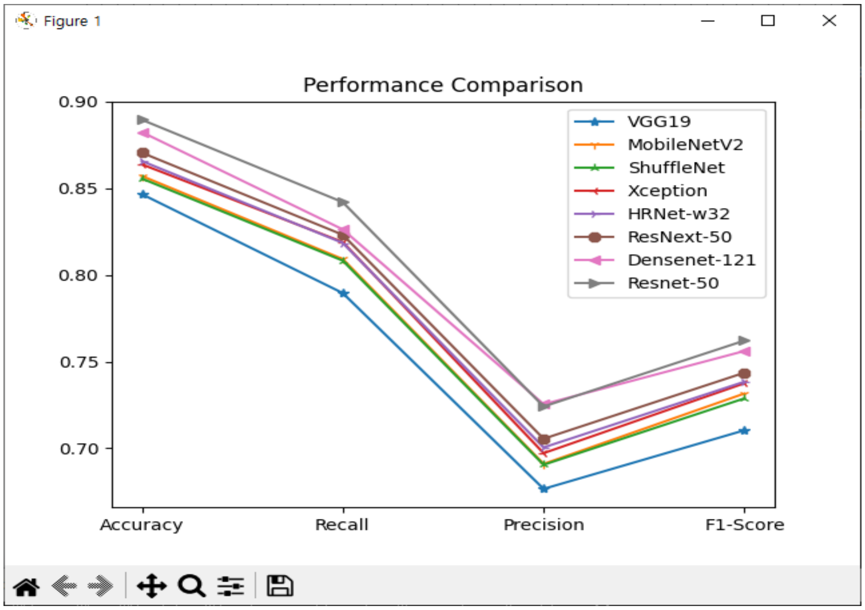


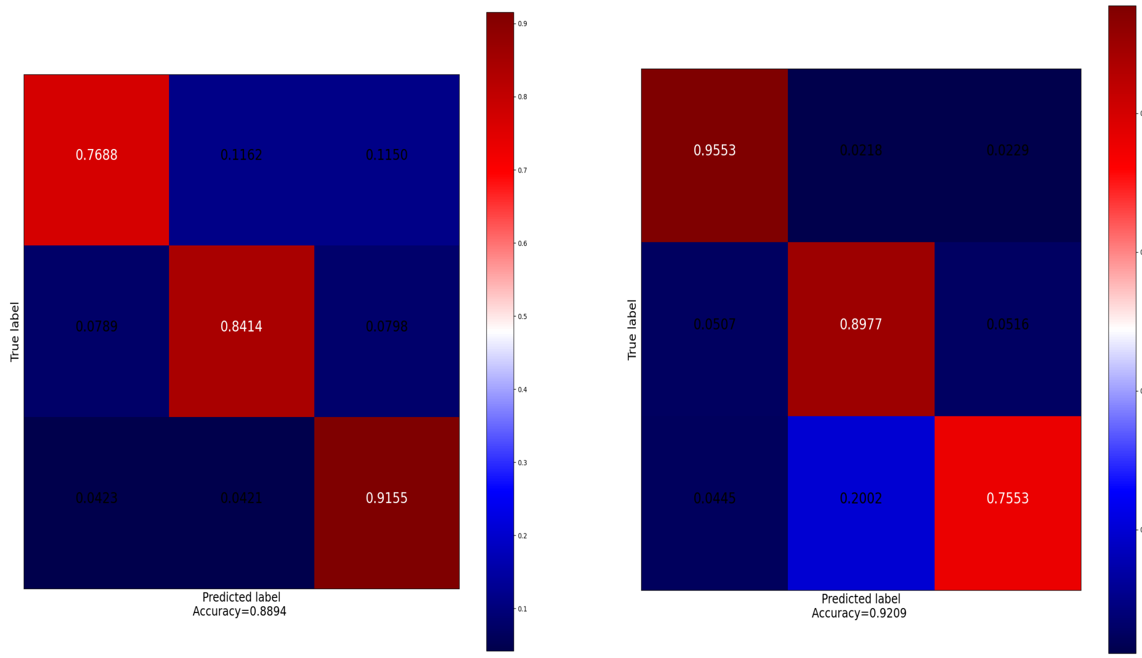
Figure 6. Inference metric evaluation and comparison (disorder classification)

As shown in Table 1 and Figure 6, all trained models utilizing the eight selected architectures achieved a minimum accuracy of 84% which showcases remarkable results. VGG19 (Simonyan et al., 2014) demonstrated the lowest performance, attributed to its relatively shallow network depth. Conversely, MobileNetV2 (Sandler et al., 2018), ShuffleNet (Zhang et al., 2018), Xception (Fran et al., 2017), and HRNet-w32 (Wang et al., 2020) exhibited comparatively higher performance due to their deeper network architectures. Densenet-121 and ResNet-50 yielded the most accurate results which benefit from their extremely deep network structures.

Furthermore, I assessed the performance of the severity classification task for ResNet-50, which achieved the highest accuracy in the disorder classification task, as summarized in Table 2. The result indicates an accuracy of approximately 92%, highlighting a remarkable performance in severity classification as well.

Table 2. Inference metric evaluation (severity classification)

	Accuracy	Recall	Precision	F1-Score
Resnet-50 (He et al. 2016)	0.9245 (±0.0012)	0.8798 (±0.0008)	0.8705 (±0.0007)	0.8750 (±0.0009)



(a) (b)

Figure 7. Confusion matrix result. (a): disorder classification, and (b): severity classification

Figure 7 represents the confusion matrix of the proposed method which proves a high true positive rate that attests to the consistency and robustness of the approach. In Figure 7(b), an analysis of severity classification performance indicates that the trained model tends to misclassify the moderate and severe classes which represents a limitation of the proposed method. This issue will be addressed in future studies through the utilization of image segmentation techniques focused on retinal blood vessel analysis to enhance overall performance.

Lastly, I conducted a data augmentation experiment to investigate how modifying the pixel distribution of input retinal samples impacts the accuracy of the proposed method.

Table 3. Data augmentation study

Data Augmentation	Accuracy
Baseline	0.8894 (±0.0012)
Horizontal Flip	0.8754 (±0.0009)
Brightness	0.8767 (±0.0011)
Grayscale	0.8481 (±0.0008)
Color Jitter	0.8387 (±0.0009)
Histogram Equalization	**0.8948** (±0.0011)

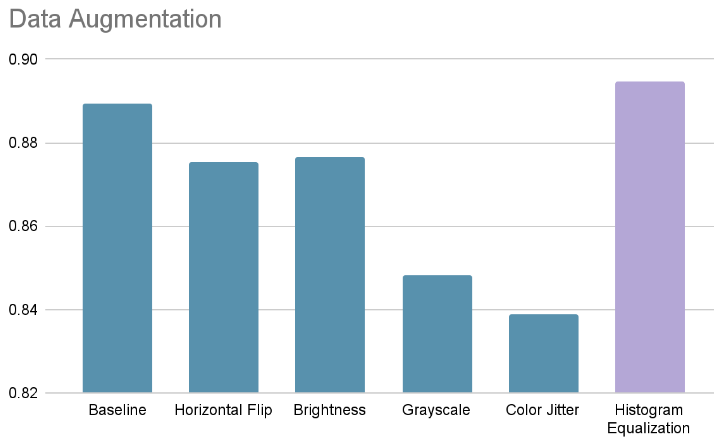


Figure 8. Data augmentation study

Table 3 and Figure 8 summarize the findings from the data augmentation study. Initially, I trained the ResNet-50-based proposed method without employing any data augmentation technique referred to as the baseline. I trained the same architecture model using five popular data augmentation techniques, including horizontal flip, brightness adjustment, grayscale transformation, color jitter, and histogram equalization.

The incorporation of additional geometric information, such as horizontal flips, did not contribute to an increase in accuracy. Similarly, modifications at the pixel-level intensity through brightness adjustment, grayscale transformation, or color jitter did not result in improved accuracy. However, histogram equalization, which adjusts image contrast by redistributing intensity values across the entire range, demonstrated increased accuracy. I attribute this improvement to its ability to normalize retinal images into a specific color range.

Conclusion

In this research, I proposed a neurodevelopmental disorder diagnosis system that leverages retinal images and employs convolutional neural networks for classification. The objective was to contribute to the early detection and diagnosis of Attention-Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) during childhood developmental stages. The comprehensive exploration of this innovative approach has yielded valuable insights and promising outcomes. The proposed system demonstrated robust performance, achieving at least 84% accuracy across various convolutional neural network architectures, as detailed in Chapter 4. Furthermore, the severity classification task for ResNet-50, the top-performing model in the disorder classification task, displayed an impressive accuracy of approximately 92%. While the proposed approach showcased promising results, limitations were identified, particularly in the misclassification of moderate and severe severity levels. In the future, I plan to develop image segmentation techniques, specifically focusing on the analysis of retinal blood vessels, to mitigate this issue and enhance overall performance.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Cheung, C. Y., Ran, A. R., Wang, S., Chan, V. T., Sham, K., Hilal, S., ... & Wong, T. Y. (2022). A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study. The Lancet Digital Health, 4(11), e806-e815.
Fran, C. (2017). Deep learning with depth wise separable convolutions. In IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.48550/arXiv.1610.02357

Hasan, M. M., Phu, J., Sowmya, A., Meijering, E., & Kalloniatis, M. (2023). Artificial intelligence in the diagnosis of glaucoma and neurodegenerative diseases. Clinical and Experimental Optometry, 1-17.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). https://doi.org/10.48550/arXiv.1512.03385

Russo, A. (2020. Oct 15), "*SMARTPHONE-BASED RETINAL SCREENING DEVICE*": Science Gallery Dublin
https://dublin.sciencegallery.com/seeing-1/d-eye

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). https://doi.org/10.48550/arXiv.1801.04381

Sharma D, (2023, Jun 14). "*Image Classification Using CNN | Step-wise Tutorial*": analyticsvidhya
https://www.analyticsvidhya.com/blog/2021/01/image-classification-using-convolutional-neural-networks-a-step-by-step-guide

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., ... & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3349-3364. https://doi.org/10.48550/arXiv.1908.07919

Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6848-6856).
https://doi.org/10.48550/arXiv.1707.01083