HIGH SCHOOL EDITION
Journal of Student Research

Volume 14 Issue 1 (2025)

Rapid Screening of Attention-Deficit/Hyperactivity Disorder using Fundus Photography with Retinal Vessel and Optic Disc Morphology

Jooha Lee[1] and Sherrie Lah[#]

[1]Crean Lutheran High School, USA
[#]Advisor

ABSTRACT

Attention-Deficit/Hyperactivity Disorder (ADHD) is a specific type of neurodegenerative disorder increasingly prevalent in younger age groups. At the present time, the number of children diagnosed with ADHD is growing at a faster rate, and this trend makes an effective diagnosis method and treatment plan necessary. Traditional screening methods of ADHD for children include analyzing electroencephalography for brain activities, the checklist method that notes any behaviors of patients that may resemble those with ADHD, and the red circle green box method, which classifies behaviors that may indicate ADHD as red circles and others as green boxes to evaluate patients. However, these methods of diagnosis tend to be subjective and therefore have limitations in accuracy. This paper presents an alternative means of predicting whether a child has ADHD or not, utilizing a machine learning algorithm and a convolutional neural network architecture. The proposed model provides an innovative approach of rapid and accurate screening of ADHD by segmenting specific features of fundus photography such as the retinal vessels and optic disc. Experiments on six different convolutional neural network architectures were conducted to retrieve the highest accuracy of 88.15% from the DenseNet-201 model. Two different morphological operations were carried out in an ablation study in order to demonstrate the features' contribution to overall model performance. Thus, this model proves to be a viable biomarker that detects and assesses the severity of ADHD. With further development, it has high potential to serve as a valuable tool that is both accurate and widely accessible for all people.

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a mental disorder that significantly affects behaviors primarily in children. ADHD is mainly categorized into two broad types: inattentive and hyperactive/impulsive. The inattentive type is characterized by symptoms such as having difficulties focusing on work, avoiding activities that may require mental or physical effort, and failing to recall daily tasks. In the hyperactive/impulsive type of ADHD, people often fidget with their hands, have difficulty staying still, or interrupt others in conversations and activities. At the current state, an estimated seven million, or about 11% of children aged three to seventeen years in the United States are diagnosed with ADHD, with its numbers growing at an increasing speed every year (CDC 2024).

ADHD is known to be a lifelong condition, with its symptoms persisting more severely in some people than others. Although there are diverse types of medication that may alleviate and manage severe symptoms of the disorder, ADHD is not known to have a complete recovery. However, recent studies have revealed that timely diagnosis at an earlier age makes a more effective and thorough treatment possible, alleviating the lifelong symptoms of the disorder. Research conducted by Zahmacioglu et al. compared two groups of children: one group was diagnosed with ADHD and treated at an earlier age, while the other received comparatively late screening and without the initiation of treatment (Zahmacioglu et al. 2017). Results showed that children who were diagnosed at a later age scored higher on loneliness scores and lower on self-esteem scales, further approving that early detection and treatment significantly reduces the negative effects of ADHD on children's psychological development.

With that, it is crucial that ADHD is diagnosed in a timely and accurate manner. Thus, I propose an automated system that is capable of an early and precise screening of ADHD. The proposed system develops upon previous research by specifically utilizing ophthalmologic features that were shown to be highly correlated with the occurrence of ADHD. The proposed system comprises a segmentation module and an ADHD severity classification module. The segmentation module isolates medically important areas of ophthalmologic images including the optic disc and retinal blood vessels. The ADHD severity classification module then analyzes these isolated images to assess the severity of ADHD.

This paper is structured as follows: Chapter 2 discusses related works, Chapter 3 explains the proposed network in detail, Chapter 4 attests the superiority of the proposed network through diverse experiments, and Chapter 5 concludes the paper with a summary.

Related Work

Fundus Image

A fundus refers to a photograph of the posterior segment of the eye, containing ophthalmologic features such as the optic disc, optic cup, retinal blood vessels, and the retina. A fundus photograph is taken with a specialized fundus camera, which directs a light source through the pupils of the eyes to view the inner, posterior part of the eye. A series of mirrors and lenses located inside the fundus camera allows the photograph to contain a wide and clear view of the ocular components. Fundus photographs can be taken by diverse types of fundus cameras including confocal scanning laser ophthalmoscope based fundus cameras, mydriatic fundus cameras, and even some smartphone applications, as shown in Figure 2. These images are frequently examined and analyzed by medical professionals to detect possible eye-related diseases such as blindness, glaucoma, and retinopathy.

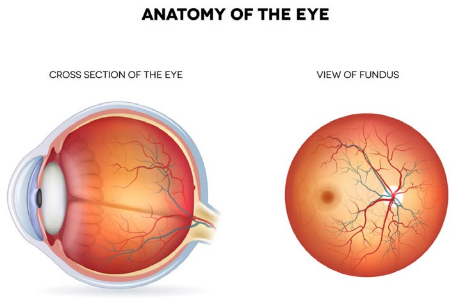


Figure 1. Example of a fundus photography (Ostrowski 2023)

Figure 2. Fundus photograph taken with a smartphone (Charters 2022)

Screening for Neurodevelopmental Disorder using Fundus Imaging

Recent studies have revealed that fundus photography may be capable of screening neurodevelopmental disorders such as autism, attention deficit/hyperactivity disorder (ADHD), and schizophrenia. This claim was derived from the fact that the eye is one of the closest organs located near the brain, and therefore may have a high probability of reflecting abnormalities associated with these neurological disorders.

One specific research conducted by Kim et al. from Yonsei University, South Korea analyzed the relationship between autism spectrum disorder (ASD) and its diagnosis through fundus imaging. The researchers conducted diverse experiments with machine learning models comparing various features of retinal photographs such as the optic disc, retinal nerve fiber layer, and retina between patients with and without ASD. Results revealed that there was a significant difference in the ocular features of those who were diagnosed with the disorder. Specifically, the trained algorithm reflected the optic disc area to be an important feature of distinguishing whether a person was likely to have ASD, suggesting that the optic disc of ASD patients was noticeably different from that of people without ASD in terms of appearance. The paper concludes that fundus photography may be a viable biomarker for diagnosing not only ASD, but other neurological disorders as well. Further implications involved the prediction of the severity of specific neurological conditions in individual patients (Kim et al. 2023).

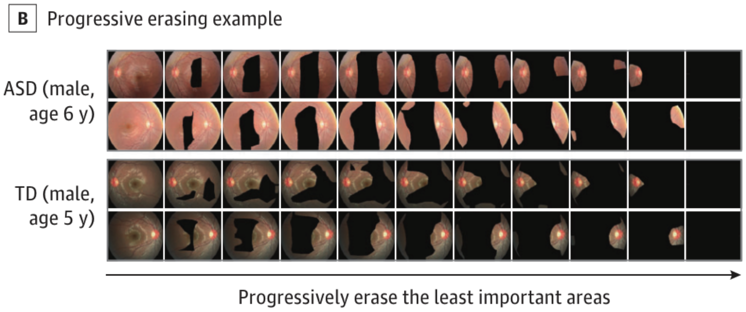


Figure 3. Differences in the ocular features of a patient diagnosed with ASD and without ASD (Kim et al. 2023)

Proposed Method

This paper is designed to prove how ophthalmologic features in fundus photography such as the retinal vessels and the optic disc could be utilized as a biomarker to diagnose and assess the severity of attention-deficit/hyperactivity disorder (ADHD). It further demonstrates the feasibility of this approach and how it could be widely applicable in clinical settings.

In order to fulfill this, I propose an artificial network consisting of two stages: optic disc and retinal vessel segmentation and ADHD severity prediction. In the first stage, fundus photography is used as an input to produce a binary segmentation map, highlighting only the retinal blood vessels and the optic disc. Then, the fundus photograph emphasizing these features ultimately predicts ADHD severity of the patient in the second stage. Furthermore, morphological operations including dilation and erosion are applied to track the varying performance of the model in relation to the contribution of individual features, testifying the model's feasibility of being a medical biomarker for ADHD diagnosis and severity assessment.

Optic Disc and Vessel Segmentation

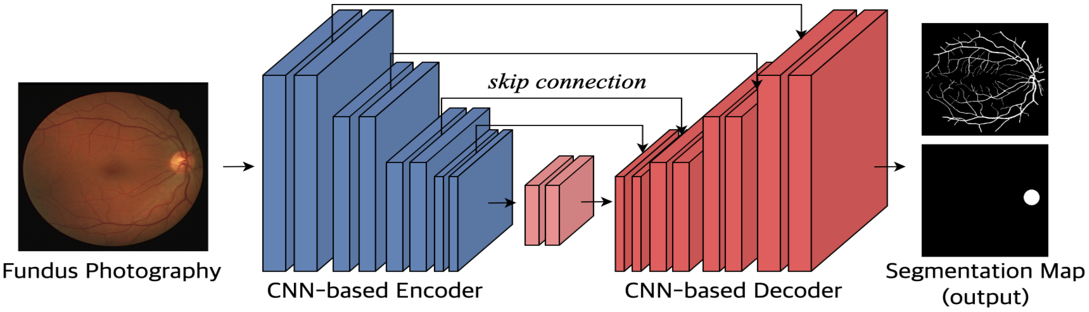


Figure 4. The proposed optic disc and retinal vessel segmentation network

The first stage of the proposed network handles the task of optic disc and retinal vessel segmentation. It is composed of a U-net architecture, often used for handling medical datasets. RGB fundus images are inputted into a CNN-based encoder, where information about its geometric features such as the retinal vessels and optic disc are extracted. By obtaining information from this data, a CNN-based decoder produces two different segmentation maps in the form of binary masks, containing the retinal vessel and optic disc information, respectively. Only the vessel and optic disc parts of the original fundus photography are highlighted in white, with a pixel value of 1. The other insignificant features of the fundus are shown in black, or with a pixel value of zero, representing the background.

During the training process, a pixel-wise cross-entropy loss function is used to track and analyze the model's performance. The network classifies each pixel into one of three categories: retinal vessel, optic disc, and the remaining background features of the fundus. This classification produces a probability predicted by the model, which is then compared to the ground truth through pixel-wise calculation. A cross entropy value closer to zero indicates higher accuracy. Utilizing the loss function allows the observance of the model's accuracy and performance as it changes over iterations.

Equation 1: Pixel-wise cross-entropy loss function

$$L_{ce} = -log_e(\hat{p})$$

Here, p_hat denotes the predicted probability of the network, among the three categories of vessel, optic disc, and the background. For instance, if the categories are noted as [vessel, optic disc, background], one pixel showing the optic disc would have the ground truth vector of [0 1 0]. Say that the proposed network predicted the pixel to be [0.8, 0.2, 0], with 20% chance of the pixel being in the true optic disc category. Then, the loss L_{ce} would be calculated as follows: $-\log_e(0.2) \approx 1.609$.

ADHD Severity Prediction

The second stage mainly predicts the severity of ADHD in individual patients through the information extracted from the fundus photography. The segmentation maps highlighting the optic disc and vessel area from the first stage are combined together. The binary mask is then applied to the original fundus photography, employing pixel-wise multiplication. Through this process, the retinal vessels and the optic disc retain their original pixel values in RGB, while other parts of the fundus remain with a pixel value of 0. This ensures that the retinal vessels and optic disc are the only features that are being emphasized while also maintaining their distinct RGB values. Figure 5 shows the resulting fundus photography that will be used for the second stage of the proposed model.

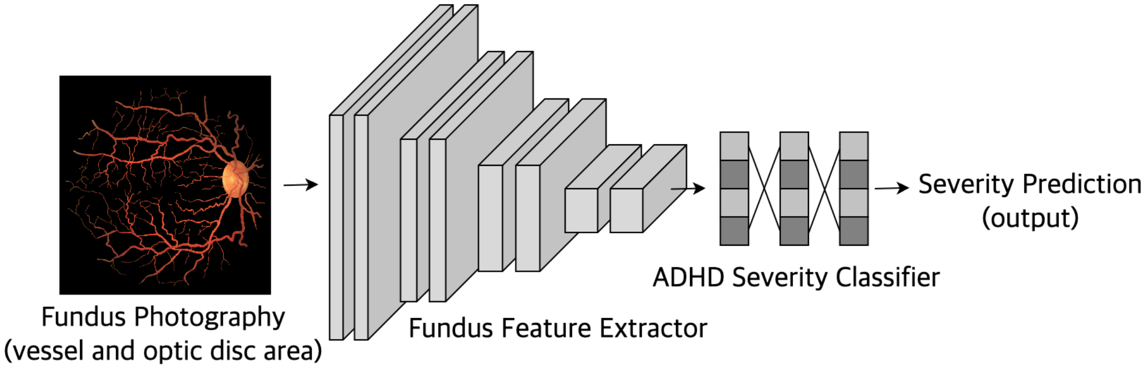
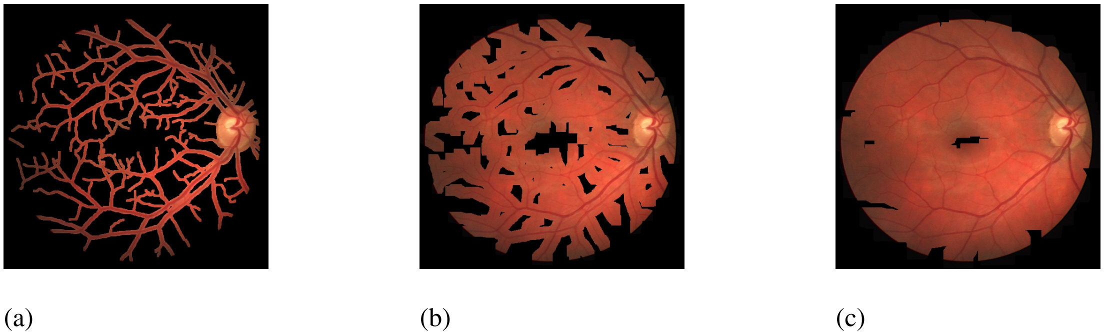


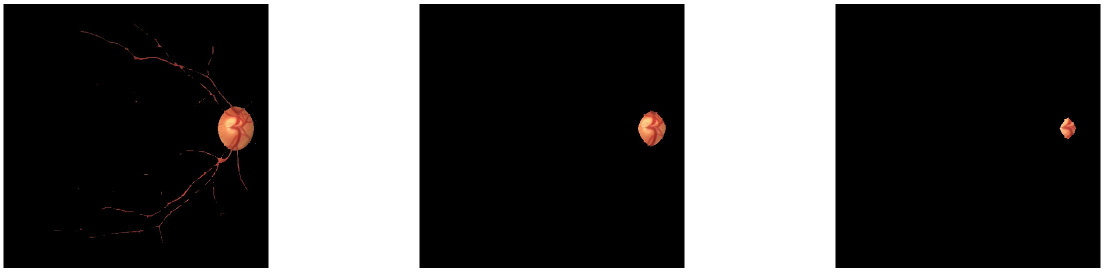
Figure 5. The proposed optic disc and retinal vessel segmentation network

The image is inputted into the fundus feature extractor, then through the ADHD severity classifier. The network analyzes important ocular features, then produces an ADHD severity prediction based on this inspection. The output, or the severity prediction, calculates the probability of the patient's ADHD severity among four categories: normal, mild, moderate, and severe. Say the categories are shown in the form of a vector, [normal, mild, moderate, severe]. If a patient's fundus had a ground truth vector of [1, 0, 0, 0], the patient would not have ADHD, and is considered normal.

The model's performance is evaluated through a cross entropy loss function, identical as the one used in the first stage. The ground truth vector and the model's predictions are compared through pixel wise calculation, allowing for the assessment of the model's performance.

Furthermore, morphological operations including dilation and erosion are applied in order to find out which features affect the model's performance the most. Dilation is the process of expanding input features at a certain scale, while erosion acts as the opposite, shrinking input features. For instance, Figure 6 (a) is an example of the original segmented fundus image, with the retinal vessel and optic disc features. Figure 6 (b) and (c) shows the images after they have been processed through dilation five and ten times, respectively. Figure 6 (d), (e), and (f) shows the fundus after they were eroded for one, five, and ten times respectively. These operations were repeated for multiple iterations, while tracking the model's variance in accuracy. Through this process, it is possible to acquire information about how different features contribute to the model's performance, with detailed results explained in Chapter 4.





(d) (e) (f)

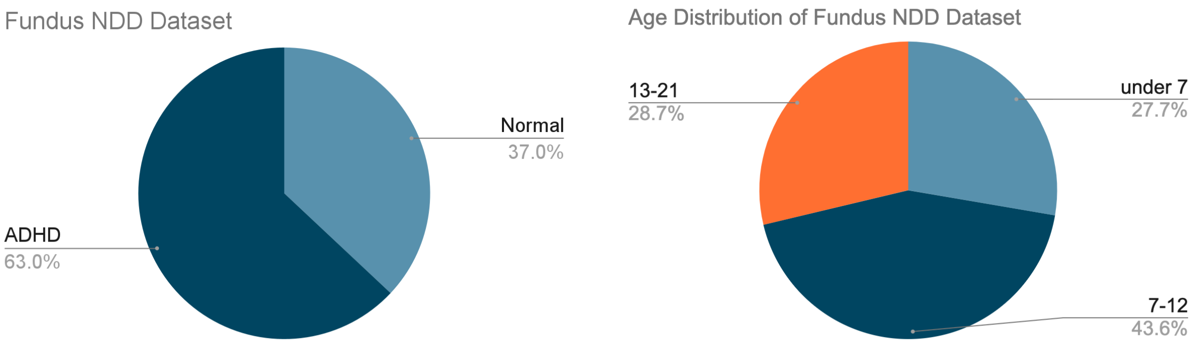
Figure 6. Segmented fundus images going through dilation and erosion operations
(a): original segmented fundus input, (b): dilation x 5, (c): dilation x 10, (d):erosion x 1, (e) erosion x 5, (f) erosion x 10

Experimental Results

Dataset

The dataset utilized for the proposed method is called the Fundus NDD Dataset, with 100,395 fundus samples from diverse patients (AI hub 2024). As indicated in Figure 7 (a), a total of 37,145 samples, or about 37.0% of the total samples are from patients with normal development. The remaining 63.0%, or 63,250 samples of the fundus are from patients diagnosed with ADHD.

Figure 7 (b) reflects the age distribution of this dataset, largely divided into three categories. Children under the age of 7, categorized under early childhood, made up 27.70% of the total samples. Middle and late childhood with ages between 7 and 12 took up about 43.57%, and the remaining 28.73% were comprised of teenagers from ages 13 to 21.



(a) (b)

Figure 7. Data distribution of the Fundus NDD Dataset
(a): ADHD and normal development distribution, (b): age distribution

Evaluation

Evaluation Metrics

In order to evaluate the performance of different models, evaluation metrics of accuracy, recall, precision, F1-Score, and confusion matrices were utilized.

	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Figure 8. Confusion matrix

A confusion matrix, mainly employed in classification tasks, is a visualization of comparing the ground truth labels with the model's predictions. If the ground truth categories are divided into a positive class and the rest as negative classes, correctly predicted positive samples would be denoted as true positive (TP), while samples that were incorrectly predicted as positive would be denoted as false positive (FP). Similarly, samples from the negative classes that are correctly identified would be true negative (TN) and samples that are incorrectly predicted as negative would indicate false negative (FN).

Equation 2. Accuracy

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

Utilizing these four values, the following scores of accuracy, recall, precision, and F1-Score can be calculated.

Accuracy is the percentage of correctly predicted samples out of the total samples in the dataset. As indicated in Equation 2, it is computed by adding up all the correctly predicted samples (TP and TN), then dividing it by the total number of samples.

Equation 3. Precision

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

Precision, also known as the positive predictive value (PPV), is the proportion of correctly predicted positive samples out of all samples that the model has predicted to be of the positive class. In simpler words, the score evaluates how many of the model's positive predictions are actually positive. This value is calculated by dividing the TP value by the sum of TP and FP.

Equation 4. Recall

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

Recall, or the true positive rate (TPR), is calculated by dividing TP by the sum of TP and FN. This score analyzes the number of correctly predicted positive samples out of all positive samples. It asks the question, how many of the positive samples did the model guess correctly.

Equation 5. F1-Score

$$F1-score = \frac{2 \cdot R \cdot P}{R + P}$$

Here, ***R*** denotes the recall value while ***P*** denotes precision. The harmonic mean of precision and recall scores is known as the F1-Score. This metric is used because it is an effective method of balancing the two scores. In some cases, precision scores may be misleading partly due to the fact that only the true positive values are considered. Thus, analyzing the F1-Score that balances precision and recall aids in better evaluating the network.

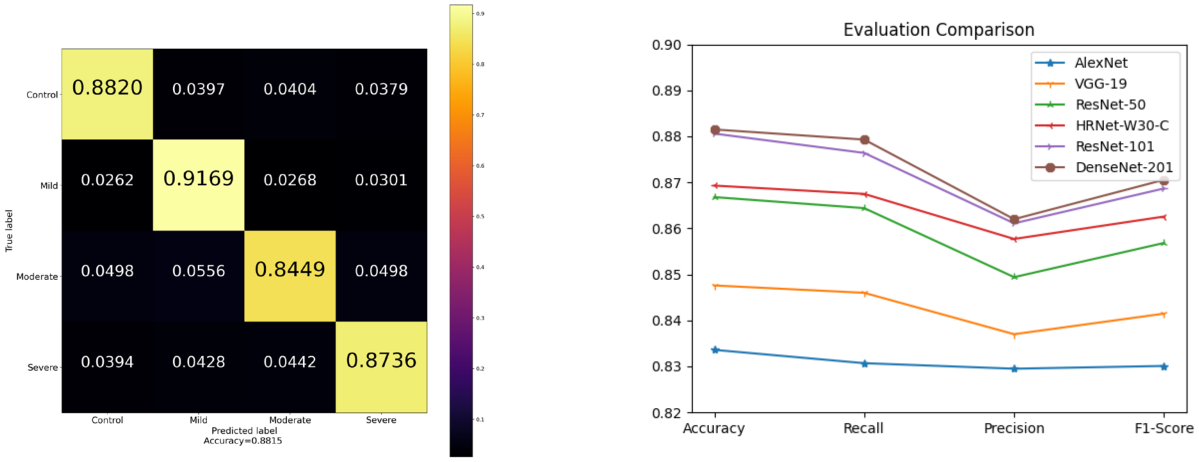
Performance Comparison

Experiments on six different models were conducted, and each of their accuracy, recall, precision, and F1-Score was recorded in Table 1. Overall, DenseNet-201 performed best compared to all other models, with an accuracy of 88.15. Using this network, a confusion matrix was plotted, comparing the ground truth labels with the model's predictions. Each axis comprised four categories: Control, Mild, Moderate, and Severe. As shown in Figure 9 (a), the model performed especially well in the Mild category, while its accuracy was comparatively lower in the Moderate category.

Evaluation comparison was further conducted through a graphical visualization, presented in Figure 9 (b). The line of DenseNet-201 was the highest compared to the rest of the networks, closely followed by ResNet-101. However, it is important to note the comparatively low precision scores than other evaluation metrics in all models. This seemed to be partly due to the training dataset imbalance, where there were considerably more data samples with patients diagnosed with ADHD compared to Normal, as shown in Figure 7 (a). Thus, the model might have tended to predict samples as positive without thoroughly reviewing the features given.

Table 1. Table representation of experimental results

	Accuracy	**Recall**	**Precision**	**F1-Score**
AlexNet (Krizhevsky et al. 2012)	83.36	83.07	82.95	83.01
VGG-19 (Simonyan et al. 2014)	84.76	84.60	83.70	84.15
ResNet-50 (He et al. 2016)	86.68	86.44	84.94	85.68
HRNet-W30-C (Wang et al. 2020)	86.93	86.75	85.77	86.26
ResNet-101 (He et al. 2016)	88.06	87.64	86.11	86.87
DenseNet-201 (Huang et al. 2017)	88.15	87.93	86.20	87.06



(a) (b)

Figure 9. (a): Confusion matrix visualization of experimental results and (b): Graphical representation of experimental results

Ablation Study

To understand and analyze each feature's importance and contribution to the model's predictions, an ablation study was conducted with two different methods. The first operation that was conducted is called dilation, which enlarges the input image at a certain scale for a chosen number of iterations. As shown in Table 2, for three different models, a total of eight dilations were implemented. The accuracies that corresponded to different numbers of dilations as well as the baseline input and full fundus photography were compared. The results indicate that there was no significant difference in accuracy as the number of dilation operations changed, with a maximum difference of 0.03. Thus, the enlargement of retinal vessels and optic disc features did not seem to improve nor worsen the performance of the model. Figure 10 that depicts these results in a graphical form visually highlights the insignificance of difference in accuracy.

Table 2. Result of the implementation of dilation operations

Dilation Operation	**Accuracy (baseline)**	**Accuracy (dilation-2)**	**Accuracy (dilation-4)**	**Accuracy (dilation-6)**	**Accuracy (dilation-8)**	**Accuracy (fundus)**
HRNet-W30-C	86.93	86.91	86.9	86.93	86.92	86.92
ResNet-101	88.06	88.05	88.06	88.06	88.07	88.08
DenseNet-201	88.15	88.15	88.14	88.14	88.14	88.14

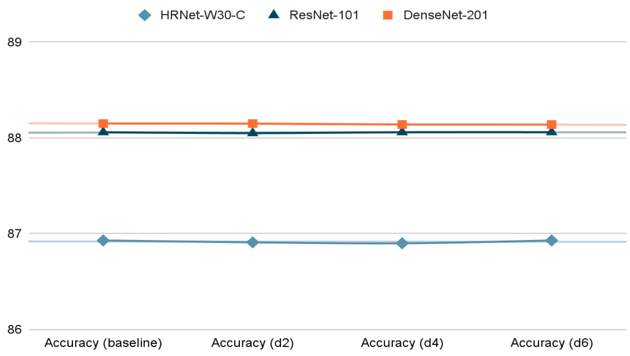


Figure 10. Graphical representation of the implementation of dilation operations in three different models

Similar to the first method, the second operation of erosion is also a type of morphological processing that shrinks the features of the input fundus at a certain scale. The same experimental method as the dilation operation was carried out, comparing the accuracy of the baseline model to four different iterations of erosion operations: two, four, six, and eight times of erosion applied respectively. The results noted in Table 3 were in great contrast with the previous operation. As more erosion operations were applied to the baseline input, the accuracy of all three models decreased substantially. Figure 11 depicts the drastic drop in accuracy, where a mere implementation of two erosions resulted in an accuracy of less than 50 percent. Such outcome implies that the retinal vessels that were eroded at a greater level compared to the optic disc as shown in Figure 6 (d)-(f) plays a great role in aiding the model's predictions.

Figure 12 sums up the ablation study by comparing the two different morphological operations as well as the baseline and original inputs. It denotes a graph that plots the accuracy of three different models for the original fundus, a modified image with two dilation operations done, the baseline input with the segmented optic disc and retinal vessel, and the modified image with two erosion operations done, respectively.

Table 3. Result of the implementation of erosion operations

Erosion Operation	Accuracy (baseline)	Accuracy (erosion-2)	Accuracy (erosion-4)	Accuracy (erosion-6)	Accuracy (erosion-8)
HRNet-W30-C	86.93	37.35	25.23	19.42	18.23
ResNet-101	88.06	38.21	26.72	20.05	19.24
DenseNet-201	88.15	38.32	27.23	20.42	19.55

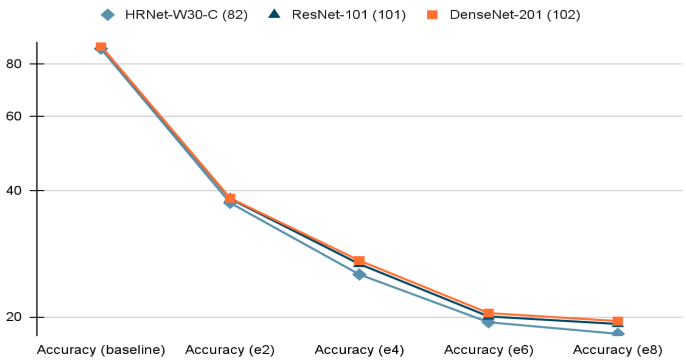


Figure 11. Graphical representation of the implementation of erosion operations in three different models

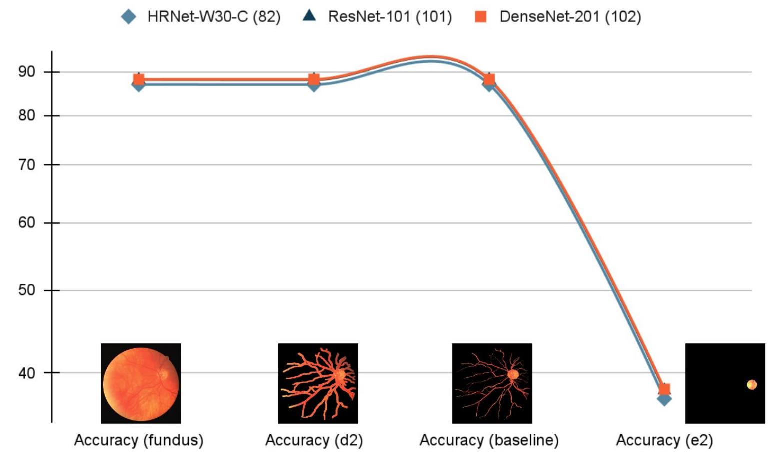


Figure 12. Compilation of the original and modified inputs and their corresponding accuracies

Conclusion

In this research, I proposed an artificial model that is capable of diagnosing whether a patient has ADHD, a specific type of neurodegenerative disorder, as well as assessing its severity in individual patients by analyzing morphological ocular features such as the retinal vessels and the optic disc from fundus photography. I conducted an experiment that modified the original fundus photography into a segmented map containing the retinal vessel and optic disc information, which was further inputted into six different CNN architectures with varying depths. Results demonstrated that DenseNet-201 scored the highest accuracy of 88.15 as well as retaining the highest scores for three other evaluation metrics out of all models tested. In order to confirm which features of the fundus had the highest contribution in aiding the model's predictions, an ablation study leveraging two morphological operations of dilation and erosion was done. Especially focusing on the erosion operation, results pointed out that the loss of retinal vessel information due to repeated erosions substantially reduced model accuracy. Therefore, it was inferred that the retinal vessels are important features that have high correlation to the occurrence of ADHD.

Once further developed, this proposed network may be a valuable tool for clinicians in aiding their tasks of diagnosing ADHD as well as assessing its severity in diverse patients in a more accurate and timely manner. Furthermore, it offers an accurate, cost and time effective, and easily accessible means of diagnosing ADHD, especially for people who may not have access to medical resources and professionals.

Acknowledgments

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

References

AI Hub. (2024, Mar 19). "*Fundus NDD Dataset*": AI Hub. https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=221

Baldi, P. (2012, June). Autoencoders, unsupervised learning, and deep architectures. In Proceedings of ICML workshop on unsupervised and transfer learning (pp. 37-49). JMLR Workshop and Conference Proceedings.

CDC (2024, May 23). "*Data and Statistics on ADHD*": CDC https://www.cdc.gov/adhd/data/?CDC_AAref_Val=https://www.cdc.gov/ncbddd/adhd/data.html

Charters, Lynda. (2022, Dec 27). "*ICYMI: Smartphone-based fundus camera provides option for glaucoma screening*" https://www.ophthalmologytimes.com/view/icymi-smartphone-based-fundus-camera-provides-option-for-glaucoma-screening

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

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708). https://doi.org/10.48550/arXiv.1608.06993

Kim, J. H., Hong, J., Choi, H., Kang, H. G., Yoon, S., Hwang, J. Y., ... & Cheon, K. A. (2023). Development of deep ensembles to screen for autism and symptom severity using retinal photographs. JAMA Network Open, 6(12), e2347692-e2347692.

Ostrowski, Stacey. (2023, Aug 1). "*Fundus*": All About Vision" https://www.allaboutvision.com/eye-care/eye-anatomy/fundus/

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

ZAHMACIOĞLU, O., & Kilic, E. Z. (2017). Early diagnosis and treatment of ADHD are important for a secure transition to adolescence. Anatolian Journal of Psychiatry/Anadolu Psikiyatri Dergisi, 18(1).