HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 4 (2024)

Enhancing Autism Spectrum Disorder Screening through Learning Geometric Features of Retinal Blood Vessels Using Graph Representation

Junseok Lee[1] and Tajvir Singh[#]

[1]North London Collegiate School Jeju, Republic of Korea
[#]Advisor

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurological and developmental disorder that has effects on how people behave and interact. It is a lifelong condition, and those born with it show symptoms typically during one's early childhood. Over the past two decades, the cases and prevalence of ASD have been on a steady increase, with those born in the United States in 2012 being around 4 times more likely to have ASD compared to those born 20 years before. The treatment of ASD is shown to be more effective at earlier ages, which is why being able to perform an early diagnosis is crucial for those born with the condition. However, most of the current tests for child ASD diagnosis rely on the analysis of behavioral patterns by other people. This becomes more of an issue the younger the child gets, because at the youngest ages, there are not many clear actions or traits that a baby can physically show, making early subjective judgment vague and unreliable . To address the aforementioned problem, I proposed a rapid ASD screening framework that utilizes geometric patterns from non-invasive retinal blood vessel scans. The framework contains two modules, the segmentation network and the ASD prediction network. The segmentation network processes the raw retinal images and generates a binary vessel segmentation map, which is then used for the ASD prediction network for a final prediction of the severity, divided into 4 stages. Through experimental results, it is demonstrated that the proposed approach is feasible by achieving an accuracy of 94.2% on a public dataset.

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by a range of symptoms that impact social interaction and communication and cause varying degrees of sensitivity to sensory stimuli. ASD typically becomes evident in early childhood, with many children showing symptoms between 12 and 18 months of age, and it persists throughout a person's life. The severity of symptoms can vary widely among individuals with ASD. Over the past two decades, the prevalence of autism has steadily increased worldwide. In the United States, children born in 2012 were found to be more than four times as likely to have ASD compared to those born in 1992 (Richter, 2024). A similar increase has been observed in South Korea, where the number of ASD cases rose nearly 40% between 2017 and 2021 (Statistics Bank, 2022). In the United States, ASD is commonly diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) by the American Psychiatric Association. Screening often involves checklists and questionnaires, such as the Ages and Stages Questionnaire (ASQ) and the Modified Checklist for Autism in Toddlers (M-CHAT), which generate a total score based on a child's behavior as reported by their legal guardian or parent.

Whilst these methods are sound in many aspects, they also contain many issues that may introduce errors to the accuracy or effectiveness of the diagnosis. One large concern is the reliance on subjective reports from parents or guardians, which can introduce bias to certain decisions, especially if the caregiver lacks awareness or understanding of typical development milestones for a child.

In this research, I proposed a rapid ASD screening framework based on a medically driven approach utilizing retinal blood vessel patterns. The framework is composed of two modules: a retinal vessel segmentation network and ASD prediction network. The retinal vessel segmentation network processes retinal images to generate isolated vessel segmentation maps which are then used to train the ASD prediction network.

Related Work

Retina Photography

Retina photography, also known as fundus photography or imaging, is a specialized form of photography used to capture detailed images of the interior surface of the eye, mainly the retina. The retina is a light-sensitive layer at the back of the eye that converts light into neural signals. Photography is done through a flash camera paired with a specialized microscope that lets it capture the interior layer of the eye via a series of lenses. This can be done with both an actual fundus camera, as well as a portable set of commercial lenses paired with a smartphone. As such, the procedure can be carried out at a very low cost, and is an accessible form of medical imaging that can be done outside of professional clinics, as it is a non-invasive procedure that has virtually no risks involved.

(a) (b)

Figure 1. Example of retina photography and retinal vessel segmentation
(a): retina photography and (b): vessel segmentation map

The retinal blood vessels seen in a retinal photograph enter the eye along with the optic nerve, which is directly connected to the brain. This places the retinal blood vessels among the blood vessels closest to the brain. As such, whilst not being directly connected, changes or abnormalities in the retinal blood vessels can be indicators of health issues affecting the brain. The relationship between these vessels and the presence of ASD is currently an ongoing subject of medical research, with many studies suggesting that there may be differences in the retinal blood vessel patterns of developing individuals with ASD and without. Inspired by previous studies in this field, I developed an ASD screening system that employs a medically driven approach which leverages retinal blood vessel information.

Image Classification

Image classification is a task in vision recognition that aims to assign a class or label to one entire image. It takes in a single-object image as input and outputs a prediction about which predefined class the image belongs to as shown in Figure 2.

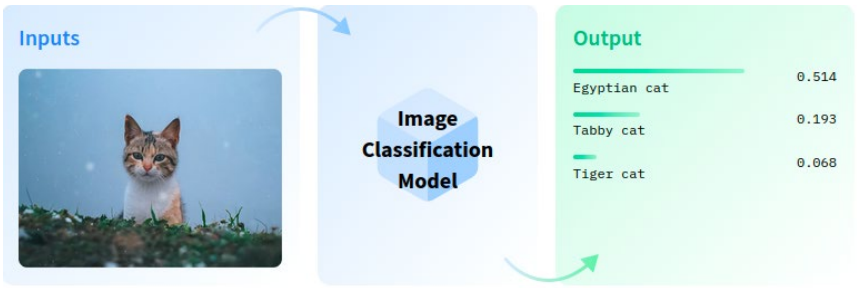

Figure 2. Explanation of image classification (Hugging Face 2023)

Convolutional neural networks specifically designed to handle data in the form of grids, mainly images, by having filters slide across every position on the image. Different filter sizes allows CNNs to learn discriminative features of the image at different levels of detail, giving them the ability to distinguish between objects with different features. As such CNNs are widely used for various computer vision tasks, most notably image classification. Similarly, ASD screening can be considered a type of image classification system where the input is a retinal image and the output is one of four ASD severity categories: normal (healthy control group), mild, moderate, and severe. Detailed information about the proposed framework is provided in Chapter 3.

Proposed Method

Retinal Vessel Segmentation Network

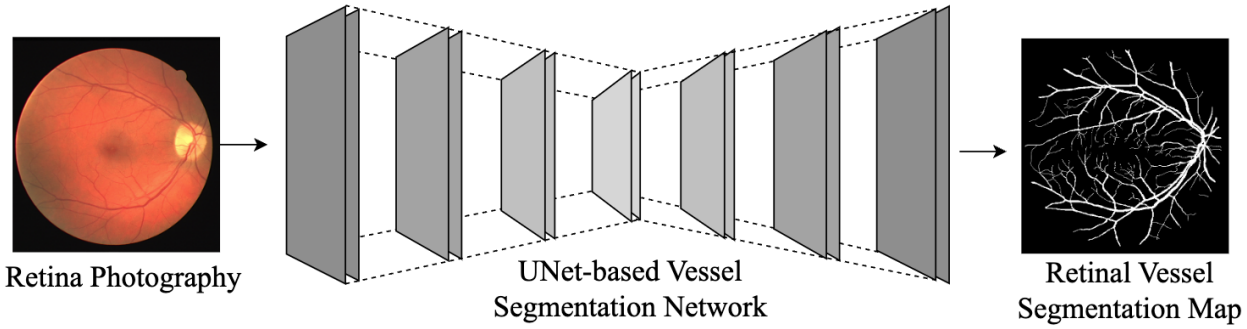

Figure 3. Architecture of the retinal vessel segmentation network

Figure 3 illustrates the overall architecture of the proposed retinal vessel segmentation network. The vessel segmentation network is composed of an encoder and a decoder. The encoder takes a retinal photograph as input and generates feature maps which captures the different details of the input image. The decoder takes these feature maps to produce the final vessel segmentation map.

To implement the proposed encoder and decoder, I utilized convolutional neural networks based on the U-Net model(Ronneberger et al. 2015), which is commonly used in biomedical image segmentation. The encoder includes downsampling layers to capture increasingly higher level features of the original retinal photograph whilst leading to a reduction in its spatial dimensions. On the other hand, the decoder involves upsampling layers to reconstruct the final segmentation map with the original resolution of the input retina image.

To train the networks, I utilized pixel-wise binary cross entropy as the segmentation task involved the separation of the input image into two different categories, hence requiring the use of binary classification. Each image consists of the background, which is the back of the retina, and the foreground being the vessel structure itself.

Equation 1: Pixel-wise binary cross entropy

$$L_{(x,y)} = -[y(x,y) \times ln(\hat{p}(x,y)) + (1 - y(x,y)) \times \ln(1 - \hat{p}(x,y))]$$

Equation 2: Pixel-wise binary cross entropy (total image pixel)

$$L = -\frac{1}{XY}\sum_{x \in X}\sum_{y \in Y} y(x,y) \times \ln(\hat{p}(x,y)) + (1 - y(x,y)) \times \ln(1 - \hat{p}(x,y))$$

Here, y(.) and $\hat{p}$(.) denote the true value and the predicted value for the pixel of that position, respectively. Equation 1 is the equation for the binary cross entropy loss of an individual pixel. Because the true value of the results has to be either a 1 or a 0, one of the added halves of the equation will result in 0, allowing it to perform different calculations for the loss depending on the true output.

Autism Spectrum Disorder Prediction Network

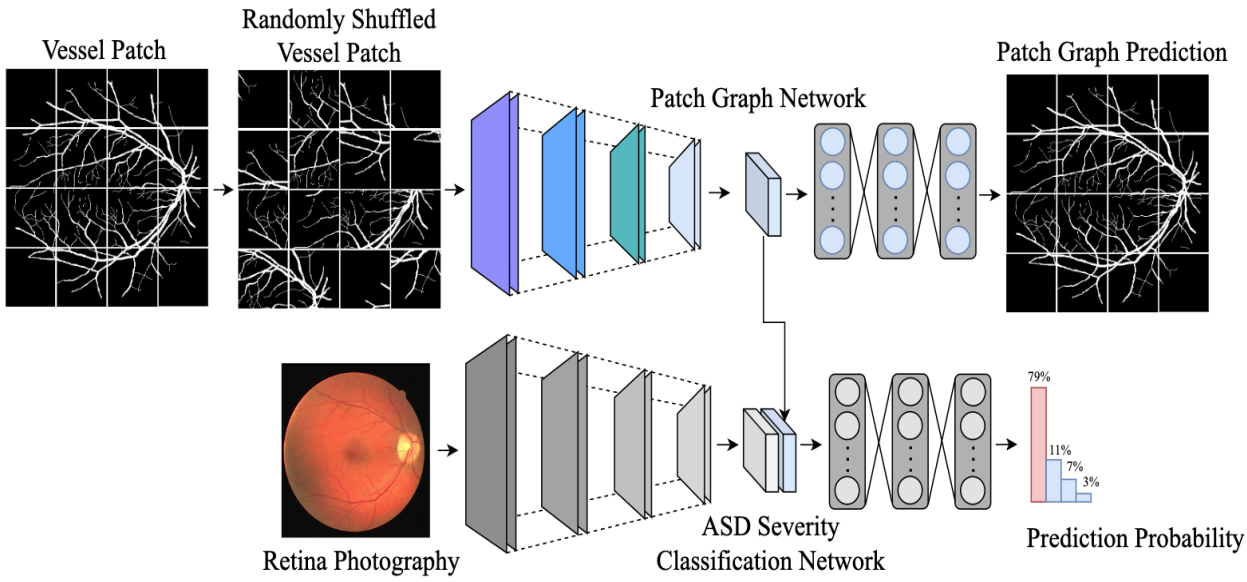

Figure 4. Architecture of the autism spectrum disorder prediction network

The proposed method, as shown in Figure 4, makes use of a new network trained to reconstruct a shuffled vessel patch image to learn the geometric features of the vessels in greater depth. The feature maps from the patch graph network are used alongside the original severity classification network so that more detailed information about the vessels' structure can be used in the training process. By doing so, the aim is to enhance the accuracy of the testing results through the application of more detailed data during training.

To train the final classification network, I utilized cross entropy loss functions as shown in Equations 3 and 4. For the overall loss of the patch graph predictions, the natural logarithm for the probability of correct prediction is

calculated for each individual segment of the image. The negative average for all the segments is the final loss value of the entire patch graph. The loss value of the final severity prediction for the retinal image is calculated as the negative natural logarithm of the probability for the true class of ASD severity.

Equation 3: Cross entropy loss function for path graph prediction

$$L_{graph} = -\frac{1}{N^2}\sum_{i \in N^2} \ln(\hat{q}_i)$$

Here in equation 3, N denotes the number of horizontal or vertical segments created after being divided into squares. As such, N^2 is the total number of divided squares in the image. $\hat{q}$ denotes the probability predicted for the actual correct value of that square.

Equation 4: Cross entropy loss function for severity prediction

$$L_{severity} = -ln(\hat{r})$$

In equation 4, $\hat{r}$ stands for the probability calculated for the true severity of ASD in the given image.

Experimental Results

Autism Spectrum Disorder Dataset

The total sample size of data used for training and testing was 57,195 retinal images. From here, 35.06% were diagnosed with some form of ASD, while the other 64.94% were part of the control group. Of this population, 27.7% were children below the age of 7, 43.57% were between 7 and 13, whilst 28.73% were between 13 and 21. 80% of the total data was used in training, and the rest was used during testing.

Model Evaluation

Table 1. Evaluation comparison for five convolutional neural network architectures

	Accuracy	Recall	Precision	F1-Score
VGG-16 (Simonyan et al. 2014)	0.8846	0.8508	0.8356	0.8431
ResNet-18 (He et al. 2016)	0.8977	0.8528	0.8354	0.8440
ResNeXt-50 (Xie et al. 2017)	0.9149	0.8643	0.8474	0.8558
Xception (Fran et al. 2017)	0.9272	0.8661	0.8532	0.8596
DenseNet-264 (Huang et al. 2017)	0.9420	0.8887	0.8716	0.8801

The testing of the proposed model was conducted using 5 different convolutional neural networks of increasing depths. 4 different metrics of quality were recorded for the tests on each network. The model with the greatest depth, DenseNet-264, performed the best in all metrics, and the model with the least depth, VGG-16, performed the worst out of all the models, clearly showing that an increase in depth leads to more accurate results. For all models, the recall, which is the amount of correctly predicted cases of positive ASD compared to all cases of positive ASD, was somewhat higher than the precision, which is the amount of correctly predicted positive ASD cases over the predictions of

positive ASD. This suggests that the model will sometimes incorrectly predict cases of ASD even for those in the control group.

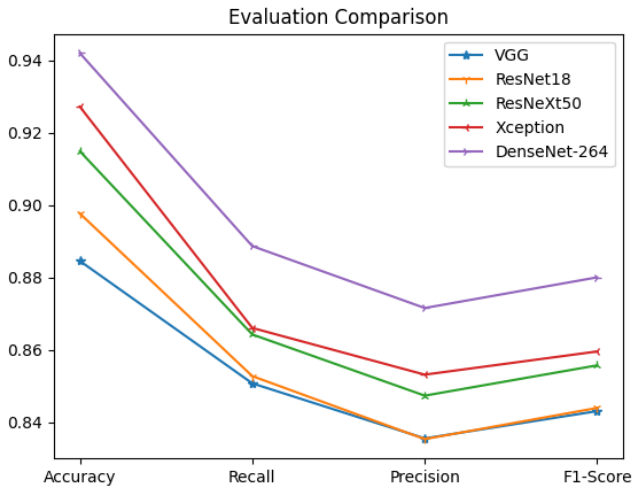

Figure 5. Evaluation comparison for five convolutional neural network architectures

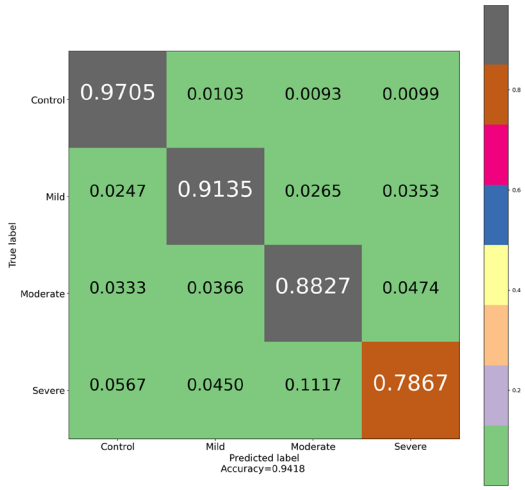

Figure 6. Confusion matrix evaluation

To further analyze the results, a confusion matrix was created for the model with the highest overall accuracy, DenseNet-264. From this, we can clearly see that there is an inversely correlated relationship between the severity of ASD and the percentage of correct predictions for that case. For most cases, the category with the highest percentage of predictions, asides from the correct choice, is generally an option that is higher in severity compared to the other predictions. This supports the previous statement that the model will tend to exaggerate the severity of ASD cases.

Table 2. Ablation study result (vessel graph feature)

	Accuracy (proposed)	**Accuracy (baseline)**
VGG-16	0.8846	0.8524

ResNet-18	0.8977	0.8592
ResNeXt-50	0.9149	0.8792
Xception	0.9272	0.8855
DenseNet-264	0.9420	0.9011

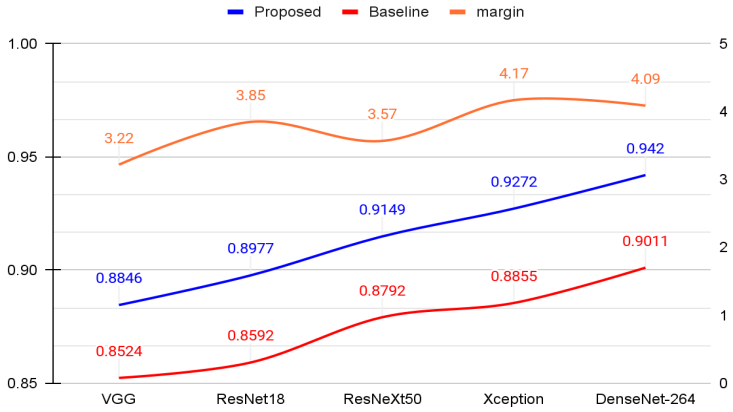

Figure 7. Ablation study result (vessel graph feature)

To measure how much the changes in the proposed method had a positive effect on the results, the same test with all the different networks was conducted for both a model with the proposed method and without. The accuracy was recorded above for both. For all depths, the proposed model performed better than the baseline model by a margin of approximately 3-4% in terms of accuracy. Although there were a few slight dips in between, this margin also showed signs of slight increase along with the depth of the model.

Conclusion

After conducting a series of tests, the following results were obtained. Table 4.2 is an overall evaluation of tests conducted using varying depths of CNNs, using the proposed method of classification. The table keeps the results of each through various different metrics for the measurement of quality. Generally, all results related to the measurement of accuracy increased as the models' depth increased. The confusion matrix of the results shows that, whilst generally accurate, the accuracy of the model decreases as the condition becomes more severe. Whilst the rate of correctly predicting the state of a control group member is 97%, this drops to below 80% when attempting to correctly diagnose patients with severe ASD. The ablation study clearly demonstrates that the proposed approach is roughly 3-4% more accurate compared to the baseline. This margin also seems to make a slight increase as the depth of the models do so as well.

Acknowledgments

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

References

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

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

Hugging Face. (2023, Sep 5). "*Image Classification*": Hugging Face. https://huggingface.co/tasks/image-classification

Richter, F. (2024, Apr 2). "*The Rising Prevalence of Autism*": Statista. https://www.statista.com/chart/29630/identified-prevalence-of-autism-spectrum-disorder-in-the-us/

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

Statistics Bank. (2022, Aug 30). "*Autism in 2.6% of children aged 7 to 12 in South Korea: General social statistics*": Statistics Bank. https://www.xn--989a170ahhpsgb.com/15/?q=YToxOntzOjEyOiJrZXl3b3JkX3R5cGUiO3M6MzoiYWxsIjt9&bmode=view&idx=12724994&t=board

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500). https://doi.org/10.48550/arXiv.1611.05431

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.