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

Multi-Resolution Image Features of Retinal Images and Optic Nerve Head for Biomarker Identification in Attention-Deficit/Hyperactivity Disorder

Taehyeon Hwang[1] and Elizabeth McCook[#]

[1]Urbana High School, USA
[#]Advisor

ABSTRACT

Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder characterized by a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development. Over the past 20 years, the number of teenagers diagnosed with ADHD has increased dramatically. This trend is not limited to the United States but is observed globally. Early diagnosis of ADHD is important because it allows for earlier treatment which can significantly improve symptoms and overall outcomes. Traditionally, diagnosing ADHD has relied heavily on self-checklists and observation from parents or teachers. However, these methods are often inaccurate, unscientific, and prone to subjective error. To solve this problem, in this research, I proposed a machine learning-based systematic ADHD diagnosis approach. The proposed system is developed with multi-resolution retinal image features and optic nerve head segmentation. Multi-resolution technique enables the model to learn from the input data in various perspectives. This results in rich feature extraction, in which the model gains deeper understanding and captures underlying patterns in the data. Optic nerve head segmentation enables the model to analyze one of the significant biomarkers of ADHD carefully in depth and have a synergistic effect in increasing the accuracy of the model when it is used with the multi-resolution technique. The experimental result of the proposed system achieved promising accuracy of 86.88% on ADHD retina image dataset.

Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder, which can affect functioning and development, characterized with inattention, hyperactivity, and impulsivity. It is often diagnosed in childhood and affects their daily life including academic, occupational, and social functioning. Several factors such as genetics, brain structure, environment, and diet are suspected as the cause of ADHD. Earlier researches have found that genetics plays the most significant role in developing ADHD. Early diagnosis and intervention can significantly improve the management of symptoms and enhance the quality of life for individuals with ADHD because it is possible to perfectly cure it in childhood, while it is not possible to cure and have to take medicine for the rest of their lives in adolescence.

Traditionally, Magnetic Resonance Imaging (MRI) has been used to diagnose ADHD, but in most of the cases, a hospital with an MRI device is not prevalent and even if there is, it's costly. Thus, accessibility to those high quality medical care is very limited. Additionally, other diagnosing methods such as EEG, which measures brain signal of the patient; checklist, which measures patient's behavior; and computer aid, which compares the patient with known parameters of ADHD, are used. However, these methods are often costly, time-consuming, hard to access, and prone to error. Also, without any scientific reasons in diagnosis, most of the parents tend to think their child is just more active than others and do not use drugs on time.

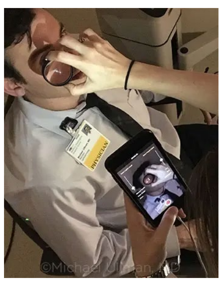
Recent studies have introduced new methods that utilize fundus images for the rapid screening of ADHD. Fundus images, which capture the back of the eye, are relatively easy to obtain compared to other medical images.

They can even be captured at home using a special lens attached to common smartphones. Kim et al. proposed a machine learning-based system for diagnosing Autism Spectrum Disorder (ASD) using fundus images. Their method successfully demonstrated the feasibility of using fundus images for screening ASD, a common neurodevelopmental disorder similar to ADHD. Inspired by their study, I propose a network for screening ADHD based on fundus images. The proposed network employs a medically-driven approach by leveraging optic nerve head information to enhance accuracy.

Background Knowledge

Fundus Image

The fundus image is a photograph of the back of the eye. Currently, fundus image is a rising method for screening neurodevelopmental disease. It is a valuable medical image since it is a non-invasive method to capture blood vessels that are directly connected to the brain, containing information that can be used in diagnosing brain diseases.



(a)

(b)

Figure 1. Taking a fundus image and an example fundus image. (a): A smartphone-based funduscopy (Haddock et al. 2018) and (b): Fundus Image

In figure 1(a), taking a fundus image with a smartphone and a lens is shown. Turn on video mode and a flash, holding a lens with another hand, start recording with a zoom feature to focus. Then, capture a screenshot from the video to capture a fine picture of the fundus. This high accessibility is also a reason why fundus image is medically valuable, it can be easily taken in households without expensive medical devices.

The fundus image mainly captures three features: optic disc, optic cup, and retinal blood vessel. In figure 2(b), an example of a fundus image is shown. Optic disc is a circular area at the middle of the eye, optic cup is a smaller circular area inside the disk with a brighter color, and retinal blood vessels are lines that stretch from the disc and cup. It is found through clinical experiments that patients with ADHD tend to have more curved retinal blood vessels and smaller optic disc to cup ratio.

Object Segmentation

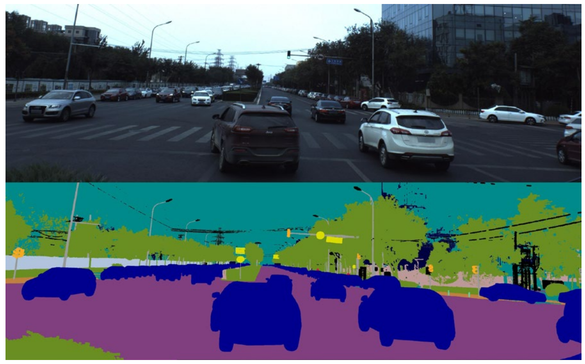


Figure 2. CVPR 2018 WAD Video Segmentation Challenge (Alibaba Cloud 2021)

Object segmentation is a process of extracting valuable objects for a certain task. Because AI does not have insights as humans do, it is important to have it able to identify some of the key features from the input data to perform its task. An example of object segmentation is shown in figure 2. The AI identifies objects such as cars, trees, traffic lights, traffic signs, and streetlights. One of the most famous implications of object segmentation is Tesla's autonomous driving technology. In the case of diagnosing ADHD, object segmentation is used to identify key features such as retinal blood vessels, optic disc, and optic cup from the fundus image. Once these features are clearly defined, the AI examines specific structure details and interprets through to classify whether the patient is ADHD or not.

Proposed ADHD Classification Network

This chapter provides a detailed explanation of the proposed method including loss function and training strategy. The proposed system is composed of two modules: optic nerve head segmentation and multi-resolution based ADHD classification.

Optic Nerve Head Segmentation

Optic nerve head segmentation is a process of extracting the optic disc, where retinal blood vessels converge to exit the eye and transfer visual information to the brain, from a fundus image. As acknowledged before, the optic nerve head is a significant biomarker for identifying ADHD, so by isolating it, the trained model can learn closer about characteristics in morphological features with wider perspectives.

For the segmentation process, I used a state-of-the-art segmentation architecture proposed by Ronneberger et al. shown in figure 3. Segmentation architecture is divided mainly into two processes: encoding and decoding. In the encoding process, the input data gets compressed into lower dimension through downsampling; while, in the decoding process, the data gets back to its original dimension through decoding. Due to these processes, the architecture gets its U-shape, and is called as U-Net. Also, the architecture has skip connections that prevents loss of information layer to layer during the encoding process, and outputs more accurate results.

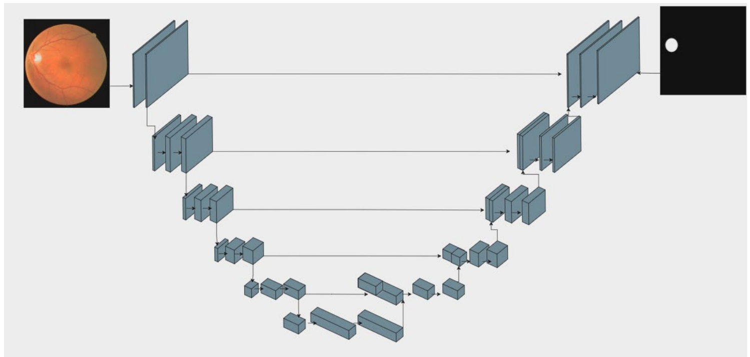


Figure 3. U-Net based optic disc segmentation network (Ronneberger et al. 2015)

With the segmented optic nerve head image, I developed a distinct network in my model to specifically analyze and capture differences in the structure of optic nerve head between an ADHD patient and healthy individuals. This advanced analysis increases the accuracy of the model's classification.

Equation 1: Segmentation loss

$$L_{seg} = -\frac{1}{WH}\sum_{w\in W}\sum_{h\in H} y(w,h)\times\ln\left(\hat{p}(w,h)\right) + (1-y(w,h))\times\ln\left(1-\hat{p}(w,h)\right)$$

For training the segmentation network, I utilized segmentation loss function which calculates how accurately the optic nerve head was segmented from the input retina image data. Equation 1 is the mathematical representation of segmentation loss function. Here, $\boldsymbol{y}$ denotes the ground truth value of the pixel being foreground or background and $\hat{\boldsymbol{p}}$ denotes the predicted value of the trained model.

Proposed Multi-Resolution Based ADHD Classification System

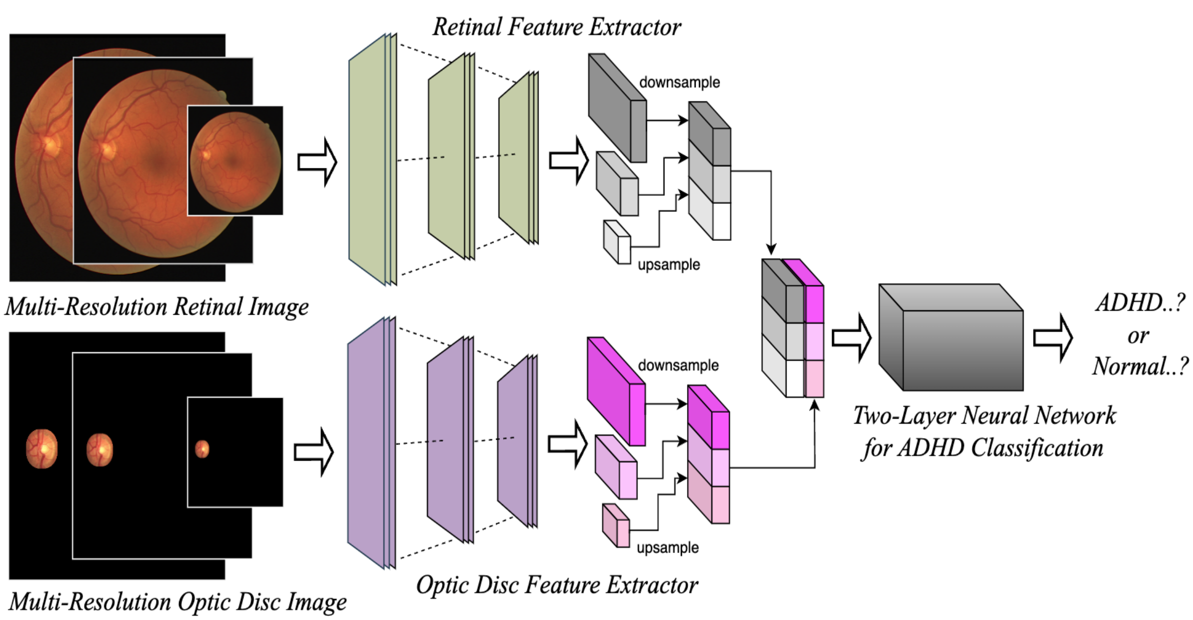


Figure 4. Overall Architecture of the Proposed Method.

In figure 4, the overall architecture of the proposed model is shown. The proposed model takes the fundus image as an input and outputs whether the patient is ADHD or not. The proposed method consists of three modules: retinal feature extractor, optic disc feature extractor, and two-layer neural network for ADHD classification.

First, a multi-resolution retinal image, the same image with three different sizes, is created from the input fundus image in three different sizes of 512 by 512, 256 by 256, and 128 by 128. These images are then placed in the retinal feature extractor to extract feature maps. Because feature maps from small and large images have different sizes, upsampling for a small feature map and down sampling for a large feature map is conducted in order to concat three different feature maps from multi-resolution retinal image into one.

At the same time, a multi-resolution optic disc image, also in three different sizes with the same ratio, is placed in the optic disc feature extractor. Upsampling for small feature maps and downsampling for large feature maps is conducted to concatenate them into one feature map.

Next, a feature map from retinal feature extraction and optic disc feature extraction are concatenated into one. Then, the concatenated feature maps are placed into a two-layer neural network to classify whether the owner of the input fundus image is ADHD or not. This unique architecture of having two networks for feature extraction of multi-resolution retinal and optic disc image allows the model to examine from a wider range of perspectives with one image, which leads the model to have rich information. Also, having a distinct network to analyze the optic neural head, which is a key component for identifying ADHD, allows the model to have improved accuracy.

Equation 2: Cross-entropy loss function

$$L_{cross} = -ln\ P$$

For training the proposed model, I utilized the cross-entropy loss function, which measures how accurately the model has predicted compared to the ground truth value in the range of 0 to infinity with a value closer to 0 being more accurate. The mathematical expression of cross-entropy loss function is shown in equation 2. Here, P denotes the probability value of the predicted presence of ADHD.

Experimental Results

Retinal Image Dataset

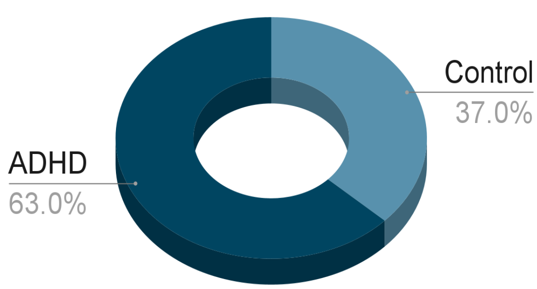


Figure 5. Category distribution of the dataset

The dataset used in this research was collected from the hospitals in South Korea, containing a total number of 100,395 retinal images of people with age ranging from 3 years old to 20 years old. As shown in figure 5, 37%(37,145) of retinal images were of normal people and 63%(63,250) of images were of people with ADHD (AI Hub 2024).

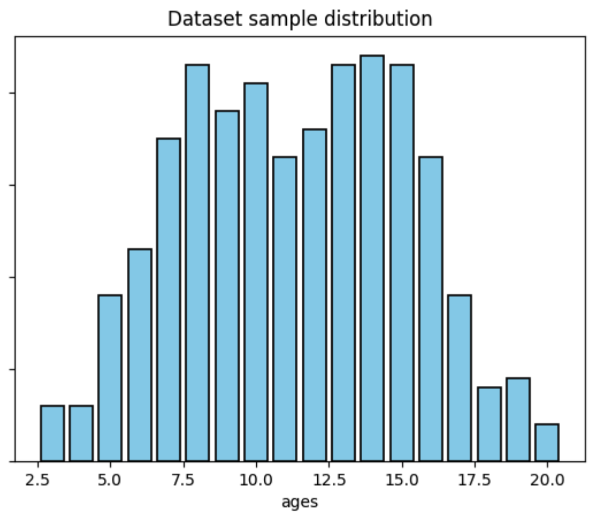


Figure 6. Sample age distribution of the dataset

In figure 6, a distribution of retinal images within different ranges of age groups is shown. Most of the data is collected from elementary school kids and teenagers. The dataset is appropriate to be used in the research as it has sufficient number of retinal images from different age groups especially from young kids to teenagers that the research is targeting to diagnose ADHD to prevent childhood ADHD to develop into adolescence ADHD, which is incurable.

Evaluation Results

Table 1. Accuracy of the model with different networks

	Accuracy	Recall	Precision	F1-Score
ResNeXt-50 (Xie et al. 2017)	81.37	81.47	80.28	**80.87**
Xception (Fran et al. 2017)	81.51	81.52	80.46	80.99
ResNeXt-101 (Xie et al. 2017)	82.82	82.77	82.12	82.12
DenseNet-264 (Huang et al. 2017)	84.86	85.07	83.66	**84.36**

In table 1, the accuracy of four different experiments using convolutional neural networks with different depth of layers are compared. Recall is how much the model correctly predicted positive among all the samples, precision is how much the model correctly predicted positive among all positive samples, and F1-score is recall and precision combined. ResNeXt-50 with 50 layers had accuracy of 81.37%, Xception with 71 layers had accuracy of 81.51%, ResNeXt-101 with 101 layers had accuracy of 82.82, and DenseNet-264 with 134 layers had accuracy of 84.86%.

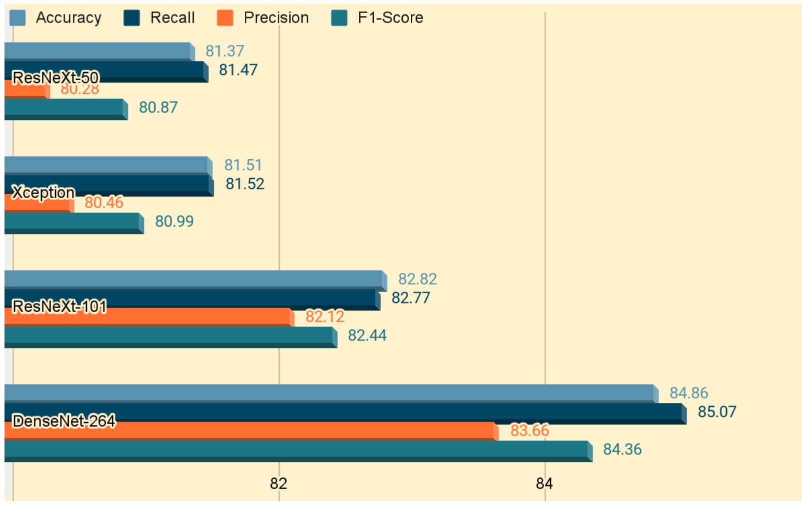


Figure 7. Accuracy, recall, precision, and F1-score of the model with different networks

As shown in figure 7, it is observed that as the network gets deeper, the accuracy of the model increases. Also, there was a pattern in which all of the networks had lower precision compared to recall. This is a sign that the networks are working well. Lower precision is better than lower recall because it is better to diagnose ADHD to a person who does not than diagnosing normal to a person who has ADHD. For example, if the model has determined positive on a patient, but seems to not have ADHD, then additional tests can be conducted for more accurate diagnosis; however, if the model determines negative on a patient who has ADHD, then consequences for failing to early diagnose the disease would hurt the patient.

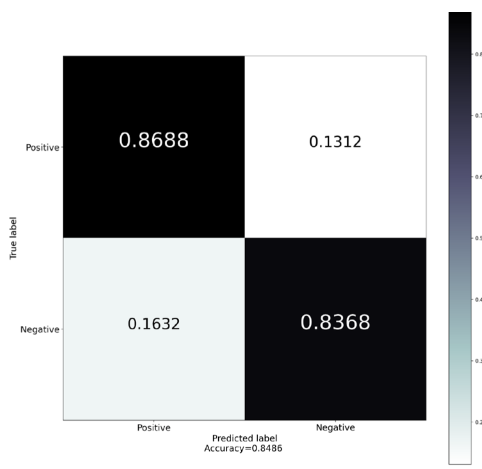


Figure 8. Confusion Matrix

In figure 8, the confusion matrix of the proposed model is shown. There are true labels, which are actual values of the data, and predicted labels, which are predicted values of the data by the model. 0.8688(86.88%) in the up-left box represents accuracy in which the model accurately diagnosed ADHD patients, and 0.8368(83.68%) in the down-right box represents accuracy in which the model accurately diagnosed people without ADHD.

Table 2. Ablation study results

	Accuracy (proposed)	Accuracy (fundus)	Accuracy (optic disc)
ResNeXt-50	81.37	78.32	75.52
Xception	81.51	78.76	76.82
ResNeXt-101	82.82	79.92	76.93
DenseNet-264	84.86	80.87	77.02

In table 2, the result of an ablation study conducted with each network is shown. Three different experiments were conducted: one with all the proposed modules, one with a network only trained with a fundus image, and one with a network only trained with an optic disc image. Experiment with all the proposed method, which trains the model with both fundus and optic disc image, had the highest accuracy among all experiments.

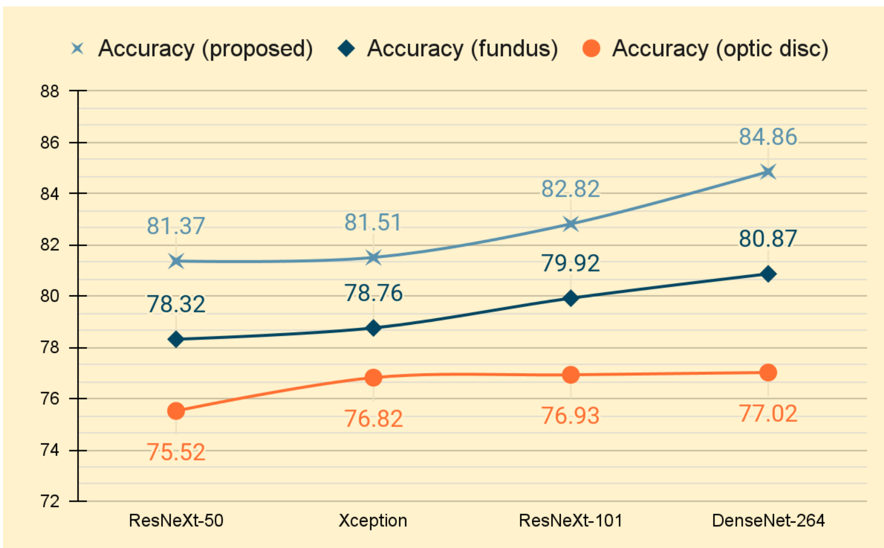


Figure 9. Graph of ablation study results

Here, in figure 9, table 2 is visualized. It is observed that models trained with optic disc image only had the lowest accuracy, and had small growth in accuracy as the networks got deeper because there are limited features for the AI to learn. On the other hand, the models trained with only fundus images had second highest accuracy, and showed a bigger increase in accuracy as the networks got deeper because there are more features to learn about. Still, it doesn't account information in the optic disc enough. Finally, the proposed method, which accounts information from both fundus and optic disc images had the highest accuracy, also it suggests that there is a synergy of training those both data as it showed the biggest improvement in its accuracy as the networks got deeper.

Conclusion

In this research, I proposed using multi-resolution image features of retinal images and optic nerve heads as an input data for the model, and through a series of training, the highest accuracy of 84.86% was obtained. Four different convolutional neural network architectures were used to test the proposed method, and as the layer of the network was deeper, the higher the accuracy was. So, among the networks used in the research, DenseNet-264 with the deepest layer of 134 layers had the highest accuracy. Utilizing an additional network for optic nerve head, or optic disc, information highly contributed in increasing the accuracy of the proposed model, and it is proven through ablation study that average 6.065% increased. When the model was trained with both fundus and optic nerve head images, the synergistic effect in which the model learned from the whole view of the fundus and closely analyzed the optic nerve head allowed the model to capture underlying patterns that improved significantly as the network got deeper. In the future, I hope the proposed model and the findings in the research contribute to the development of ADHD diagnosis using machine learning, ultimately preventing developing adolescence ADHD by early diagnosis in childhood.

Acknowledgments

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

References

AI Hub. (2024, Aug 13). "*Neurodevelopmental disorder fundus dataset*": AI Hub. https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=data&dataSetSn=71516

Alibaba Cloud. (2021, Mar 23). "*CVPR 2018 WAD Video Segmentation Challenge*": Alibaba Cloud https://tianchi.aliyun.com/dataset/95516

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

Haddock, L. J., Hendrick, A. M., Pan, C. K., & Ullman, M. (2018). Smartphone funduscopy: a high-tech, low-cost imaging alternative. EyeNet Mag, 2018, 29-31.

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

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. https://doi.org/10.48550/arXiv.1505.04597