HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

Age-guided Cardiovascular Disease Diagnosis from Retinal Images: A Multitask Deep Learning Framework

Ellen Ryu[1], Seunghyuk Lee[2], Youngjae Sim[3] and Devon Marks[#]

[1]Seoul International School, Republic of Korea
[2]Daegu Il Science High School, Republic of Korea
[3]Korea International School, Republic of Korea
[#]Advisor

ABSTRACT

One of the major global health concerns today is cardiovascular disease. A cardiovascular disease is a disease that involves the heart and blood vessels, its most common causes being smoking, alcohol, high blood pressure and blood cholesterol, etc. Added to the fact that it is a major global health issue, the fact that it is incurable once diagnosed makes the disease significantly more severe. Hence, an early diagnosis is crucial in preventing the disease and, if detected, mitigating it. Traditionally, detection for cardiovascular diseases was usually done by computerized tomography scans; however, these methods are too costly and not easily accessible for the general population. Because of these shortcomings of the traditional diagnosis method, there is a need for a more automated, accessible early diagnosis system, and that is the center of this research. In this study, we propose a novel machine learning-based approach for the detection of cardiovascular diseases using retinal images. The proposed method comprises three modules: a feature extractor, an age prediction network, and a cardiovascular disease diagnosis network. We consider the diagnosis of cardiovascular diseases as a classification task, aiming to assess the severity level of the condition based on the retinal image. The proposed method classifies retinal images into one of four categories, each corresponding to a specific level of disease severity. Through extensive experimentation, the results demonstrate that the proposed method outperforms previous state-of-the-art approaches. We expect that the proposed method holds the potential to become a novel machine learning-powered biomarker for diagnosing cardiovascular diseases.

Introduction

Cardiovascular Disease

Cardiovascular disease (CVD) is a disorder that occurs in the heart and blood vessels. CVD normally occurs when there is a build-up of fatty deposits in blood vessels that supply the heart and brain, which leads to the blockade of the flow of blood to the heart (Gaziano et al. 2006). As a result, the most common symptoms are detrimental heart attacks and strokes, which can severely impact one's health. However, CVD has become a growing issue as the number of patients has rapidly increased in the past few years. In 2019, about 17.9 million people died from CVDs, responsible for 32% of the cause of global deaths. CVD has become a serious issue because many people are prone to this disease. Smoking, unhealthy diets, physical inactivity, and obesity are common causes (Jager et al. 2011). Though CVD has become prevalent, it is an untreatable disease. Therefore, early diagnosis is crucial as it can prevent the disease from getting worsening.

Diagnosis of Cardiovascular Disease

Traditionally, diagnosing CVDs mainly relied on computerized tomography (CT) scans, specifically employing the Coronary Artery Calcium Scoring (CACS) method. CACS is a medical imaging technique used to assess the amount of calcium deposits in the coronary arteries, which are the blood vessels that supply the heart muscle with oxygen and nutrients as shown in Figure 1. The presence of calcium deposits in these arteries is a sign of atherosclerosis, a condition characterized by the buildup of plaque (which includes calcium) on the artery walls. High levels of calcium deposits in the coronary arteries can indicate an increased risk of coronary artery disease or other cardiovascular problems.

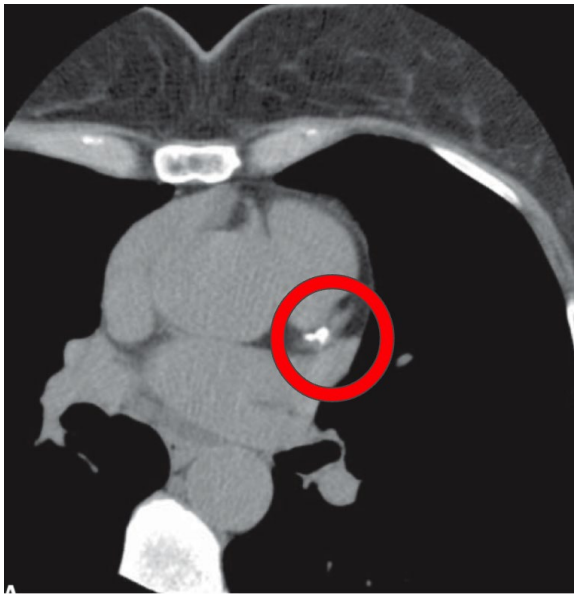

Figure 1. Real life CT scanned image with calcium object detected (red circle). (Sekikawa et al. 2012)

Typically, the CACS results are expressed numerically. A CACS score of 0 means that there is no detectable calcium in the coronary arteries. This suggests that there is little to no plaque buildup in the coronary arteries, indicating a low risk of coronary artery disease or cardiovascular events. Low CACS scores are a favorable result and are typically associated with a lower risk of heart problems. A CACS score of 100 indicates a moderate amount of calcium deposits in the coronary arteries. This suggests the presence of some plaque buildup and an increased risk of CAD compared to a score of 0. However, it is still considered a relatively low to moderate risk, and preventive measures such as lifestyle changes and medications may be recommended. A CACS score of 400 or higher signifies a significant amount of calcium deposits in the coronary arteries. This indicates a higher risk of coronary artery disease and cardiovascular events. Individuals with a CACS score of 400 or higher are typically considered at high risk, and more aggressive management strategies, such as medication, lifestyle changes, or further diagnostic tests, may be recommended.

While CT scans have been widely employed as a biomarker for CVD diagnosis, they do come with several drawbacks. These disadvantages include cost and accessibility issues. The expense of a CT scan can vary significantly, ranging from $300 to $6,750, and additional expenses can accrue when further investigations are necessary (Sekikawa et al. 2012). Moreover, the high cost of CT machines themselves has limited their availability in local hospitals, contributing to the complexity of the diagnosis process. Consequently, the combination of high costs and limited accessibility has posed challenges for achieving early CVD diagnosis. Hence, there is a pressing need to explore alternative biomarkers that can facilitate accurate early detection of CVD.

Retinal Image

A retinal image is a digital representation of the human retina, which is the light-sensitive tissue at the back of the eye. Retinal images are typically captured using specialized medical equipment, such as fundus cameras or ophthalmic imaging devices. These images provide a detailed view of the retina's structure, including the blood vessels, optic nerve head, macula, and other anatomical features. They can be useful for diagnosing various eye conditions and systemic diseases that may manifest in the retina, such as diabetic retinopathy, glaucoma, age-related macular degeneration, and hypertensive retinopathy.

Figure 2. Retinal image medical equipment

Retinal images provide a non-invasive way to observe the intricate structures inside the eye, including blood vessels and the optic nerve, without the need for surgical procedures. Retinal imaging offers a safer and less invasive alternative, allowing ophthalmologists and medical professionals to visualize these critical components of the eye in a way that is comfortable for the patient and does not carry the risks associated with surgery or invasive techniques.

Due to these advantages, several research studies have been conducted with the aim of developing a biomarker for the diagnosis of CVD. Wong et al. proposed a CVD diagnosis system based on convolutional neural networks. Their method establishes a correlation between retinal images and CVD diagnosis, demonstrating the feasibility of implementing such a system (Wong et al. 2022) .

Proposed Approach

Inspired by their work, we propose a novel cardiovascular disease diagnosis framework from retinal image. The proposed framework takes the retinal image as input and outputs categorization based on the severity of CVDs. Additionally, we introduce an age-guided pipeline to enhance the accuracy and stability of the inference process. The structure of this research paper is as follows: Chapter 2 provides a comprehensive explanation of the proposed framework, Chapter 3 assesses the effectiveness of the framework through extensive experiments, and Chapter 4 offers a summary of the research findings.

Proposed Cardiovascular Disease Detection

The outlined framework comprises three modules: the retinal feature extractor, age prediction network, and cardiovascular disease (CVD) diagnosis network. The following subheadings will provide a comprehensive explanation of each module.

Cardiovascular Disease Detection Network

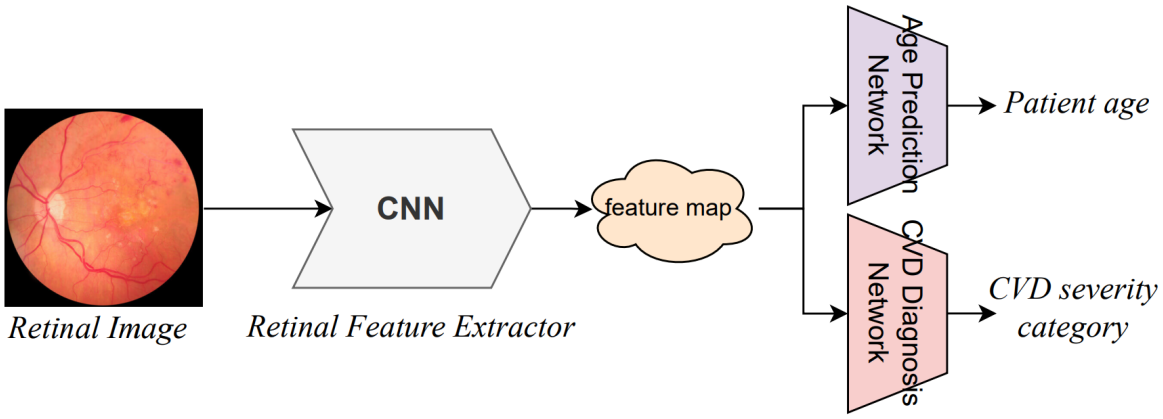

Figure 3. Architecture of the proposed cardiovascular disease detection network

Figure 3 provides an overview of the proposed network architecture. Initially, the input retinal image, denoted as $I \in R^{HW}$, is processed by the retinal feature extractor to generate a feature map referred to as Z. This retinal feature extractor, denoted as RFE, can be defined as follows: RFE: $I \rightarrow Z$, where $Z \in R^{D}$ represents the resulting feature map. This feature map is then fed to two tasks within the network: age prediction and CVD diagnosis. In the case of age prediction, denoted as APN, the relationship is defined as APN: $Z \rightarrow \widehat{age}$, where $\widehat{age}$ corresponds to the predicted age of the patient. On the other hand, for CVD diagnosis, denoted as CDN, the relationship is defined as CDN: $Z \rightarrow P$, where P stands for the patient's severity of CVD. To summarize, the initial retinal image undergoes processing through the retinal feature extractor, yielding a feature map that is later utilized by both the age prediction network and the CVD diagnosis network to predict the patient's age and assess their CVD severity, respectively.

Loss Function

To train the proposed cardiovascular disease detection network, we employed two types of loss functions: the L-1 loss function and the cross-entropy loss function.

Equation 1: L-1 loss function

$$L_{1} = \sum_{i=1}^{n} \left| age_i - \widehat{age}_i \right|$$

Here, age_i and $\widehat{age}_i$ denote the patient's actual age and predicted age respectively. If the predicted age is equivalent to the patient's actual age, the absolute difference between the age_i and $\widehat{age}_i$ will be 0, representing no loss. On the contrary, if there is a difference, the higher the difference, the greater the loss.

Equation 2: Cross-entropy loss function

$$L_{CE} = -log_e(P)$$

Where, P represents the patient's severity of CVD. The cross-entropy loss function helps represent the loss for $0 \leq P \leq 1$.

Equation 3: Total loss function

$$L_{total} = L_{CE} + \alpha L_{1}$$

Here, **α** acts as the weight that accentuates the loss function found by the cross-entropy loss function, as the severity of CVD is the more essential prediction of the proposed model. Hence, the value of **α** is less than 1, which diminishes the value of the L-1 loss function.

Implementation Details

In the implementation, we employed Resnet-101 (He et al. 2016) as the foundational architecture for the retinal feature extractor. The ImageNeet pretrained weight is utilized for faster training process. For the development of both the age prediction network and the CVD diagnosis network, a two-layered neural network was employed. The training process involved utilizing the stochastic gradient descent optimizer (Ruder et al. 2016) over 110 epochs, with a learning rate of 0.0001 and a batch size of 512. The NVIDIA Titan RTX (Sanzharov et al. 2019) device was selected as the hardware for training the network.

Experimental Result

Dataset

The dataset used in this research contains retinal images of 37,295 patients. The average age of the patient is 59.3years old. Among the 37,295 patients there are 24,827 male patients which is 66.57% of the total dataset. There are 12,468 female patients, which is 33.43% of the total patients. Each retinal image sample is assigned to one of four categories representing specific levels of cardiovascular disease (CVD) severity. Approximately 50.49% (18,829 samples) are classified into level-1, where the coronary artery calcium score (CACS) is observed to be zero. In the second category, level-2, approximately 26.74% (9,972 samples) have a CACS greater than zero but equal to or less than 100. The third and fourth categories, level-3 and level-4, comprise approximately 13.05% (4,867 samples) and 9.72% (3,627 samples) of the dataset, respectively. Level-3 includes samples with a CACS greater than 100 and equal to or less than 400, while level-4 encompasses samples with a CACS greater than 400. The medical significance of these CACS values is thoroughly explained in Chapter 1.2.

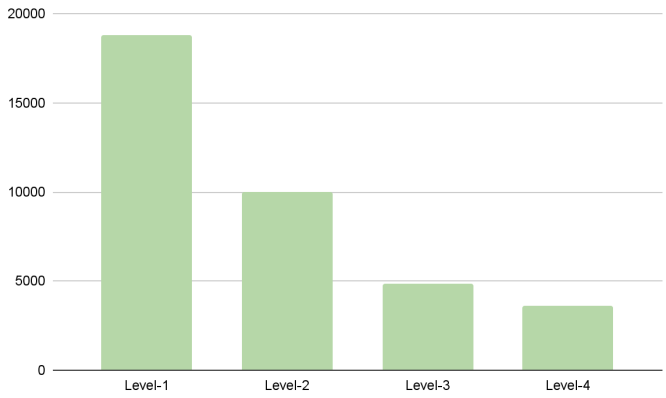

Figure 4. Sample distribution of the retinal image dataset used in this research

Evaluation Metrics

To train the proposed model, we conducted the k-fold cross validation. K-fold cross-validation is a machine learning evaluation technique used to assess the performance and generalization capabilities of predictive models. In this approach, the dataset is initially partitioned into K folds, where K represents the number of subsets.

During each iteration of K-fold cross-validation, one of the K folds is designated as the testing set, while the remaining K-1 folds serve as the training set. The model is trained on the training set, and its performance is evaluated by predicting accuracy and calculating error on the testing set. This process is then repeated K times, with each fold taking turns as the validation set. In this paper, we conducted a 5-fold cross validation.

For evaluation metrics, we used general classification metrics commonly used in machine learning for image classification tasks, including accuracy, recall, precision, and the F1-score. To gain a comprehensive understanding of the model's performance, we also utilized the confusion matrix which offers a summary of the model's predictions. An illustrative example of a confusion matrix is provided in Figure 5.

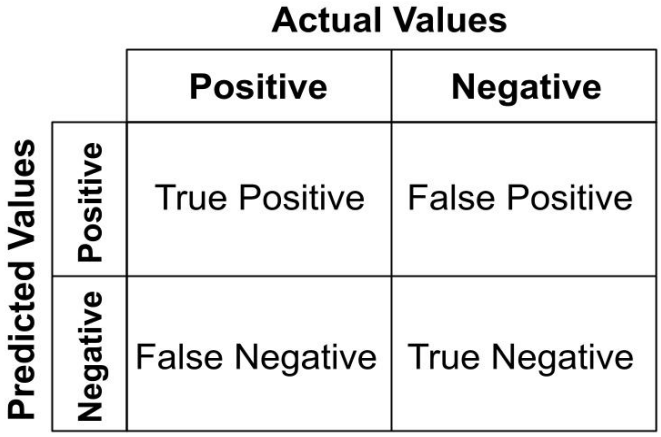

Figure 5. Illustration of confusion matrix

As shown in Figure 5, there are 4 cases that the proposed method can affect: true positive, false positive, false negative, and true negative. True positive is the case when the model predicts the positive class and the actual class is indeed positive. False positive is the case when the model predicts the positive class but the

actual class is negative. False negative is the case when the model predicts the negative class but the actual class is positive. True negative is the case when the model predicts the negative class and the actual class is indeed negative.

With these four cases, the following evaluation metrics can be calculated: the accuracy, recall, precision, and f1-score. In the following equations, true positive is denoted as *TP*, false positive as *FP*, false negative as *FN*, and true negative as *TN*.

Equation 4: Accuracy

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

Equation 5: Recall

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

Equation 6: Precision

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

Equation 7: F1-score

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

Here, the accuracy is calculated as the ratio of the number of correct predictions to the total number of predictions. The recall is calculated as the ratio of true positives to the total number of actual positive instances. The precision is calculated as the ratio of true positives to the total number of positive predictions. Finally, the f1-score is calculated as the mean of precision and recall. After the 5-fold cross validation, the averages of these four evaluation metrics will be calculated, as well the standard deviation for each, in order to evaluate the stability of the proposed method.

Evaluation Comparison

For the performance comparison, we have selected five convolutional neural network-based methods known for their competitive performance in general image classification tasks. They are VGG19 (Simonyan et al. 2014), MobileNetV2 (Sandler et al. 2018), HRNet-32 (Wang et al. 2020), DenseNet-121 (Huang et al. 2017), and ResNet-101 (He et al. 2016). The results of this comparison are presented in Table 1 and visually summarized in Figure 4.

Table 1. Comparison with state-of-the-art methods in performance

	Accuracy	Recall	Precision	F1-Score
VGG19 (Simonyan et al. 2014)	0.6857 (±0.0009)	0.6725 (±0.0009)	0.6504 (±0.0005)	0.6593 (±0.0014)
MobileNetV2 (Sandler et al. 2018)	0.7046 (±0.0011)	0.6805 (±0.0009)	0.6681 (±0.0015)	0.6842 (±0.0013)
HRNet-32 (Wang et al. 2020)	0.7427 (±0.0004)	0.7164 (±0.0005)	0.7042 (±0.0008)	0.7182 (±0.0005)

DenseNet-121 (Huang et al. 2017)	0.7758 (±0.0013)	0.7859 (±0.0008)	0.7370 (±0.0005)	0.7417 (±0.0011)
Resnet-101 (He et al. 2016)	0.7994 (±0.0007)	0.7796 (±0.0008)	0.7456 (±0.0009)	0.7594 (±0.0010)
Proposed Method	0.8188 (±0.0011)	0.8001 (±0.0006)	0.7525 (±0.0008)	0.7716 (±0.0007)

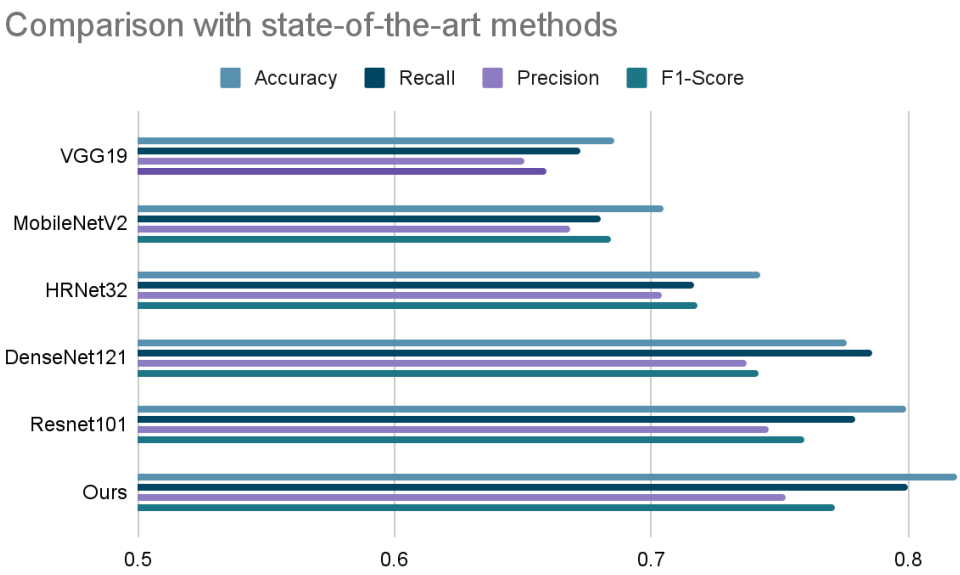

Figure 6. Comparison with state-of-the-art methods in performance

As Table 1 and Figure 6 show, shallow networks like VGG19 and MobileNetV2 had lower performance scores, almost all below 70%. Relatively deeper networks like HRNet32, DenseNet121, and Resnet-101 performed significantly better, exceeding 70% and even nearing 80% in accuracy. The proposed method surpassed all of the other methods in all of the four factors, performing over 13% better than VGG19 in accuracy. As a result, the proposed method can be seen performing more effectively than other state-of-the-art methods. We attribute this superior performance to the proposed approach of jointly training the age prediction network and the CVD diagnosis network. The effectiveness of the proposed age-guided approach is examined in Chapter 3.4.

Ablation Study

To demonstrate the effectiveness of the proposed age-guided approach, we conducted an ablation study. To ensure a fair comparison, we trained a CVD diagnosis network without the age prediction module, utilizing the exact same training protocol outlined in Chapter 2.3. This model is referred to as the "ablation model." Subsequently, we compared this ablation model with the full model, which was trained using the proposed age-guided approach.

A summary of the ablation study results is presented in Table 2, and Figure 5 provides a visual representation of these findings.

Table 2. Ablation study result

	Accuracy	**Recall**	**Precision**	**F1-Score**

ablation model (without age prediction module)	0.7994 (±0.0007)	0.7796 (±0.0008)	0.7456 (±0.0009)	0.7594 (±0.0010)
full model	0.8188 (±0.0011)	0.8001 (±0.0006)	0.7525 (±0.0008)	0.7716 (±0.0007)

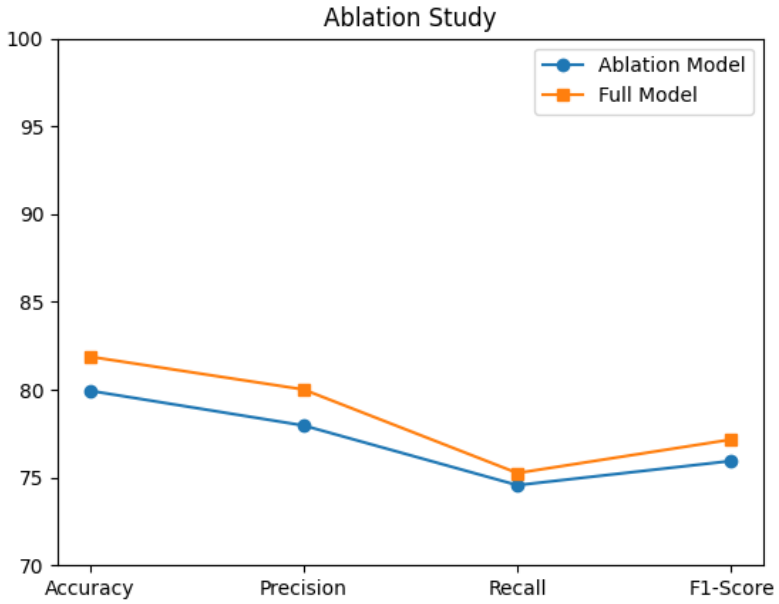

Figure 5. Ablation study (orange line: full model, and blue line: ablation model)

Table 2 and Figure 5 clearly demonstrate that the proposed age-guided approach contributed to the performance improvement. Considering the well-established and widely recognized correlation between age and CVD, this age-guiding system has significantly enhanced the network's effectiveness.

Feature Engineering

As an additional experiment, we tested the proposed method's performance in various color inputs, such as RGB, grayscale, and YCbCr.

Table 3. Feature engineering result

	Accuracy	Recall	Precision	F1-Score
RGB	0.8188 (±0.0011)	0.8001 (±0.0006)	0.7525 (±0.0008)	0.7716 (±0.0007)
Grayscale	0.8172 (±0.0009)	0.8001 (±0.0010)	0.7499 (±0.0007)	0.7696 (±0.0009)
YCbCr	0.8207 (±0.0007)	0.8001 (±0.0009)	0.7540 (±0.0005)	0.7742 (±0.0011)

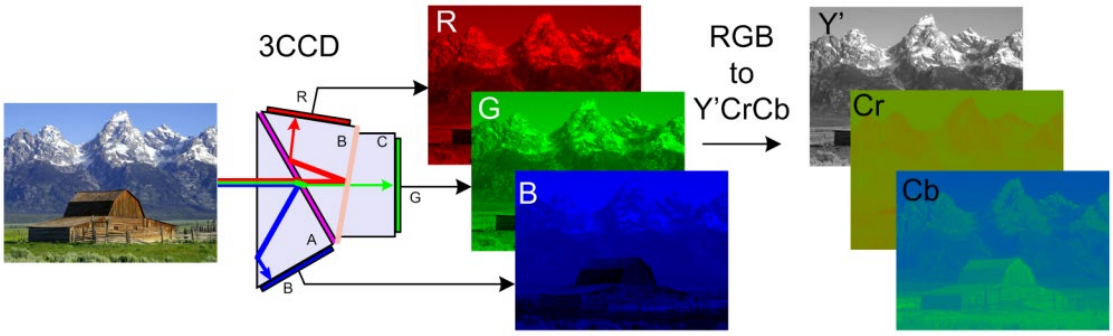

Figure 6. Example of rgb and YCbCr conversion

The original input retinal image is stored in the RGB color format. We then performed conversions to both grayscale and YCbCr color spaces. As shown in Table 3, the conversion into the YCbCr image format contributed to an improvement in accuracy. In the YCbCr color space, the luminance (Y) and chrominance (Cb and Cr) information are separated, offering particular utility for analyzing the geometric patterns of nerves within the retinal image.

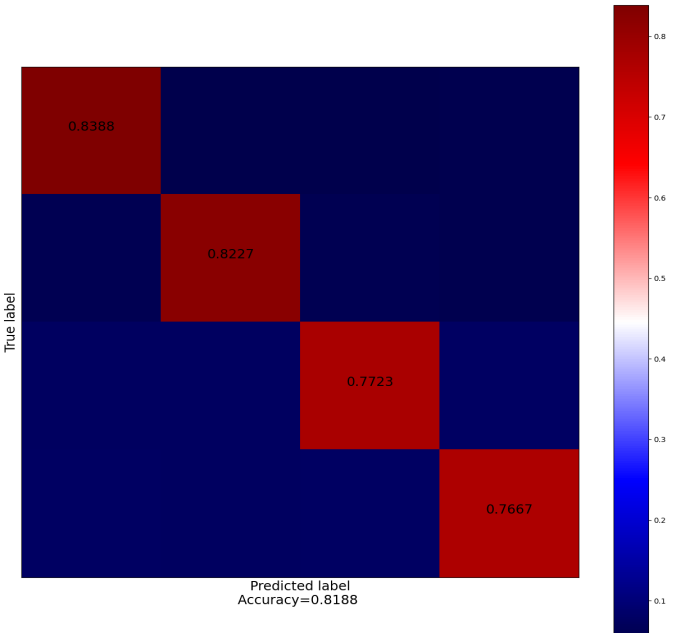

Figure 7. Confusion matrix of the proposed method

Finally, in Figure 7, the confusion matrix for the proposed methods is presented. The strong presence of diagonal elements in the matrix indicates that the proposed method consistently achieves high accuracy across all CVD severity categories.

Conclusion

In this study, we proposed a novel machine learning-based approach for the detection of cardiovascular diseases using retinal images. We considered the diagnosis of cardiovascular diseases as a classification task, aiming to assess the severity level of the condition based on the input retinal image. In addition, we proposed jointly

training the age prediction network and the CVD diagnosis network. The effectiveness of this approach is clearly demonstrated through comprehensive experiments. Furthermore, we explored the utility of the YCbCr color space in analyzing retinal images. By separating luminance and chrominance information, we gained valuable insights into the geometric patterns of nerves within the retinal images. In summary, this research study has demonstrated the viability of utilizing retinal images as biomarkers for diagnosing CVD. For future research, we will investigate the correlation between CVD prognosis and retinal images.

Acknowledgments

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

References

Gaziano, T., Reddy, K. S., Paccaud, F., Horton, S., & Chaturvedi, V. (2006). Cardiovascular disease. Disease Control Priorities in Developing Countries. 2nd edition.

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

Jager, K. J., Lindholm, B., Goldsmith, D., Fliser, D., Wiecek, A., Suleymanlar, G., ... & Dekker, F. W. (2011). Cardiovascular and non-cardiovascular mortality in dialysis patients: where is the link?. Kidney international supplements, 1(1), 21-23.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747. https://doi.org/10.48550/arXiv.1609.04747

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

Sanzharov, V. V., Gorbonosov, A. I., Frolov, V. A., & Voloboy, A. G. (2019, December). Examination of the Nvidia RTX. In CEUR Workshop Proceedings (Vol. 2485, pp. 7-12). http://dx.doi.org/10.30987/graphicon-2019-2-7-12

Sekikawa, A., Curb, J. D., Edmundowicz, D., Okamura, T., Choo, J., Fujiyoshi, A., ... & Ueshima, H. (2012). Coronary artery calcification by computed tomography in epidemiologic research and cardiovascular disease prevention. Journal of epidemiology, 22(3), 188-198.

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

Wong, D. Y., Lam, M. C., Ran, A., & Cheung, C. Y. (2022). Artificial intelligence in retinal imaging for cardiovascular disease prediction: current trends and future directions. Current Opinion in Ophthalmology, 33(5), 440-446. https://doi.org/10.1097/ICU.0000000000000886