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

CardioXNet: Representation Learning for Accurate Cardiovascular Disease Diagnosis from X-Ray Images

Jiwon Hwang[1] and Jesse Klingebiel[#]

[1]Kent School, USA
[#]Advisor

ABSTRACT

Cardiovascular Disease (CVD) is a prevalent, incurable condition affecting the heart and blood vessels. Due to its significant impact on mortality in the United States, there is a pressing need for enhanced risk stratification methods. Coronary Artery Calcium Score (CACS), reliant on CT scans, is most commonly employed as a risk stratification method but suffers from limitations, including accessibility and early detection challenges. To address the problem, this research study proposes a representation learning-based CVD diagnosis framework utilizing X-ray images, offering expedited and earlier detection, as well as improved accessibility. This system comprises two distinct stages: representation learning to extract CVD-related features and transfer learning to train a CVD classifier. Representation learning enhances the quality of extracted features, thereby leading to more precise results in the subsequent CVD diagnosis network. Comprehensive experiments and training validate the efficacy of the proposed method, demonstrating its superiority over the existing methods. These promising results suggest the potential utility of X-rays as a valuable biomarker for diagnosing CVD disease.

Introduction

Problem Definition

Cardiovascular disease (CVD) is an incurable cardiac disorder typically associated with atherosclerosis and an elevated risk of blood clots. It has emerged as the leading cause of mortality among both men and women in the United States, primarily due to sedentary lifestyles and unhealthy dietary habits that lead to high blood pressure, high cholesterol, and diabetes, which in turn increases the risk of CVD. Given the escalating prevalence of CVD, implementation of effective CVD diagnosis methods has become crucial. Early detection of CVD holds a particular significance since it helps doctors select appropriate treatment at an early stage and helps to prevent further deterioration of the victim's health. While the Coronary Artery Calcium Score (CACS), a CT scan-based risk stratification assessing calcium in the atherosclerotic plaques in the coronary arteries, has conventionally served as a widely adopted risk stratification system, its reliance on costly CT scans limits public accessibility. Conversely, using machine learning can increase the accuracy of detection, contribute to early diagnosis, and improve accessibility by using radiology.

Previous Method

There are previous cases where scientists have facilitated AI for CVD risk stratification systems. Wong et al. proposed a risk stratification system using a deep-learning algorithm on retinal photography that predicted systemic biomarkers for CVD (Wong et al. 2022). Their research's primary aim was to make associations between systemic disease with unobservable retinal features. The research developed 47 deep-learning algorithms to estimate 47 systemic biomarkers.

Rim et al. proposed a risk stratification system using a deep-learning algorithm on retinal photographs that predicted systemic biomarkers for CVD (Rim et al. 2021). Their research aims to make associations between systemic disease with unobservable retinal features. They provided a more accessible CVD diagnosis method by using retinal photographs instead of CT scans. Rim et al. also proposed a simplified cardiovascular disease risk stratification system, which is based on deep-learning-predicted CAC from retinal photographs (Rim et al. 2020). This utilizes the risk stratification system they developed previously and develops it further to a model that predicts the probability of the presence of CAC based on retinal photographs. Another group of scientists proposed a CVD diagnosis method using retinal photography. In 2019, De Vos et al. proposed a method that can perform CACS directly in multiple types of CT by using an unsupervised deep learning atlas-registration method. Their method detected CACS by registering input CT to CT atlas image (De Vos et al. 2019).

Proposed Method

Improving on the previous methods, I propose a new method CardioXNet, an AI-powered biomarker for CVD diagnosis using X-ray images, to rectify the inaccessibility and the accuracy of previous CVD diagnosis methods. The proposed method will input X-ray images and output the predicted severity of CVD after CACS analysis. To develop this model, I propose the feature-swapping mechanism that disentangles CVD-related features from the input X-ray image. Previously, the methods proposed incorporated supervised learning. However, when non-diverse data was provided, the method caused bias, failing to detect the target features. In other words, features extracted from the input images were entangled, which caused the trained network to be biased on the dataset. Compared to the previous method, the proposed feature-swapping mechanism leverages the limited data. That is, the feature-swapping mechanism disentangles the entangled feature map and identifies the relevant feature.

Related Work

Cardiovascular Disease

Cardiovascular Disease(CVD) is a general term for fatal heart disease, often associated with atherosclerosis. Atherosclerosis is a term used to describe a condition that develops when plaque builds up in the walls of the arteries. This plaque narrows the arteries, making it harder for blood flow, and imposes a high risk of heart attack on the victim. CVD includes conditions such as coronary artery disease, heart failure, stroke, and peripheral artery disease, among others. These conditions often arise due to a combination of genetic, environmental, and lifestyle factors, making them complex and multifactorial.

Traditionally, diagnosis of CVD is possible after a certain amount of time when the presence of the disease is obvious. Most typically, Coronary Artery Calcium Scoring(CACS), which analyzes the calcium deposits in the plaque in the arteries using special Computed Tomography (CT) scans, is used to diagnose CVD. Once the scanner takes multiple pictures of the heart, it is combined to reveal the calcium deposits that are

represented as white clots (Rim et al. 2021). Seen in Figure 1. are calcium plaques represented as white packs, indicated by the arrows. Once identified, the predefined software logic quantifies a score based on the calcification. Although accurate, this method is inaccessible to the public and fails to early diagnose CVD. As a result, there has been a growing interest in leveraging the power of artificial intelligence to enhance the detection, early diagnosis, and prediction of CVD.

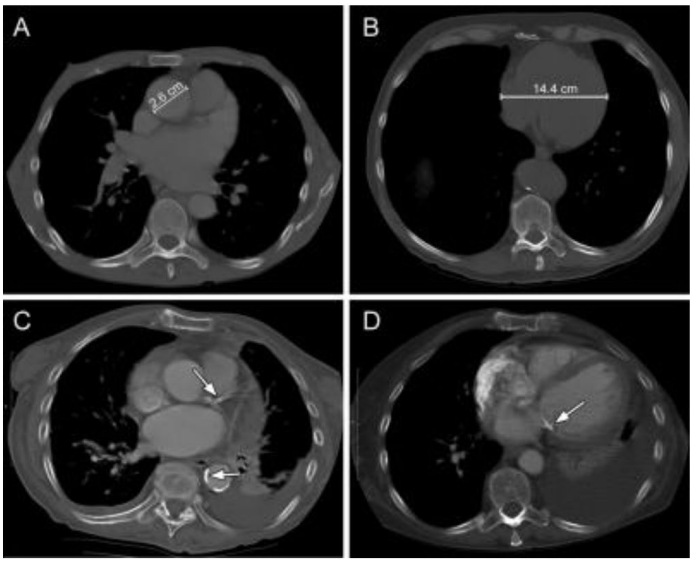


Figure 1. CT Scans of Calcium Plaques

In this research, I propose an AI-powered biomarker that diagnoses CVD for early diagnosis. Further details on the system will be explained in Chapter 3.

X-Ray Image

X-ray images are radiographs that are capable of creating an image of dense tissues and structures inside the human body using X-rays. X-ray imaging can be accessed commonly in many hospitals and therefore has low cost unlike CT scanned images, which are inaccessible and expensive. Recent advancements in artificial intelligence, particularly in the field of deep learning, have shown promising results in improving the accuracy and efficiency of cardiovascular disease detection from medical images. Deep learning models can be trained on extensive medical image samples, incorporating thousands of X-ray images annotated with corresponding clinical information which is CACS calculated by the predefined logics. Matsumoto et al. (Matsumoto et al. 2020) proposed a deep learning algorithm that diagnosed heart failures using X-ray images. They successfully relabeled 260 normal and 378 heart failure images and obtained an 82% accuracy rate. Their method tends to suffer from the overfitting problem and yields unsatisfactory results primarily due to the limited size of their dataset.

In this research paper, I aim to leverage and expand upon machine learning techniques to improve the accuracy of CVD detection. I plan to utilize a large-scale dataset, including external validation. The detailed description of the proposed approach will be explained in Chapter 3.

Image Classification

Image classification refers to a computer vision task involving inputting images and categorizing them into a set of categories as shown in Figure 3. It is often used to develop an automated program that analyzes the visual images that have been inputted and then assigns them into corresponding categories. This process is typically implemented by using Linear Classifiers(LN), Neural Networks (NN), and Convolutional Neural Networks (CNN).

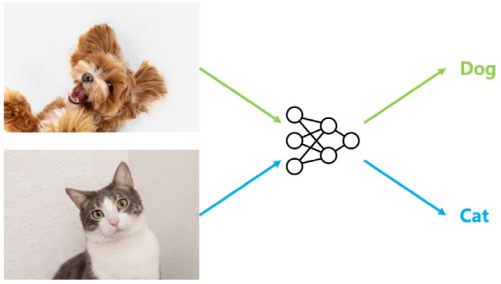
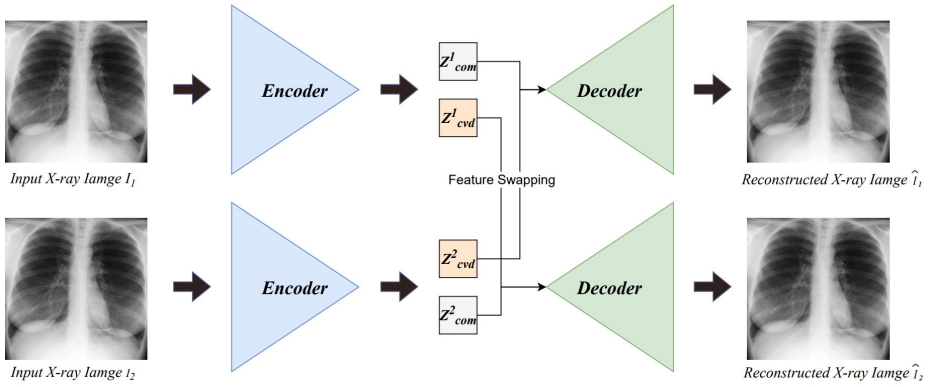


Figure 2. An Example of Image Classification

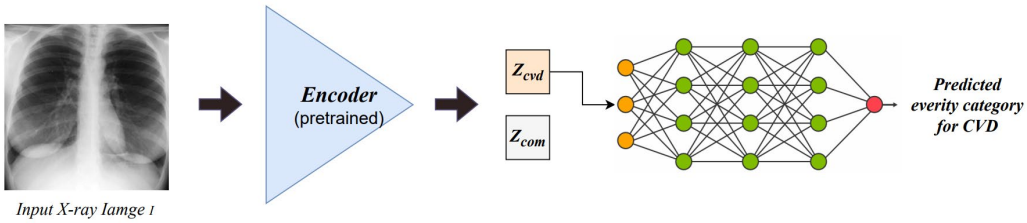
Among them all, CNN shows comparable performance in many computer vision tasks. Unlike other systems, CNNs excel in capturing intricate patterns and hierarchical representations from images, enabling them to learn complex features effectively. Prominently applied CNNs include AlexNet (Krizhevsky et al. 2012), VGGNet (Simonyan et al. 2014), and ResNet (He et al. 2016).

These have numerous applications in the medical field due to their efficiency and high accuracy. Doctors would input the X-ray, CT, and MRI images of their patients to a certain image classification system and it will output whether the disease being identified is present in the patient or not. For instance, retinal photographs are inputted into the classification system to diagnose diabetes, and MRI scans are inputted to detect the presence of Alzheimer's disease. In this research, I approach the CVD diagnosis system as an image classification problem, where the goal is to categorize x-ray images into specific ranges of CACS.

Proposed Method



(a) X-ray Representation Learning



(b) Transfer Learning (cardiovascular disease diagnosis module)

Figure 3. The Architecture of the Proposed Cardiovascular Disease Diagnosis System

In this chapter, a comprehensive explanation of the proposed framework will be presented, including the design of the convolutional neural network and the medical rationale behind it. The proposed framework consists of two stages: the first stage focuses on representation learning to disentangle CVD-related features, while the second stage involves transfer learning to train a CVD classifier. Representation learning serves the purpose of disentangling the CVD-related features from the input X-ray image and transfer learning is used to output specific categories of the inputted image.

Representation Learning for CVD-related Features

This process serves the purpose of disentangling CVD-related features from a feature map. In representation learning, two CNNs are used to process images showing a low CACS radiology image and a high CACS radiology image to output a high CACS radiology image and a low CACS radiology image correspondingly.

Each image is inputted to the encoder, which will produce a feature map. The CVD-related feature is disentangled and inputted into each other's decoder using a feature-swapping mechanism. For instance, if the images inputted were 100 and 300, the CVD-related feature of 100 is processed through the 300 decoder. Once processed, the CNN with low CACS input outputs a high CACS radiology image, and the CNN with high CACS input outputs a low CACS radiology image.

To train the proposed autoencoder architecture, I employ the L1 loss function, denoted as Equation 1.

Equation 1: L1 Loss Function:

$$L_1 = \frac{1}{XY} \sum_{x}^{X} \sum_{y}^{Y} \left| I(x,y) - \hat{I}(x,y) \right|$$

Here, X and Y denote the width and height of the input image while I(x,y) represents the pixel intensity at the coordinates (x,y). The loss function measures the average absolute difference between a reconstructed X-ray image and its original image. The best-case scenario for the L1 loss function is when the reconstructed image matches its original image perfectly, resulting in an L1 loss of zero.

Transfer Learning (Fine-Tuning for CVD Classification)

The main goal of this stage is to output a category for the input radiology image. In transfer learning, a pre-trained encoder of an autoencoder is used for fine-tuning CVD classification.

X-ray image of a patient is inputted into the pre-trained encoder, which produces a feature map. CVD-related features are disentangled and processed through the neural network, outputting a category assigned by CACS analysis.

To train the proposed CVD diagnosis network, I employed the cross-entropy loss function, denoted as Equation 2.

Equation 2: Cross-Entropy Loss Function:

$$L_i = -\ln P\left(Y = y_i \middle| X = x_i \right)$$

Here, P denotes the probability distribution corresponding to the ground truth of the input image. The cross-entropy loss function evaluates the dissimilarity between the predicted probability assigned to each class by the model and the true probabilities associated with those classes. The best-case scenario occurs when the predicted probability distribution aligns perfectly with the true probability distribution of class labels, resulting in a loss value of zero.

Experimental Results

Dataset

In this chapter, I will provide detailed information about the dataset I used to train the proposed method. The dataset consists of X-ray images of 43,962 patients with an average age of 56.8 years. 37.01% of the patients are female, 62.99% male as shown in Table 1.

Table 1. Category distribution of the dataset used in this research.

Patients	43,962
Average Age	56.8
Female	16,241 (37.01%)
Male	27,651 (62.99%)
CACS 0 (Absent)	20,642 (47.03%)
CACS >0 (Discrete)	11,484 (26.16%)
CACS >100 (Moderate)	7,149 (16.29%)
CACS >400 (Accentuated)	4,617 (10.52%)

The X-ray samples are categorized into 4 labels based on their Coronary Artery Calcium Score (CACS). The 4 labels are Absent, Discrete, Moderate, and Accentuated. Patients with a CACS of 0 are categorized as Absent, indicating an extremely low possibility of CVD presence. Patients with CACS between 0 and 100 are categorized as Discrete, showing a low CVD presence. Patients with CACS between 101 and 400 are categorized as Moderate, which acknowledges the potential risk of further development of CVD and considers its potential for re-categorization into a higher level. Patients with CACS of more than 400 are categorized as Accentuated, indicating a high presence of CVD. The proposed method outputs one of these 4 categories for the inputted X-ray image.

Experimental Protocol

For the experiment, 5-fold cross-validation is conducted to assess the performance of the proposed method. This method divides the dataset into 5 groups, then tests each group and trains the rest. In each of the experiments, the accuracy of the testing group is presented, and ultimately, these accuracies are averaged to present

the accuracy of the model. I used standard deviation to prove that the accuracy was stable, meaning that accuracies from each experiment were consistent.

To assess the performance of the proposed method, I used the 4 most popular evaluation metrics: accuracy, precision, recall, and f1-score as shown in Equation 3-6.

Equation 3: Accuracy:

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

Where, TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively.

True positive occurs when the classifier correctly predicts a positive output for an instance, and the actual outcome is indeed positive. True negative occurs when the classifier correctly predicts a negative output, and it is negative.

False negative describes a situation where a classifier incorrectly predicts a negative outcome for an instance when the actual outcome is positive. False positive describes a situation where a classifier incorrectly predicts a positive outcome for an instance when the actual outcome is negative.

Equation 4: Recall:

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

Recall is used to calculate the proportion of actual positive instances among all the positive samples. It evaluates the performance of the classifier in finding positive outcomes.

Equation 5: Precision:

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

Precision evaluates the classifier's ability to predict the positive outcomes correctly. It represents a ratio of the correct positive values to all the instances that the classifier predicted as positive.

Equation 6: F1-Score:

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

F1-score measures the harmonic mean between Precision and Recall. It provides a balanced measure of a classifier's performance by considering both false positives and false negatives.

Comparison

Table 2. Performance comparison.

Model Architecture	Accuracy	Precision	Recall	F1-Score
VGG19 (Simonyan et al. 2014)	0.6694 (± 0.0005)	0.6192 (± 0.0007)	0.6371 (± 0.0009)	0.6317 (± 0.0010)

MobileNetV2 (Sandler et al. 2018)	0.6855 (± 0.0007)	0.6419 (± 0.0011)	0.6640 (± 0.007)	0.6442 (± 0.0009)
Xception (Fran et al. 2017)	0.7245 (± 0.0011)	0.6871 (± 0.0012)	0.6894 (± 0.0008)	0.6941 (± 0.0014)
HRNet-w32 (Wang et al. 2020)	0.7504 (± 0.0013)	0.7310 (± 0.0009)	0.7275 (± 0.0012)	0.7204 (± 0.0010)
DenseNet-121 (Huang et al. 2017)	0.7794 (± 0.0009)	0.7414 (± 0.0008)	0.7705 (± 0.0011)	0.7545 (± 0.0009)
Resnet-101 (He et al. 2016)	0.7884 (± 0.0010)	0.7492 (± 0.0008)	0.7714 (± 0.0006)	0.7628 (± 0.0004)
Proposed Method (Resnet-101 based)	0.8372 (± 0.0007)	0.7873 (± 0.0010)	0.8000 (± 0.0008)	0.7924 (± 0.0009)

Table 2 shows a comparison of the performance of the proposed methods with that of state-of-the-art methods. This is conducted to prove the improved performance of the proposed method compared to the previous methods. I trained each of the model architectures on the same dataset using 5-fold cross-validation and used the evaluation metrics to assess their performance. The results showed a clustering effect: networks with a similar number of convolutional layers showed similar performance. Shallow networks, such as VGG19 and MobileNetV2, had low accuracies of approximately 67-68%. Networks with more layers, Xception and HRNet-w32, showed comparable performances with a 5% difference at most. Deeper networks, such as DenseNet-121 and Resnet-101, had higher accuracies than the others but still showed similar performances with almost the same accuracy rate. The proposed method showed a significantly higher accuracy and better performance overall. I attribute this superiority to the proposed feature-swapping mechanism-based approach.

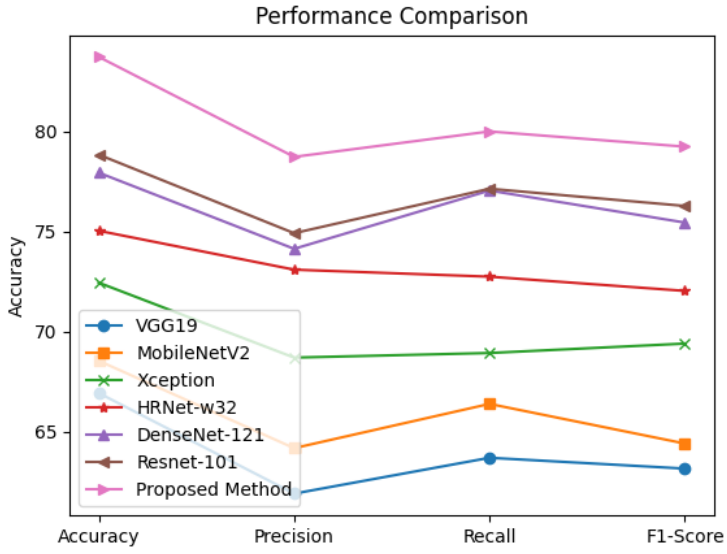


Figure 4. Performance comparison (line graph)

Figure 5 visualizes the performance of each model in a line graph form. It can help our understanding of the performance by representing each model in 7 distinct lines. From the graph, it can be observed that the results are consistent.

Architecture Replacement

Table 3. Architecture replacement

Model Architecture	Accuracy	
	baseline	**proposed method applied**
VGG19	0.6694	0.7009 (+3.15%)
MobileNetV2	0.6855	0.7173 (+3.18%)
Xception	0.7245	0.7740 (+4.95%)
HRNet-w32	0.7504	0.7953 (+4.49%)
DenseNet-121	0.7794	0.8259 (+4.65%)
Resnet-101	0.7884	0.8372 (+4.88%)

In this experiment, I aim to prove that the proposed method shows good performance independent of the model architecture. The proposed method was trained using the feature swapping mechanism, and it was applied to each model architecture using the transfer learning method. Comparing when the model architectures were trained on a baseline model, and when the proposed method was applied, every model improved in its performance by 3-4% consistently. This experiment proved that the proposed method could be added independently regardless of the model architectures.

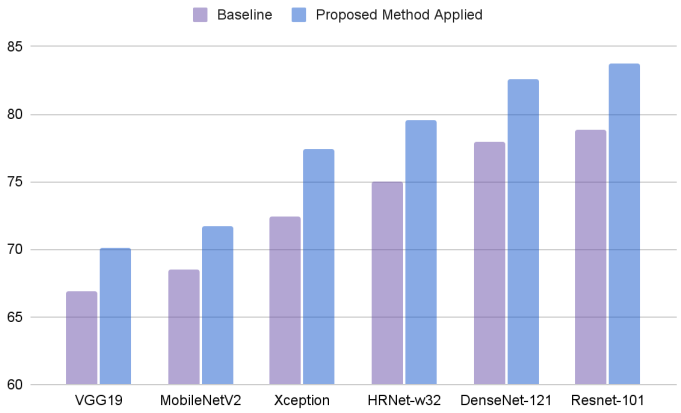


Figure 5. Accuracy comparison on architecture replacement

Architecture Replacement

Table 3. Architecture replacement

Augmentation Method	Accuracy	Precision	Recall	F1-Score
baseline	0.8372	0.7873	0.8000	0.7924
Sharpness	**0.8541**	**0.7925**	**0.8176**	**0.8072**
Gaussian Noise	0.8045	0.7702	0.7852	0.7675
Histogram Equalization	0.8245	0.7608	0.7842	0.7904
CLAHE	0.8186	0.7704	0.7682	0.7753

I modified the images' sharpness, gaussian noise, histogram equalization, and CLAHE then compared it to the baseline performance. Sharpness refers to the degree to which an image appears as well-defined; Gaussian noise involves adding random noise that abides by a Gaussian distribution; histogram equalization improves the contrast and visibility of details in an image by redistributing the intensity values of the image's pixels; and Contrast-Limited Adaptive Histogram Equalization(CLAHE) improves the visibility of a region with varying illumination levels. However, the accuracy only improved when the sharpness of the image was modified.

By applying sharpness, the details of subtle cardiovascular areas on an X-ray image were emphasized, and this improved the overall performance of our trained network.

Architecture Replacement

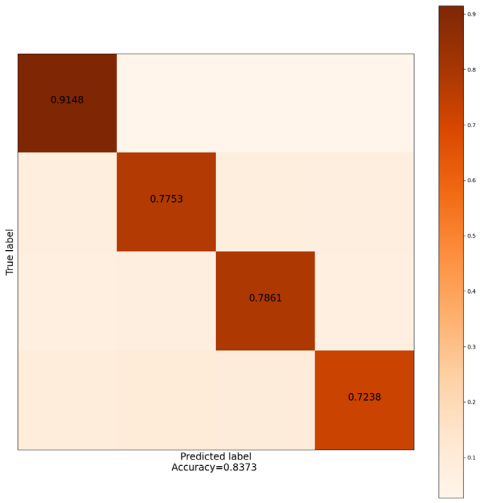


Figure 6. Confusion matrix of the proposed method

To evaluate the performance of the model, I used the confusion matrix. A confusion matrix is an evaluation matrix that visually summarizes correct and incorrect predictions. The representation of the results is easy to understand, and it is widely used to evaluate image classification systems. Here, it is noticeable that the ratio of diagonal components of the matrix is higher than the rest, implying a high performance for all 4 CACS categories. This robust result proves the overall high performance of my model.

Conclusion

In this paper, I proposed CardioXNet, an AI-powered biomarker for CVD diagnosis using X-rays, bolstered by a feature-swapping mechanism enhancing the performance of image classification. Developed with two Autoencoders and transfer learning, CardioXNet consistently outperformed existing state-of-the-art models, improving their accuracy by an average of 4.22%. Beyond quantitative results, I highlight the method's applicability to enhance accuracy in various image classification systems. The model is also suitable for use at local hospitals, enabling patient access to CVD diagnosis without expert intervention. Future works aim to refine and implement a practical system that can be applied in local hospitals.

Acknowledgments

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

References

De Vos, B. D., Wolterink, J. M., Leiner, T., de Jong, P. A., Lessmann, N., & Isgum, I. (2019). Direct Automatic Coronary Calcium Scoring in Cardiac and Chest CT. IEEE Transactions on Medical Imaging, 1–1. https://doi.org//10.1109/tmi.2019.2899534

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

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Matsumoto, T., Kodera, S., Shinohara, H., Ieki, H., Yamaguchi, T., Higashikuni, Y., ... & Komuro, I. (2020). Diagnosing heart failure from chest X-ray images using deep learning. International Heart Journal, 61(4), 781-786. https://doi.org/10.1536/ihj.19-714

Rim, T. H., Lee, C. J., Tham, Y. C., Cheung, N., Yu, M., Lee, G., ... & Wong, T. Y. (2021). Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. The Lancet Digital Health, 3(5), e306-e316.

Rim, T. H., Lee, G., Kim, Y., Tham, Y. C., Lee, C. J., Baik, S. J., … & Cheng, C. Y. (2020). Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. The Lancet Digital Health, 2(10), e526-e536

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

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.