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

Automated Cardiovascular Disease Diagnosis from X-Ray Images Using Convolutional Neural Networks

Yeju Kim[1] and Lindy Torres[#]

[1]West High School in Torrance, USA
[#]Advisor

ABSTRACT

Cardiovascular Disease (CVD) is a leading cause of mortality worldwide, and its early and accurate diagnosis is crucial for effective treatment and patient care. Medical imaging, particularly X-ray imaging, plays a crucial role in the detection and assessment of cardiovascular abnormalities. In recent years, Convolutional Neural Networks (CNNs) have emerged as a powerful tool in medical image analysis, demonstrating promising results in various diagnostic tasks. This research paper investigates the application of CNNs for the automated diagnosis of CVD from X-ray images. The CVD diagnosis framework proposed in this study consists of three key modules. The first module is an X-ray feature extractor built using a state-of-the-art CNN architecture. The second module is an age prediction component, which accurately estimates the age of the patients from the X-ray images. Finally, the third module is the CVD classifier, which categorizes the input X-ray images into four predefined severity categories of CVD. Through extensive experiments, the proposed method has demonstrated its capability to offer novel insights into the potential use of X-ray images for predicting systemic biomarkers in the diagnosis of CVD. I expect that the proposed CVD diagnosis method can provide a significant advancement in the field of cardiovascular healthcare by offering an accurate, efficient, and automated solution for early detection of CVD.

Introduction

Cardiovascular Disease

Cardiovascular Disease (CVD) remains a leading cause of morbidity and mortality worldwide, posing a significant public health challenge. Timely and accurate diagnosis of CVD is crucial for effective management, risk stratification, and treatment planning, thereby improving patient outcomes and reducing the burden on healthcare systems. Medical imaging, particularly X-ray imaging, plays a pivotal role in diagnosing and assessing various cardiovascular conditions.

Traditionally, the diagnosis of CVD is performed on CT (Computed Tomography) scans. It has been widely employed in clinical practice for several years. CT is a non-invasive medical imaging technique that uses X-rays to create detailed cross-sectional images of the heart and blood vessels. The process involves capturing multiple X-ray images from different angles, which are then reconstructed by a computer to create a 3D representation of the cardiovascular system. CVD diagnosis using CT scans is a valuable and widely used method, but it does come with certain limitations.

Diagnosis of Cardiovascular Disease

CT scans are more complex and expensive imaging procedures compared to X-rays. As a result, access to CT facilities may be limited in certain regions or healthcare settings, particularly in resource-constrained areas. Additionally, the cost of CT scans can be prohibitive for some patients, making it difficult for them to undergo the procedure for routine screening or early diagnosis of cardiovascular conditions.

Coronary Artery Calcium Score (CACS), measured from CT scans, has emerged as a popular method for assessing the risk of CVD. The presence and extent of CACS can provide valuable information about the overall burden of atherosclerosis and can aid in predicting a patient's risk of future cardiovascular events. However, meaningful CACS values are often associated with patients who already have advanced stages of CVD or significant atherosclerotic plaque burden. As a result, the method may not be as effective in identifying individuals in the early stages of cardiovascular disease or those at high risk but showing no apparent calcification.

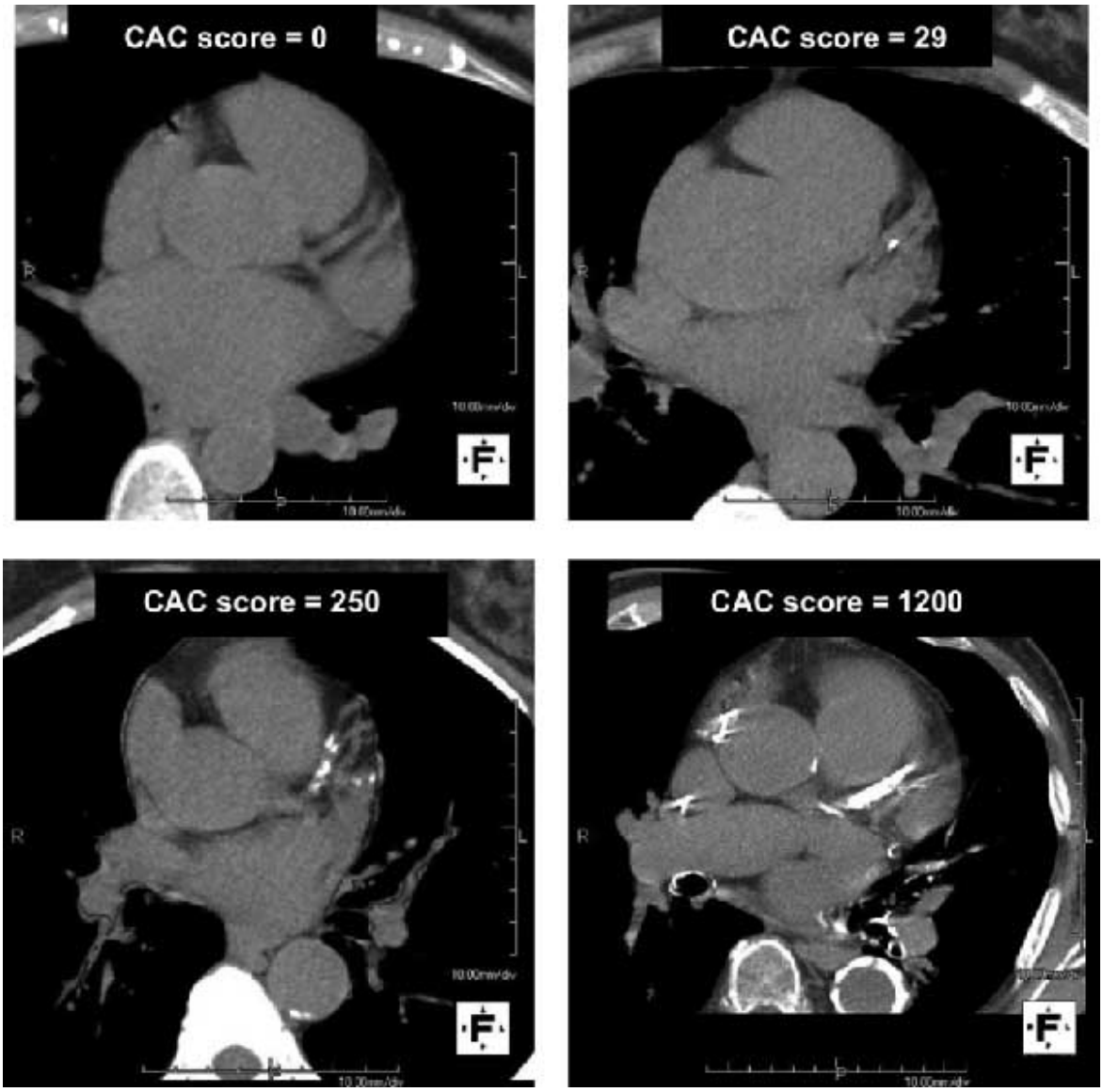

Figure 1. Example of CACS from CT scan

Proposed Method

To address the aforementioned problem, I proposed a novel CVD diagnosis method through X-ray images. The proposed method takes X-ray images as input and generates a probability-based categorization of the input X-ray images into four predefined severity categories of CVD. The input X-ray images are passed through the X-ray feature extractor, where they are transformed into feature maps, capturing essential patterns and information. These feature maps are then fed to the CVD classifier to predict the severity categories of CVD. In addition to the feature extractor and CVD classifier, I have introduced an age prediction network as part of the framework. This age prediction network provides age-aware information during the training process, recognizing that age may be correlated with the existence and progression of CVD. By incorporating age as an additional input during training, the proposed method aims to enhance the accuracy and sensitivity of the overall diagnostic process, accounting for potential age-related variations in cardiovascular conditions.

The detailed process of the proposed method and comprehensive experimental results will be presented in Chapter 3 and Chapter 4, respectively.

Related Work

Diagnosis of Cardiovascular Disease Through X-ray

Diagnosing CVD from X-ray images involves the interpretation and analysis of radiographic images of the chest to detect signs of cardiovascular abnormalities. X-ray imaging, also known as radiography, is a common and widely available diagnostic tool used in clinical practice to visualize the heart, lungs, and other structures within the chest cavity. Trained radiologists or cardiologists analyze the X-ray images to identify any signs of cardiovascular abnormalities. They look for specific features or findings that may indicate various cardiovascular conditions, such as heart enlargement, pulmonary congestion, abnormal heart shapes, or evidence of vascular abnormalities. It is important to note that while X-ray imaging is a valuable diagnostic tool, it may not be as sensitive or specific as other imaging modalities, such as echocardiography or cardiac MRI, in certain cases.

However, X-ray imaging is generally considered a more cost effective option compared to more advanced imaging modalities such as computed tomography or magnetic resonance imaging. This cost advantage has made X-ray a widely utilized input for AI-powered biomarker analysis in various types of diseases. The use of X-ray images as input data for AI-driven biomarker analysis has gained significant momentum in the medical field. It has proven particularly valuable in the assessment of various diseases, such as pulmonary conditions (e.g., pneumonia, lung cancer), skeletal disorders (e.g., fractures, arthritis), and cardiovascular diseases (e.g., coronary artery disease, congestive heart failure).

Convolutional Neural Network

Over the past decade, advancements in artificial intelligence and deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown great promise in medical image analysis, revolutionizing disease diagnosis and prognosis. CNNs consist of core building blocks known as convolutional layers, each comprising a set of learnable filters or feature detectors called kernels. These filters slide or convolve across the input image, capturing local patterns and features at various spatial locations. Through this process, the CNN can extract meaningful and hierarchical features, including CVD-related latent features, from the input X-ray image.

The hierarchical architecture of CNNs empowers them to automatically learn complex and hierarchical representations of CVD-related features present in the X-ray images. Beginning with low-level features such as edges and corners in the early layers, the network progressively learns more abstract and high-level features

in deeper layers. This ability to autonomously learn feature hierarchies is the key factor that renders CNNs so potent and effective in image recognition tasks.

In this research, I developed the proposed CVD diagnosis system heavily based on the state-of-the-art CNN architectures. As a result, the trained CNNs have become instrumental in identifying crucial cardiovascular characteristics within X-ray images, aiding in the diagnosis and management of CVD.

Proposed Approach

Architecture Overview

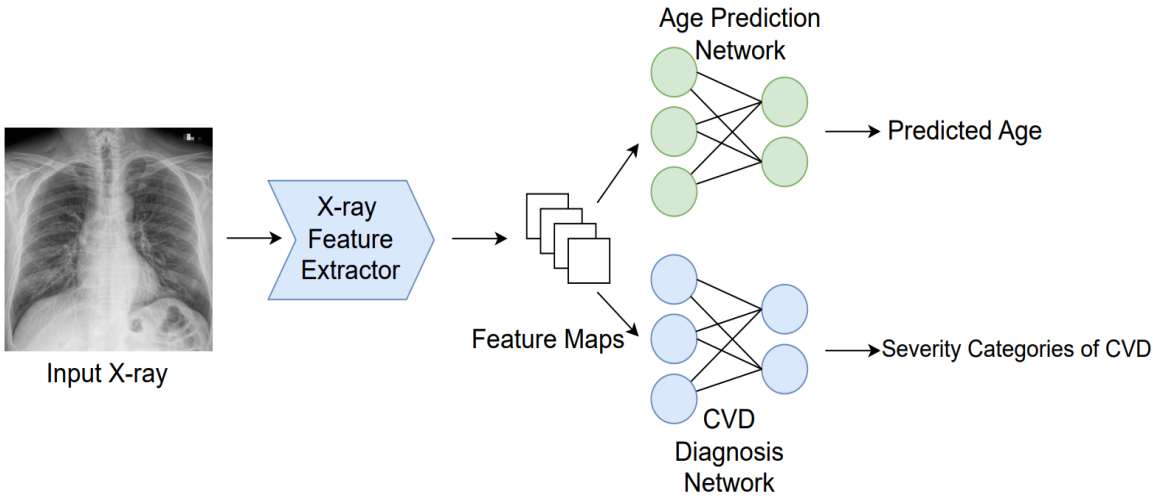

Figure 2. Overall Architecture of the Proposed Method

This chapter offers a comprehensive overview of the proposed method, including its operation and the underlying reasons behind the network's development. Figure 2 illustrates the overall architecture of the proposed methods. It consists of three modules: the X-ray Feature Extractor, the CVD Diagnosis Network, and the Age Prediction Network. Detailed information about each module will be provided in the following subchapters.

X-Ray Feature Extractor

The proposed X-ray Feature Extractor exploits a Convolutional Neural Network designed to process chest X-ray images. This network transforms the input X-ray images into feature maps that encapsulate the essential visual attributes of the images. These feature maps subsequently serve as inputs to both the age prediction network and the CVD diagnosis network. The training of the X-ray Feature Extractor involves optimizing two distinct loss functions pertinent to the downstream tasks of age prediction and classification of CVD severity categories. Throughout the training phase, the X-ray Feature Extractor learns the ability to extract crucial features, potentially harboring diagnostic cues for CVD assessment. I developed the X-ray Feature Extractor based on Resnet-50 (He et al. 2016) which shows comparable performance in image classification tasks.

CVD Diagnosis Network

The objective of the CVD Diagnosis Network is to predict the severity category of CVD. The process of assigning the input X-ray image to a specific CVD severity category hinges upon the utilization of the feature

maps derived from the X-ray Feature Extractor. For the construction of the CVD Diagnosis Network, I employed a two-layer neural network architecture. The output of the network, probability of each severity category of CVD, is then used to calculate the loss value. I utilize the cross-entropy loss function which is popularly used to train classification models. The cross-entropy loss function is explained in Chapter 3.2.

CVD Age Prediction Network

To enhance the precision of the proposed method, an Age Prediction Network is incorporated to estimate the age of patients based on the provided X-ray image. This age-aware training strategy allows the network to capture more informative feature maps that could hold relevance to CVD. Similar to the architecture of the CVD Diagnosis Network, the Age Prediction Network also adopts a two-layer neural network structure. The predicted age is compared to its ground truth in order to quantify loss value. The training procedure is explained in Chapter 3.2.

Loss Function

During the training of the proposed network, I utilized two distinct loss functions: the cross-entropy loss function for the CVD Diagnosis Network and the mean square error function for the Age Prediction Network. It is important to note that the X-ray Feature Extractor is trained by both loss functions due to the backpropagation algorithm's characteristics. The cross-entropy loss function is commonly employed for object classification tasks (Mao et al. 2023), while the mean square error function is well-suited for training regression networks that predict continuous values (Redmon et al. 2016). Equations 1 and 2 demonstrate the calculation of each respective loss function. (1)

Equation 1: Cross Entropy Loss Function

$$L_{ce} = -\log_e P$$

Here, P denotes the predicted probability of the severity categories of CVD. The loss function measures the dissimilarity between the predicted probability distribution and the actual probability distribution of classes which are the severity level of CVD. The loss value can reach zero when the predicted value perfectly aligns with its ground truth, and it can tend toward infinity in case of failure.

Equation 2: Mean Square Error Function

$$L_{mse} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2$$

Here, n represents the total number of values in the prediction, which is 1 in the proposed method. $\widehat{y_i}$ and y_i denote the predicted age and its ground truth. The mean square error function provides a way to assess how well the proposed model's predictions align with the actual values. Lower values of the function indicate that the proposed model's predictions are closer to the true values, suggesting better performance. Conversely, higher values imply that the proposed model's predictions deviate further from the actual values, indicating poorer performance. Finally, the overall loss function is constructed as a linear combination of the aforementioned individual loss functions, as shown in Equation 3.

Equation 3: Overall Loss Function

$$L = L_{ce} + \alpha L_{mse}$$

Here, *a* represents the weight value for the mean square error. Through extensive experimentation, it has been discovered that setting *a* to 0.9 yields the optimal results. These carefully chosen loss functions contribute to the effective training and optimization of the proposed neural network, facilitating accurate predictions of both CVD severity categories and age from X-ray images.

Implementation Details

In this chapter, I present a comprehensive overview of the development process behind the proposed method. The X-ray Feature Extractor in the proposed approach is based on the Resnet-50 architecture, which has proven to be effective in learning intricate features from medical images. Additionally, I implemented two neural networks: one for the Age Prediction Network and the other for the CVD Diagnosis Network. During the training process, I employed the Adam optimizer (Kingma et al. 2014) with a learning rate of 0.0001 for 100 epochs. At the 80th epoch, a learning rate decay of 0.1 was applied to fine-tune the training process. To enhance the efficiency of training, a batch size of 256 was used, and the data augmentation technique of sharpness augmentation was applied to augment the dataset.

Experimental Results

Dataset

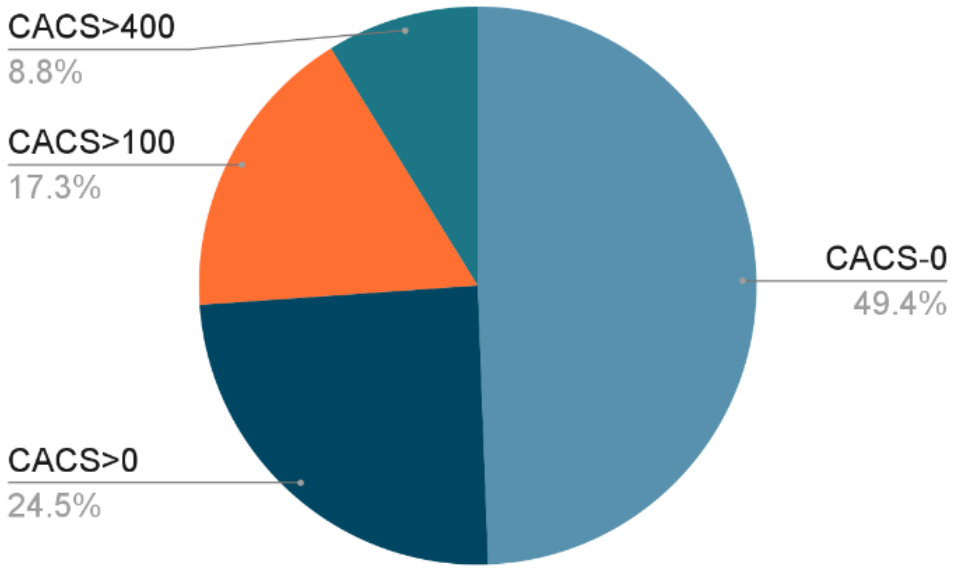

Figure 3. Label distribution of the dataset used in this research

In this section, I present a comprehensive description of the dataset used to train the proposed network. The dataset comprises a total of 39,592 X-ray samples. Each individual X-ray sample is categorized into one of four distinct groups based on its corresponding coronary artery calcium score. The distribution of these labels is illustrated in Figure 3. In terms of demographic annotations, approximately 61.67% of the samples pertain to males, while the remaining 38.32% belong to the female category. The collective average age of the patients within the dataset registers at 57.4 years.

Comparison with State-of-the-Art Method

Table 1 and Figure 4 presents a comparison of performance against state-of-the-art methods. For this comparative analysis, I selected a range of models, including VGG19 (Simonyan et al., 2014), MobileNetV2 (Sandler et al., 2018), DenseNet (Huang et al., 2017), Vision Transformer (Dosovitskiy et al., 2020), Swin Transformer (Liu et al., 2021), and ResNet-50 (He et al., 2016), all of which have demonstrated comparable performance in image classification tasks.

Table 1. Performance comparison with the state-of-the-art methods

Architecture	Accuracy	Precision	Recall	F1-Score
VGG19	0.7163	0.6785	0.6990	0.6755
MobileNetV2	0.7280	0.6789	0.7088	0.6986
DenseNet	0.7666	0.7179	0.7442	0.7308
Vision Transformer	0.7892	0.7385	0.7724	0.7658
Swin Transformer	0.7998	0.7611	0.7897	0.7695
Resnet-50	0.7886	0.7408	0.7871	0.7590
Proposed Method (Resnet-50 based)	**0.8362**	**0.7788**	**0.8190**	**0.7937**

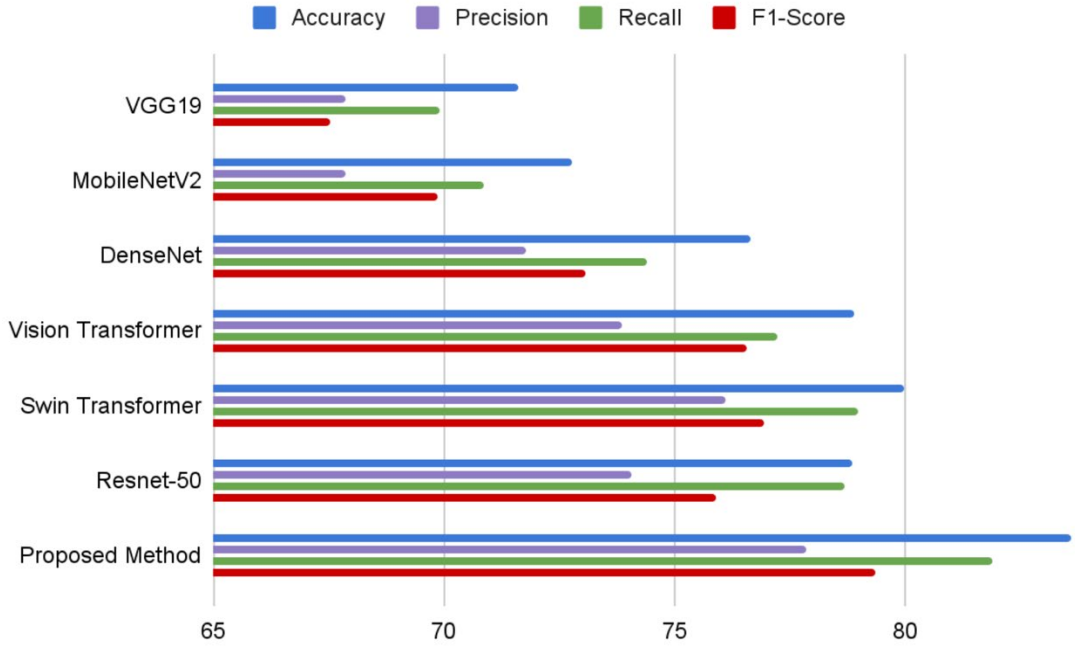

Figure 4. Performance comparison with the state-of-the-art methods

The evaluation metrics employed encompass accuracy, precision, recall, and the F1-score, widely acknowledged for quantifying the classification model's performance. In terms of results, VGG19 and MobileNetV2, due to their shallower network layers, exhibit comparatively lower accuracy. Notably, both transformer-based methods achieve accuracy levels comparable to the proposed approach. However, these methods essentially have high computational cost due to its unique data processing operations.

While Resnet-50 achieves an accuracy of 78.86, the proposed method proves its superiority by achieving the highest accuracy of 83.62. This enhanced performance can be attributed to the joint training of the age prediction network, which potentially compels the trained network to generate more precise CVD diagnostic outcomes.

Ablation Study

I also conducted an ablation study to quantify how the proposed age prediction network contributes to the overall performance. For this experiment, the network was trained in the absence of the age prediction network, focusing solely on predicting the severity category of CVD as the baseline approach. This baseline network was subsequently compared against the full model that incorporates the proposed approach. The results are presented in Table 2 and Figure 5, offering a comparison of accuracy between the baseline and the comprehensive model.

Table 2. Ablation study result (accuracy comparison)

Architecture	Accuracy (baseline)	Accuracy (Proposed Method)
VGG19	0.7163	0.7382
MobileNetV2	0.7280	0.7485
DenseNet	0.7666	0.8045
Vision Transformer	0.7892	0.8042
Swin Transformer	0.7998	0.8097
Resnet-50	0.7886	0.8362

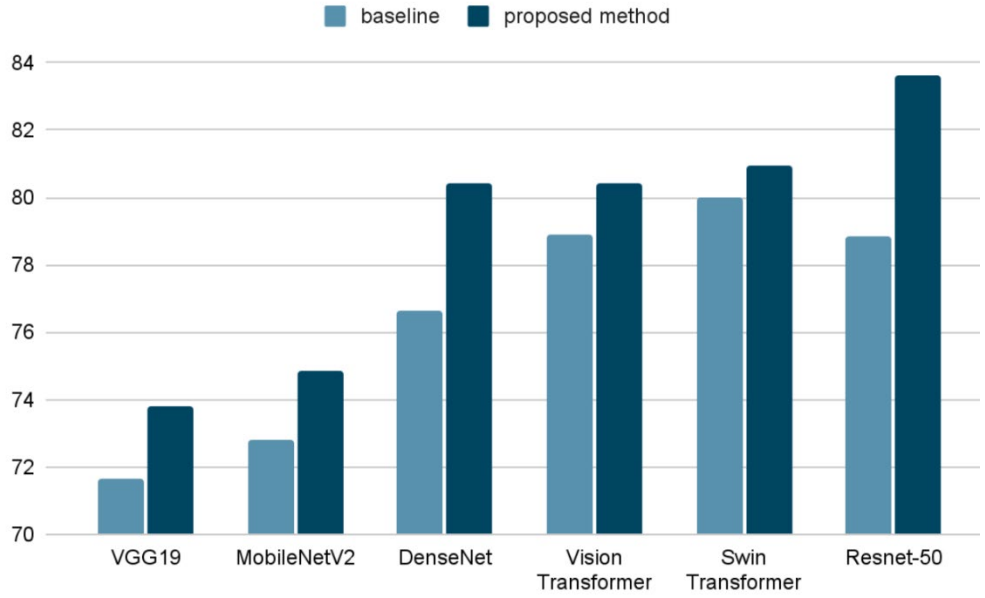

Figure 5. Ablation study result (accuracy comparison)

Application of the proposed method consistently yielded accuracy enhancements across all architectural configurations. Notably, DenseNet and ResNet-50, which have deeper network layers, exhibited substantial performance improvements when compared to shallower networks like VGG19 and MobileNetV2. This observation underscores the advantageous impact of the proposed method, particularly on models with more complex network structures.

The elements along the diagonal of this matrix provide insights into the model's proficiency in generating accurate predictions for each category. Remarkably, the proposed method maintains a consistent level of accuracy across all four CVD severity categories, thereby substantiating its robustness and reliability.

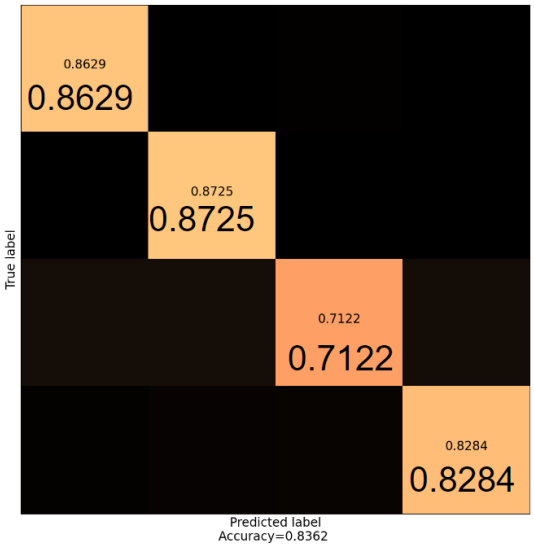

Figure 6. Confusion Matrix

Conclusion

In this research, I have explored the potential of convolutional neural networks in automating the diagnosis of CVD from X-ray images, a widely accessible and cost-effective imaging modality. Through a comprehensive analysis of convolutional neural networks architectures and optimization techniques, I proposed a novel and efficient CVD diagnosis framework that leverages the power of deep learning to extract meaningful features from X-ray images. The extensive experiments have demonstrated the superiority of the proposed method, showcasing its accuracy and robustness in categorizing X-ray images into four predefined severity categories of CVD. The results of this study have significant implications for the field of cardiovascular healthcare. The ability to automate CVD diagnosis using convolutional neural networks has the potential to transform clinical practice by providing medical professionals with a valuable decision-support tool. By expediting the diagnostic process and increasing accuracy, the proposed framework can aid in early detection, risk stratification, and personalized treatment planning for patients with cardiovascular conditions.

Acknowledgments

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

References

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. https://doi.org/10.48550/arXiv.2010.11929

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

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022). https://doi.org/10.48550/arXiv.2103.14030

Mao, A., Mohri, M., & Zhong, Y. (2023). Cross-entropy loss functions: Theoretical analysis and applications. arXiv preprint arXiv:2304.07288. https://doi.org/10.48550/arXiv.2304.07288

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788). https://doi.org/10.48550/arXiv.1506.02640

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

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980