HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 2 (2024)

Machine Learning-Based Biomarker Identification for Cardiovascular Disease Prediction: Combining X-Ray with Electrocardiogram Analysis

Heeyoon Choi[1] and Gregory Rose[#]

[1]Shanghai American School Puxi, China
[#]Advisor

ABSTRACT

Cardiovascular disease (CVD) is a major health problem worldwide. As there currently is no cure for cardiovascular disease, early diagnosis is imperative in maintaining a high quality of life for patients. Traditional method of diagnosis for cardiovascular disease involves assessing a coronary artery calcium score (CACS) through computerized tomography (CT). However, this method cannot achieve early diagnosis of cardiovascular disease. This study aims to find a novel biomarker for cardiovascular disease using X-rays and electrocardiograms (ECG). A machine learning-based approach will be used to create a multimodal algorithm that can diagnose the severity of CVD through using CACS benchmarks. Numerous experimental results indicate that the proposed multimodal approach enhances the accuracy of diagnosing the severity of CVD. The proposed multimodal approach was evaluated with several state-of-the-art convolutional neural network architectures, resulting in an impressive accuracy rate of 93.8% on a large-scale CVD dataset comprised of 21,625 patient samples. These findings demonstrate the feasibility of utilizing biomarkers such as ECG and X-ray for diagnosing CVD.

Introduction

Cardiovascular disease (CVD) is a major health problem worldwide. CVD refers to a group of conditions that affect the heart and blood vessels, often leading to various complications. CVD is indeed not curable in the traditional sense, meaning there is no treatment that can completely eliminate the risk of developing CVD or reverse all its effects once it has occurred. Therefore, early diagnosis and appropriate management are crucial for improving the quality and length of life for patients with CVD.

Traditional method of diagnosis for cardiovascular disease involves the calculation of a coronary artery calcium score (CACS) through computerized tomography (CT). CACS is indicative of coronary artery disease because calcium deposits in arterial walls of the heart is a sign of plaque, which is a sign of coronary artery disease. However, coronary artery calcium is only visible on CT when there already is plaque buildup, meaning that early diagnosis is difficult. Moreover, the traditional diagnostic method involving CT scans is expensive and time-consuming.

This study aims to find a novel biomarker for CVD using X-ray and electrocardiogram (ECG) to improve the accuracy of early diagnosis of CVD. A machine learning-based approach will be used to create a multimodal algorithm that can predict CVD, specifically predicting the degree of coronary artery classification into four classes: normal, mild, moderate, and severe.

The organization of this paper is outlined as follows: Chapter 2 offers background information to facilitate a deeper understanding of the proposed system. In Chapter 3, a detailed explanation of the multimodal

CVD diagnosis system is presented, including its architecture and training parameters. Chapter 4 showcases comprehensive experimental results, and finally, Chapter 5 offers a summary of this research.

Related Work

Chest X-Ray

Chest X-rays use X-ray light to view the structures within the thoracic cavity, which includes lungs and hearts. Chest X-ray can help diagnose various medical symptoms and diseases related to the chest area.

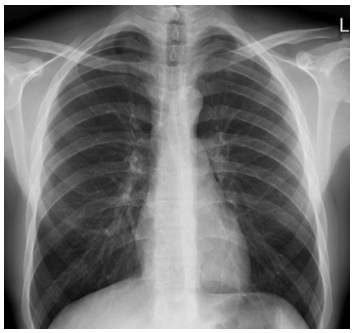

Figure 1. An example of a chest X-ray.

Previous research has employed chest X-rays in machine learning to diagnose a wide range of cardiovascular diseases. D'Ancona et al. utilized chest X-rays in deep learning with deep convolutional neural networks (DCNN) to develop a model that can detect severe coronary artery disease (D'Ancona et al. 2023). Kolossváry et al. created a deep learning model with chest X-rays that can efficiently triage emergency patients with acute chest pain, judging whether they should be deferred from further cardiovascular or pulmonary testing (Kolossváry et al., 2023). There have also been previous studies that utilized machine learning and chest X-rays to predict CACS. Kamel et al. used DCNN to develop a model that can predict CACS and CVD risk through chest X-rays; specifically, this model could conduct binary classification between a zero and non-zero CAC score, and similar binary classification with different thresholds (Kamel et al. 2021). Jeong et al. utilized radiomics score - based machine learning using random forest to predict CACS that exceeded 100 (Jeong et al. 2024).

Electrocardiogram

Electrocardiogram (ECG) is a medical test that records the electrical activity of the heart over a period of time. The heart has a natural pacemaker that generates electrical impulses which govern its movement: electrical signals flow from the sinoatrial node (SA node) via Purkinje fibers to the atrioventricular node (AV node). An ECG records these electrical signals in the heart. During an ECG, electrodes are attached to the chest, arms, and legs. These electrodes are connected to the ECG machine through wires. The ECG machine reads the electrical signals of the heart through these electrodes and records them onto a paper traveling at a rate of 25 mm/s. A signal with an amplitude of 1 millivolt (mV) moves the recording stylus 1 cm vertically. An electrical impulse traveling towards the electrode produces an upright deflection, shown as a peak going up in the ECG.

Conversely, an impulse moving away from the electrode produces a negative deflection, causing a downward peak to appear in the ECG.

ECG uses 12 leads, of which there are six chest leads (V1 to V6) and six limb leads (I, II, III, aVR, aVL, aVF). The chest leads record the electrical signals of the heart in the horizontal plane, while the limb leads record the electric signals of the heart in the vertical plane. The leads also view specific areas of the heart. Leads II, III, and aVF view the interior surface of the heart; leads V1, V2, V3, and V4 view the anterior surface of the heart; leads I, aVL, V5, and V6 view the lateral surface of the heart; leads V1 and aVR view the right atrium and the cavity of the left ventricle (Meek and Morris, 2002).

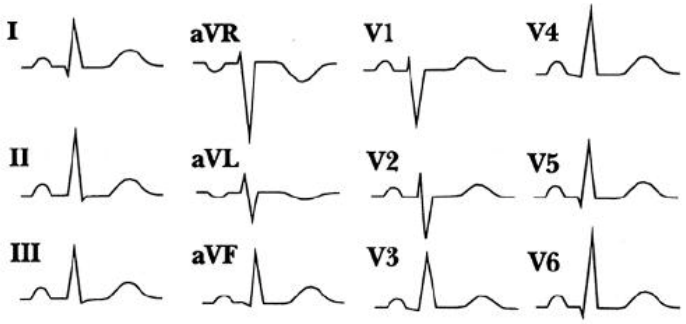

Figure 2. The waveform of the signals from each ECG node.

Research studies that utilize ECG for diagnosing medical diseases cover a broad spectrum of conditions related to the heart and cardiovascular system. Alfaras et al. introduced a machine learning-based system for heartbeat classification and arrhythmia detection, employing ECG data (Alfaras et al. 2019). Al-Zaiti et al. developed a machine learning model using ECG data to deduce an occlusion myocardial infarction (OMI) risk score (Al-Zaiti et al. 2023). Abdar et al. applied machine learning to the diagnosis of coronary artery disease (CAD) using ECG (Abdar et al. 2019). Abubaker and Babayiğit used ECG data to develop a deep learning model utilizing CNN architecture that could conduct multiple classification of CVD, including the classes of abnormal heartbeat, myocardial infarction, history of myocardial infarction, and normal (Abubaker and Babayiğit 2019). There have been studies employing ECG data and machine learning to predict CAC score. Eem et al. created a deep-learning model for the prediction of CAC score; however, the average accuracy was only approximately 72.9-80.6% (Eem et al. 2020).

In this research, I will develop a multimodal machine learning model utilizing chest X-ray and ECG that can predict the severity of the CAC score into four classes: normal, mild, moderate, severe. Chapter 3 will explain this in greater detail.

Proposed Method

This chapter will provide a detailed explanation of the proposed method. Figure 3 illustrates an overview of the proposed cardiovascular disease diagnosis system. The proposed method utilizes multimodal data which are electrocardiograms (ECG) and chest X-ray, as input for the cardiovascular disease (CVD) diagnosis system. The system consists of a 2-D convolutional neural network for processing chest X-ray and a separate 1-D CNN for analyzing ECG. The 2-D CNN generates feature maps from the chest X-ray input, while the 1-D CNN produces feature maps from the ECG input. These feature maps are then concatenated and fed into a subsequent CVD prediction neural network for further analysis.

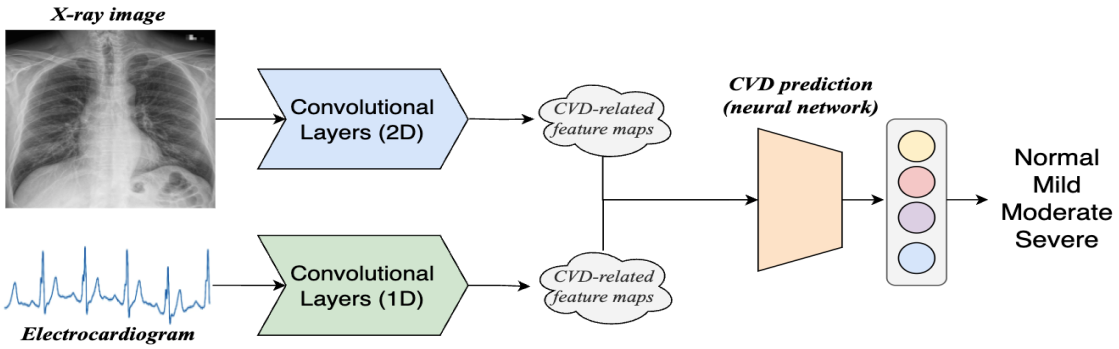

Figure 3. An overview of the proposed cardiovascular diagnosis system

I represent X-rays as $X \in R^{HW}$, where H and W denote the height and width of the input X-ray image, respectively. I denote ECGs as $E \in R^{A}$, where A denotes the time period of the ECG. For each modality, the data is first processed through convolutional layers. As X-ray images are in 2D form, it will be processed through 2D convolutional layers; on the other hand, electrocardiograms are 1D, so it will be processed through 1D convolutional layers. For every convolutional layer, the model extracts features from the data, eventually creating a CVD-related feature map for each modality. This process is represented as $CNN2D: X \rightarrow feat_X$ and $CNN1D: E \rightarrow feat_E$ for X-rays and ECG respectively. These feature maps are flattened and processed through a neural network, which predicts the severity of CVD in four categories based on predicted CACS score. Normal indicates a CACS score of 0, mild is a CACS score under 100, moderate is a CACS score under 400, and severe indicates a CACS score over 400.

To train the proposed system, I utilized the cross-entropy loss function, a commonly used function in object classification. The mathematical form of the cross-entropy loss function is in equation 1.

Equation 1: Cross-entropy loss function

$$L = -ln(p)$$

Here, p denotes the predicted probability of the model accurately classifying the object. Ideally, the loss will be zero. In the worst case, loss can be infinite.

Experimental Results

Dataset

To train and evaluate the proposed cardiovascular disease (CVD) network, I utilized a CVD dataset containing X-ray scan, ECG, and CT scan data from a total of 21,625 patients.This CVD Dataset comes from a tertiary university hospital in South Korea. The patients in this dataset did an X-ray scan and ECG for a health checkup, and then returned a few months later to take a CT scan for CACS. Hence, the chest X-rays and ECGs in this dataset can be used for the early diagnosis of severe CVD with high CACS. Of the patients in this dataset, 45.08% (9748) have a CACS of 0, 23.04% (4982) have a CACS of greater than 0 but less than 100, 20.07% (4341) have a CACS of greater than 100 but less than 400, and 11.81% (2554) have a CACS of greater than 400. These CACS boundaries are classified as normal, mild, moderate, and severe, respectively. The average age of the patients in this dataset is 56.9 years old. The dataset is 37.16% female (8036 patients) and 62.84% male (13589 patients).

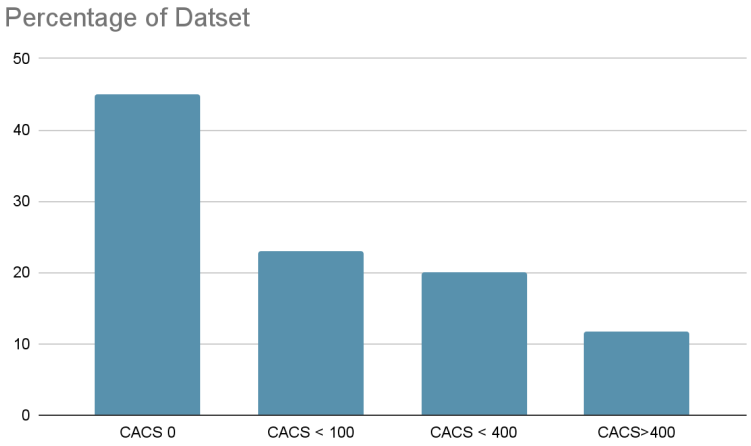

Figure 4. A graph showing the distribution of the dataset.

Inference Metric

To assess the performance of the proposed system, I employ four common object classification evaluation metrics: accuracy, recall, precision, and F1-score. The evaluation metrics use the total number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True positives are when the model correctly predicts a positive case as positive. False positives are when the model incorrectly predicts a negative case as positive. True negatives are when the model correctly predicts a negative case as negative. False negatives are when the model incorrectly predicts a positive case as negative.

Equation 2: Accuracy

$$Accuracy = \frac{TP + TN}{Total}$$

Accuracy is the ratio of correct predictions to the total number of predictions. The total number of correct predictions is the sum of all true positives and true negatives.

Equation 3: Recall

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

Recall is used to evaluate the model's performance in accurately identifying positive cases as positive. It is the ratio of true positives to the total number of positive cases in the dataset. The total number of positive cases in the dataset is the sum of the true positives and false negatives.

Equation 4: Precision

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

Precision is the ratio of true positives to all cases which the model predicted as positive. The total number of cases which the model predicted as positive is the sum of true positives and false positives.

Equation 5: F1-Score

$$F1\text{–}Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

The F1-score is the harmonic mean of the recall and the precision. It is a metric that indicates the balance between precision and recall.

Table 1. Performance comparison with state-of-the-art convolutional neural network architectures.

	Accuracy	Recall	Precision	F1-Score
VGG-19 (Simonyan et al. 2014)	0.9098 (±0.0005)	0.9104 (±0.0007)	0.8996 (±0.0009)	0.9050 (±0.0008)
MobileNetV2 (Sandler et al. 2018)	0.9105 (±0.0009)	0.9084 (±0.0011)	0.8993 (±0.0010)	0.9038 (±0.0008)
HRNet-W40 (Sun et al. 2019)	0.9124 (±0.0013)	0.9125 (±0.0015)	0.9061 (±0.0008)	0.9093 (±0.0010)
SE-ResNet-101 (Hu et al. 2018)	0.9276 (±0.0010)	0.9276 (±0.0009)	0.9198 (±0.0006)	0.9237 (±0.0013)
ResNeXt-101 (Xie et al. 2017)	0.9304 (±0.0011)	0.9306 (±0.0007)	0.9240 (±0.0008)	0.9273 (±0.0012)
ResNet-101 (He et al. 2016)	**0.9383** (±0.0007)	**0.9384** (±0.0008)	**0.9272** (±0.0010)	**0.9328** (±0.0011)

I trained the CVD network on various architectures to compare the effectiveness. Of the various architectures, VGG-19 (Simonyan et al. 2014) and MobileNetV2 (Sandler et al. 2018) are comparably shallow networks. HRNet-W40 (Sun et al. 2019) is a deeper network, and SE-ResNet-101 (Hu et al. 2018), ResNeXt-101 (Xie et al. 2017), and ResNet-101 (He et al. 2016) are very deep neural networks. It can be seen from Table1 that accuracy increases as the network becomes deeper. Overall, the lowest accuracy was 90.98%, from VGG-19, and the highest accuracy was 93.83%, from ResNet-101. These high values of accuracy indicate the feasibility of taking a multimodal approach using chest X-rays and ECG to the early diagnosis of a high CACS. The standard deviation values from K-folds cross validation is at most 1.3% with HRNet-W40 and at least 0.05% with VGG-19, which is low.

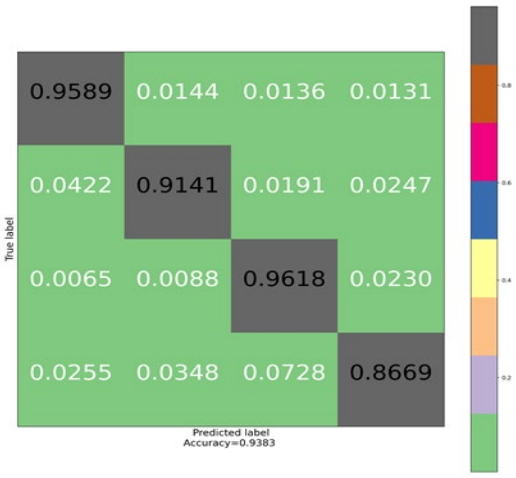

Figure 5. A confusion matrix of the proposed cardiovascular disease diagnosis system.

Figure 5 shows the confusion matrix from the training. The labels are from left to right and up to down CACS=0, CACS<100, CACS<400, and CACS>400, or normal, mild, moderate, and severe. The labels CACS<400 and CACS=0 have the greatest accuracy, followed shortly after by the label CACS<100. The CACS>400 label has a notably lower accuracy of 86.69% when data to the other labels. It can be seen that data of this label are confused with images from the CACS<400 label, showing that the model may have difficulty in making the judgment of when CACS is severe.

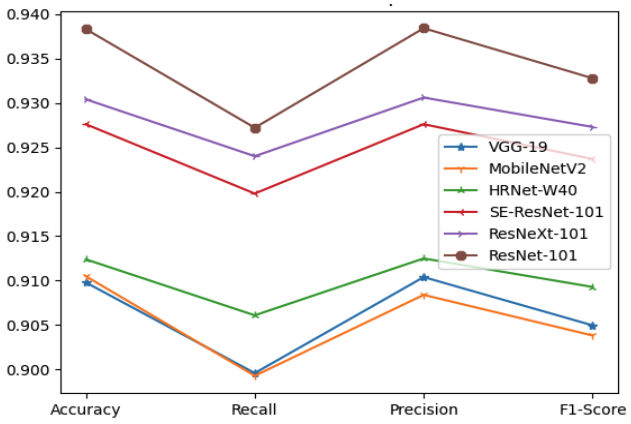

Figure 6. Performance comparison with state-of-the-art convolutional neural network architectures.

Figure 6 is a graph of each model's performance across the four inference metrics. It shows that performance increases in all inference metrics with deeper neural networks. ResNet-101 performs the best in all metrics in addition to accuracy, and the relatively shallow models, VGG-19, MobileNetV2, and HRNet-W40, show significantly lower performance compared to the deeper models.

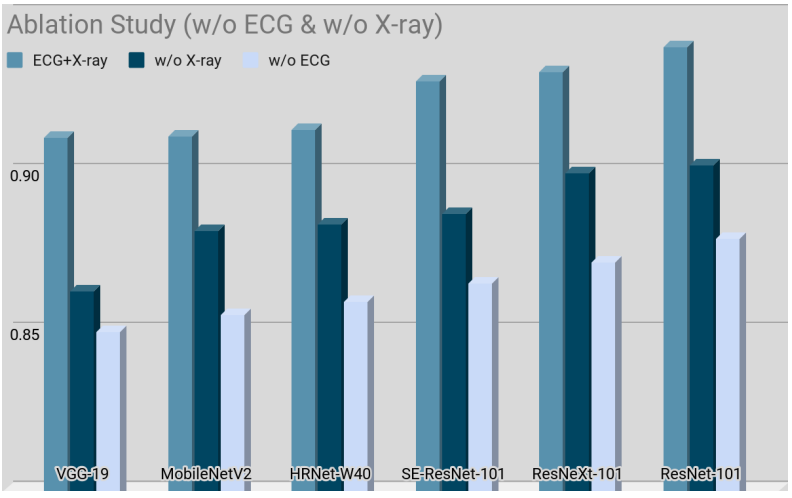

Figure 7. Results of an ablation study

Table 2. Result of an ablation study.

	Accuracy		
	Without ECG	**Widthout X-ray**	**X-ray + ECG**
VGG-19 (Simonyan et al. 2014)	0.8487 (±0.0011)	0.8617 (±0.0007)	0.9098 (±0.0005)
MobileNetV2 (Sandler et al. 2018)	0.8543 (±0.0009)	0.8804 (±0.0008)	0.9105 (±0.0009)
HRNet-W40 (Sun et al. 2019)	0.8584 (±0.0012)	0.8825 (±0.0010)	0.9124 (±0.0013)
SE-ResNet-101 (Hu et al. 2018)	0.8642 (±0.0010)	0.8861 (±0.0009)	0.9276 (±0.0010)
ResNeXt-101 (Xie et al. 2017)	0.8707 (±0.0007)	0.8986 (±0.0008)	0.9304 (±0.0011)
ResNet-101 (He et al. 2016)	0.8783 (±0.0008)	0.9014 (±0.0009)	0.9383 (±0.0007)

Figure 7 and Table 2 both show that the model's performance regardless of architecture improves significantly with the inclusion of two modals. The graph also indicates that the accuracy with only ECG is higher than that with only X-rays, showing that ECG contributes more to the model than X-rays.

Conclusion

In this study, I proposed the use of chest X-rays and ECG in a multimodal machine-learning based approach for the early diagnosis of CVD using CACS benchmarks. I trained models of various architectures on the classification task and obtained evaluation metrics, which indicated that overall accuracy exceeded 90% for all models used. Moreover, accuracy significantly improved with deeper CNNs such as ResNet-101. Afterwards, I conducted ablation studies with the same models and found that using a multimodal approach with chest X-ray and ECG greatly improves the model's accuracy, with ECG playing the more important role. Seeing such results, I expect that this multimodal machine-learning approach using chest X-rays and ECG can be a novel biomarker for CVD. For future research, I aim to use a different dataset for external validation, and use other possible biomarkers, such as retinal fundus imaging, in expanding the multimodal algorithm.

Acknowledgments

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

References

Abdar, M., Książek, W., Acharya, U. R., Tan, R. S., Makarenkov, V., & Pławiak, P. (2019, October 1). *A new machine learning technique for an accurate diagnosis of coronary artery disease*. Computer Methods and Programs in Biomedicine. https://doi.org/10.1016/j.cmpb.2019.104992.

Abubaker, M.B., and Babayiğit, B (2022). Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods. IEEE Transactions on Artificial Intelligence, 4(2), https://doi.org/10.1109/TAI.2022.3159505.

Alfaras, M., Soriano, M. C., & Ortín, S. (2019). A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. Frontiers in Physics, 103.

Al-Zaiti, S. S., Martin-Gill, C., Zègre-Hemsey, J., Bouzid, Z., Faramand, Z., Alrawashdeh, M., Gregg, R. E., Helman, S., Riek, N. T., Kraevsky-Phillips, K., Clermont, G., Akçakaya, M., Sereika, S. M., Van Dam, P., Smith, S. W., Birnbaum, Y., Saba, S., Sejdić, E., & Callaway, C. W. (2023, June 29). *Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction*. Nature Medicine. https://doi.org/10.1038/s41591-023-02396-3

D'Ancona, G., Massussi, M., Savardi, M., Signoroni, A., Di Bacco, L., Farina, D., Metra, M., Maroldi, R., Muneretto, C., İnce, H., Costabile, D., Murero, M., Chizzola, G., Curello, S., & Benussi, S. (2023, January 1). *Deep learning to detect significant coronary artery disease from plain chest radiographs AI4CAD.* International Journal of Cardiology. https://doi.org/10.1016/j.ijcard.2022.10.154

Eem, C., Hong, H. K., & Noh, Y. (2020, December 7). *Deep-Learning Model to Predict Coronary Artery Calcium Scores in Humans from Electrocardiogram Data.* Applied Sciences. https://doi.org/10.3390/app10238746

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

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

Jeong, H., Park, H.-B., Hong, J., Lee, J., Ha, S., Heo, R., Jung, J., Hong, Y., & Chang, H.-J. (2023). Identifying Coronary Artery Calcification Using Chest X-ray Radiographs and Machine Learning. *Journal of Thoracic Imaging*. https://doi.org/10.1097/rti.0000000000000757

Kamel, P.I., Yi, P.H., Sair, H.I., and Lin, C.T (2021). Prediction of Coronary Artery Calcium and Cardiovascular Risk on Chest Radiographs Using Deep Learning. Radiology: Cardiothoracic Imaging, 3(3), https://doi.org/10.1148/ryct.2021200486

Kolossváry, M., Raghu, V.K., Nagurney, J.T., Hoffman, U., and Lu, M.T (2023). Deep Learning Analysis of Chest Radiographs to Triage Patients with Acute Chest Pain Syndrome. Radiology, 306(2), https://doi.org/10.1148/radiol.221926.

Morris, F., Brady, W. J., & Camm, A. J. (Eds.). (2009). ABC of clinical electrocardiography. John Wiley & Sons.

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

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