HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 2 (2024)

Enhancing Cardiovascular Disease Detection through Deep Multimodal Fusion: Integrating Radiology and Electrocardiogram via Convolutional Neural Network

Kyungryun Kim

Harry Fried, USA

ABSTRACT

Cardiovascular Diseases (CVDs), also known as heart diseases are related to a process of atherosclerosis, which is when a plaque builds up in the arteries and blocks blood flow. CVDs are an ever-increasing health risk for the general population, and according to the American Heart Association, the number of deaths related to CVDs has surpassed the previous 910,000 all-time record from 2003. To solve this problem, numerous novel approaches have been presented for a solution to decrease such anomalies with traditional methods utilizing computed tomography scans (CT scans) to calculate coronary artery calcium. However, countless problems arise from using CT scans as CT scans are not accessible to the general public, and early diagnosis is especially difficult. So, this research paper aims to explore the potential of AI-driven biomarkers in enhancing the efficiency of CVD diagnosis. Moreover, this study presents a novel approach that uses artificial intelligence to enhance the accuracy of CVD diagnosis. Specifically, this paper suggests an innovative methodology that utilizes the classification of x-ray images that is additionally supplemented by classifying electrocardiograms. These techniques allow for analysis of multi-modal data sources allowing latent information on CVDs to be extracted. Throughout the experiments, the proposed method has proven that it is superior compared to other state-of-the-art methods, as it has achieved an accuracy of 90.49%.

Introduction

Problem

The term cardiovascular disease, or CVD, refers to diseases of the heart and blood vessels. With CVDs on the rise in America, it is important to take precautions to reduce the number of deaths due to them. Traditionally, coronary artery calcium has been measured with computed tomography (CT) scans to diagnose CVDs. This traditional method, however, has several disadvantages. First, early diagnosis of CVDs is difficult as the symptoms can be atypical. Second, CT scans are also costly and inaccessible most of the time. As a final point, the use of only health markers (biomarkers) as a measure of health is limited. To address these issues, several studies have focused on the early-stage detection of CVD using convolutional neural networks (CNN).

Previous Method

Several methods have been developed to detect CVD in patients. For instance, Poplin et al. proposed the classification of retinal photographs to predict various health markers of a person (Poplin et al. 2018). The retina is a special organ that allows us to see the microvasculature and neural tissues without invasive procedures. By analyzing the retina, deep learning models can detect systemic diseases (cardiovascular risk factors). The previous model utilized the VGG16 network to analyze retinal photographs. Weng et al. introduced a CVD diagnosis system that utilizes routine clinical data obtained from patients (Weng et al. 2017). Their research demonstrates the promise and feasibility of employing machine learning for CVD diagnosis. While researchers are currently engaged in active investigation, they still face challenges with low accuracy. There is a pressing need to develop and enhance such systems to achieve more accurate results.

Related Work

This chapter presents a comprehensive overview of the related work and existing research conducted on cardiovascular disease (CVD) detection utilizing X-ray and ECG data.

Cardiovascular Disease

Cardiovascular disease (CVD) is a leading cause of mortality worldwide, encompassing a wide range of conditions that affect the heart and blood vessels. CVD includes diseases such as coronary artery disease, heart failure, arrhythmias, and valvular heart disease. So, timely detection and accurate diagnosis of CVD are crucial for effective treatment and prevention of complications.

Various diagnostic techniques and tools are employed in the detection and evaluation of CVD. These include clinical history and physical examination, electrocardiography (ECG), echocardiography, stress testing, cardiac biomarkers, cardiac imaging (such as X-ray and magnetic resonance imaging), and invasive procedures like cardiac catheterization. While these methods have proven efficacy, there is a need for more accurate and efficient diagnostic approaches. In recent years, machine learning techniques have gained significant attention for their potential to improve CVD detection and diagnosis accuracy and efficiency. Specifically, convolutional neural networks (CNNs), a type of deep learning algorithm, have shown promising results in various medical applications, including the analysis of medical images and physiological signals.

In this research, a novel multi-modal CVD detection framework is developed utilizing both X-ray and ECG data. The proposed framework combines the strengths of these two diagnostic modalities to improve the accuracy and efficiency of CVD detection.

Object Classification

Object classification is a fundamental task in the field of computer vision, which focuses on the identification and categorization of objects within images. It plays a crucial role in a wide range of applications, including but not limited to autonomous driving, robotics, surveillance systems, image retrieval, and augmented reality. The primary objective of object classification is to empower machines with the capability to accurately recognize and assign appropriate labels or categories to objects present in visual data. Object classification has practical foundations in various domains, including facial expression recognition and cancer-type classification from microscopic images.

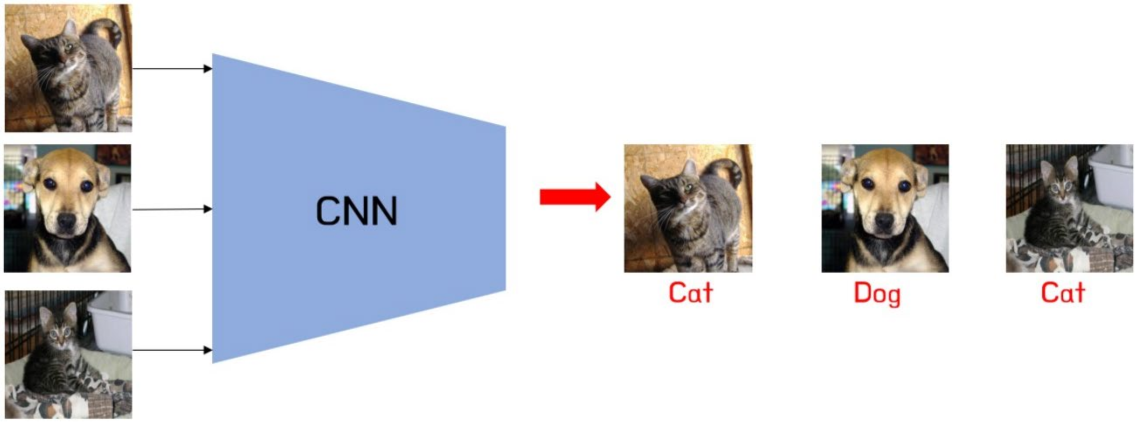


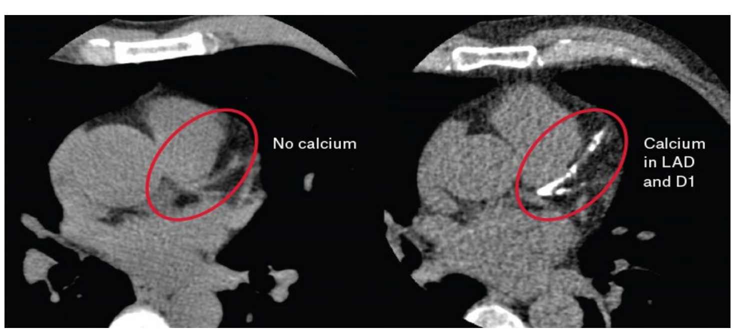
Figure 1. Object Classification of Dogs and Cats (Kumar 2022)

To implement the object classification system, numerous machine learning techniques have been proposed, including neural networks and convolutional neural networks (CNNs). CNNs have emerged as a compelling approach for object classification tasks. In this research, CVD detection is considered as an object classification task, where the model analyzes X-ray images and ECG signals as input and assigns the corresponding category along with a specific level of Coronary Artery Calcium Score (CACS). Chapter 3 provides a comprehensive explanation of the proposed method, presenting detailed information on the implementation of the proposed method and key components.

X-Ray and Electrocardiogram

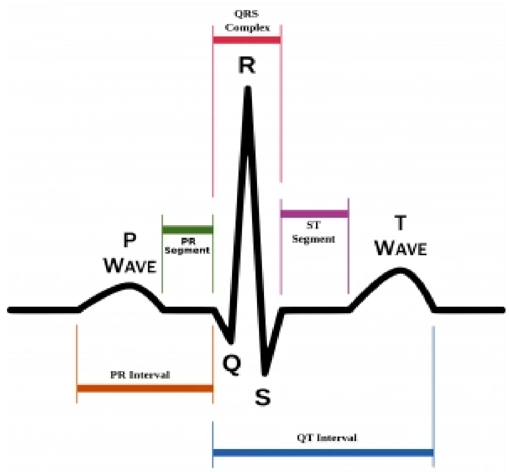
X-ray imaging is a widely used diagnostic tool in the field of medicine for capturing images of internal structures of the body. It is based on the principles of X-ray radiation, where the X-ray emits ionizing radiation, a form of electromagnetic radiation with higher energy than visible light. The radiation can penetrate through soft tissues but is absorbed differently by various body structures. For instance, denser bones and organs absorb more radiation compared to fat tissues and air-filled spaces. This is why X-ray imaging has long been used in the diagnosis of many heart-related diseases, providing valuable information about the size, shape, and condition of the heart.

On the other hand, ECG (Electrocardiography) is a non-invasive diagnostic technique used to record and analyze the electrical activity of the heart. It involves placing electrodes on the skin's surface in specific locations, which detect and measure the electrical signals generated by the heart as it contracts and relaxes. These signals, known as electrocardiograms, provide valuable insights into the heart's rhythm, rate, and overall electrical behavior.



(a)

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



(b)

Figure 2. CAC region in CT scan and ECG signal. (a): Calcium-filled artery compared to a normal artery (Chua et al. 2020), and (b): ECG signal (Ryan 2020)

These two different types of data can be an important cue for detecting CVD since they provide complementary information about the structure and function of the heart. X-ray images offer insights into the anatomical aspects of the cardiovascular system. Calcified heart vessels can be readily examined in X-ray imaging as they often appear as distinct white dots or smudges, making them easily detectable. Furthermore, ECG data contains latent information that holds the potential to detect cardiovascular disease (CVD). The electrical signals recorded in an ECG waveform can reveal important indicators of cardiac abnormalities, including irregular heart rhythms, conduction disorders, and ischemic events. By analyzing the characteristic patterns and features of the ECG signals, clinicians and researchers can uncover subtle changes or deviations that may signify the presence of CVD.

This research focuses on utilizing two distinct types of data as inputs to the proposed CVD detection framework, aiming to extract meaningful features related to cardiovascular disease for the development of an accurate diagnosis system. The subsequent chapter will provide detailed information regarding the overall architecture of the framework and the specific methodologies employed for processing these multi-modal data.

Proposed Cardiovascular Disease Diagnosis System

HIGH SCHOOL EDITION
Journal of Student Research

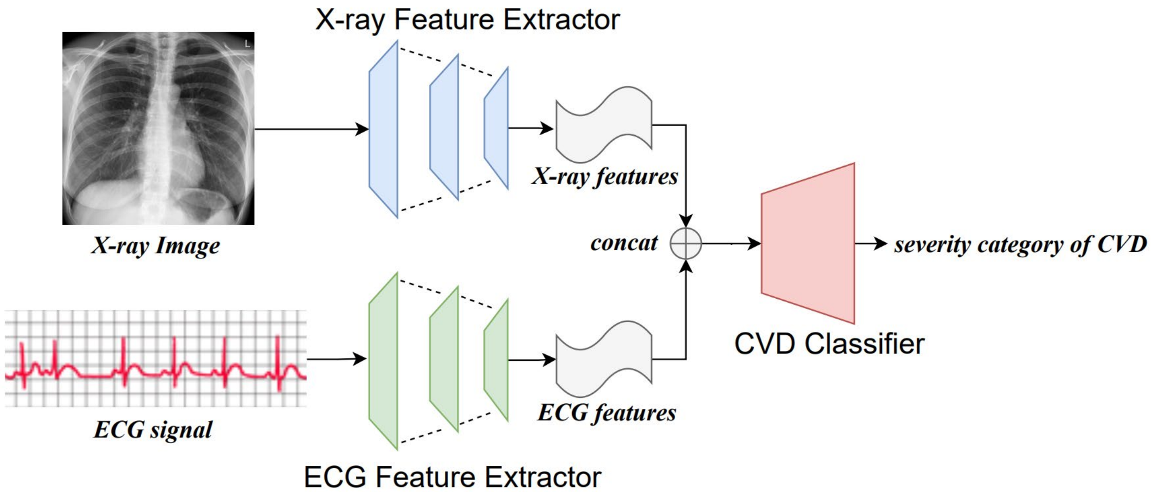


Figure 3. Overall Architecture of the Proposed Method

Figure 3 represents the overall architecture of the proposed multi-modal Cardiovascular Disease (CVD) detection system. The network consists of three modules: 2-D CNN, 1-D CNN, and the Agatston Score calculation. In the 2-D CNN module, the input image, I, is the X-ray image. In the 1-D CNN, the input is the patient Electrocardiogram (ECG) graph, namely, E. Then the X-ray feature extractor, X-Ext, and the ECG feature extractor, E-Ext both extract information from the given input data. Once that information is stored, the X-ray features, (z_x) and the ECG Features, (z_e) are combined to create a concat layer of information, Z. to determine the Agatston score, an evaluation metric for classifying the severity of coronary artery calcification (CAC) build up in arteries in a probability, p. By integrating information from X-rays and ECG data, the multi-modal CVD detection system, or CLS offers a state-of-the-art assessment of the patient's cardiac health.

Table 1. Notation used in this research paper

Common Notations
X-ray image: I
ECG graph: E
X-ray feature extractor: X-Ext
ECG feature extractor: E-Ext
X-ray features: z_x
ECG features: z_e
Concat: : z
CVD classifier: CLS

severity : *p*

Experimental Results

Dataset

In this chapter, I provide an in-depth description of the dataset employed in this research for diagnosing cardiovascular disease (CVD) using X-ray and electrocardiogram (ECG) data, coupled with severity labels based on coronary artery calcium scores (CACS). The dataset comprises a total of 18,945 samples, each accompanied by corresponding X-ray and ECG data, as well as severity labels. The dataset is intended to provide a representative distribution of CVD cases across varying age groups and genders. The average age of the individuals in the dataset is 54.7 years, and the distribution of sex is slightly uneven, with 7,541 samples from females and 11,404 samples from males.

Table 2. Dataset overview

Label	Explanation
Total Samples	18,945
Data Types	X-ray and ECG
Severity Labels	Based on Coronary Artery Calcium Scores
Average Age	54.7 years
Gender Distribution	Female: 7,541 samples and Male: 11,404 samples

Table 3. Category distribution

Severity Classes	CACS Range	Description
Class 1	CACS = 0	No detectable coronary artery calcium, low risk
Class 2	0 < CACS < 100	Detectable coronary artery calcium, elevated risk
Class 3	100 < CACS < 400	Significant likelihood of CVD
Class 4	CACS > 400	Substantial risk of severe CVD

Evaluation Protocol

To prove the efficacy of the proposed method, it is compared with previous supervised approach based state-of-the-art methods for CVD diagnosis. The 5-fold cross-validation protocol is applied to each of these methods, and their performances are evaluated and compared against the proposed method in terms of accuracy, precision, recall, and F1-score. I also conducted an ablation study to gain insights into the contributions of individual components of the proposed method to the overall accuracy.

Performance Comparison

Table 4. Performance comparison result

	Accuracy	Recall	Precision	F1-Score
VGG19 (Simonyan et al. 2014)	0.8589 (±0.0012)	0.8555 (±0.0007)	0.8160 (±0.0012)	0.8314 (±0.0013)
MobileNetV2 (Sandler et al. 2018)	0.8616 (±0.0014)	0.8595 (±0.0008)	0.8154 (±0.0009)	0.8328 (±0.0014)
EfficientNet-B7 (Tan et al. 2019)	0.8654 (±0.0007)	0.8631 (±0.0005)	0.8270 (±0.0008)	0.8460 (±0.0007)
Xception (Fran et al. 2017)	0.8791 (±0.0009)	0.8761 (±0.0013)	0.8308 (±0.0008)	0.8499 (±0.0005)
HRNet-w32 (Wang et al. 2020)	0.8952 (±0.0006)	0.8898 (±0.0010)	0.8529 (±0.0011)	0.8618 (±0.0011)
Resnet-50 (He et al. 2016)	0.9049 (±0.0007)	**0.9021 (±0.0008)**	**0.8577 (±0.0011)**	**0.8763 (±0.0010)**

To prove the model's excellent performance, I compared it to other previous state-of-the-art methods across all categories (Accuracy, Recall, Precision, and F-1 Score). Before reviewing the model's performance with others, it is important to understand the concepts in the table first. First and foremost, all the components in this table refer back to the idea of confusion matrices. Accuracy refers to the number of cases that the model predicted right (level of CaC) over the total number of cases in the dataset. Next, Recall refers to how many did the models got correct about CaC levels out of all the positive samples that the dataset had. The precision is referred to as when the model predicts that there is cAc and out of those how many had severe CaC levels. Finally, the F-1 score is the average of the recall and precision.

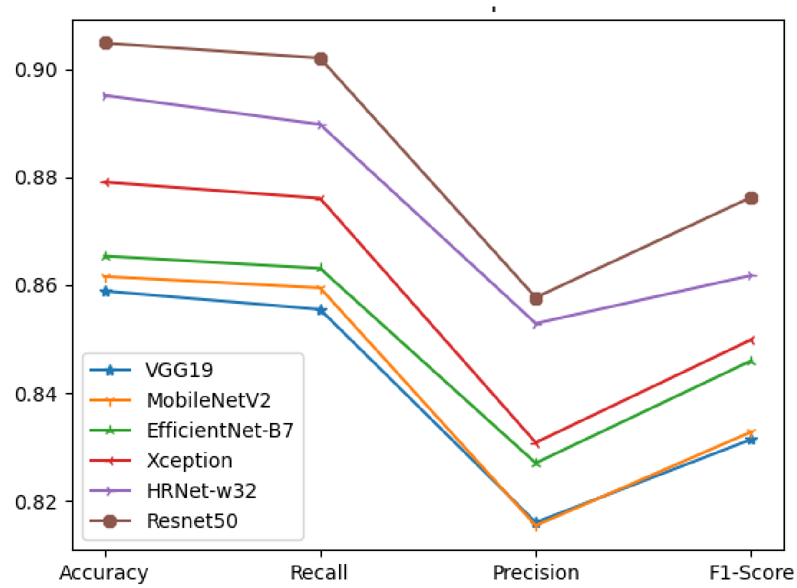


Figure 4. Performance Metrics Comparison: Accuracy, Recall, Precision, and F-1 Score across Different Networks

As shown in table 4 and figure 4, the proposed model (Resnet50) outperforms all other methods in all categories including Accuracy (90.49%), Recall (90.21%), Precision (85.77%), and F-1 Score (87.63%).

Convolutional Neural Network Architecture Replacement

Table 5. Ablation study result

	Accuracy (w/o ECG)	**Accuracy (w/o X-ray)**	**Accuracy (full model)**
VGG19 (Simonyan et al. 2014)	0.8384 (±0.0014)	0.8048 (±0.0016)	0.8589 (±0.0012)
MobileNetV2 (Sandler et al. 2018)	0.8451 (±0.0011)	0.8124 (±0.0011)	0.8616 (±0.0014)
EfficientNet-B7 (Tan et al. 2019)	0.8560 (±0.009)	0.8247 (±0.0010)	0.8654 (±0.0007)
Xception (Fran et al. 2017)	0.8674 (±0.0013)	0.8347 (±0.008)	0.8791 (±0.0009)
HRNet-w32 (Wang et al. 2020)	0.8905 (±0.0009)	0.8593 (±0.0008)	0.8952 (±0.0006)
Resnet-50 (He et al. 2016)	0.8874 (±0.0010)	0.8684 (±0.0009)	0.9049 (±0.0007)

The table 5. shows the 6 methods that we used to extrapolate the severity of the CAC score. The table also examines the idea of an ablation study, which is conducted to determine the difference in efficiency between the full model versus the model absent of the X-ray or ECG. As you can see, the Resnet-50 achieved an accuracy of 88.74% without the ECG component and an accuracy of 86.84% without the X-ray. From these results, we can assume that the ECG plays a larger role in predicting CAC levels than the X-ray and that the full model with all the components outperforms the ablation models (90.49% in accuracy).

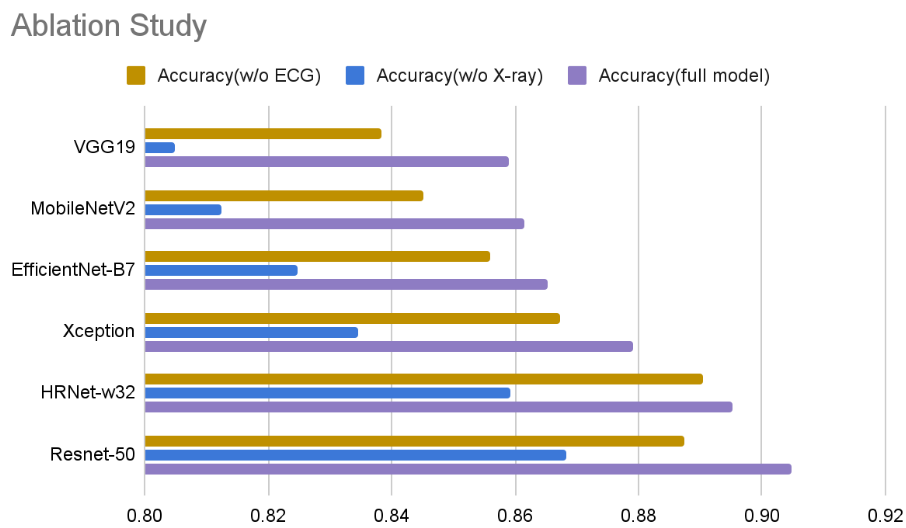


Figure 5. Ablation Study on State-of-the-Art Models

Conclusion

In this research project, I introduced an innovative training approach aimed at classifying the severity of the CAC scores in heart arteries. The proposed method encompasses the components of ECG and X-ray in order to calculate the amount of coronary artery calcium. For the network, I used Resnet-50, and as shown throughout the research, it outperforms previous state-of-the-art methods. So for the evaluation method, I used a confusion matrix and calculated the most important components that are statistically significant. The four components are recall, precision, accuracy, and F-1 score. In addition, I conducted an ablation study to determine how the efficiency of the model deteriorates when we remove one of the key components (either X-ray or ECG). In the future, I plan on introducing my model to other domains of the medical field that need this technology, including radiologists, endocrinologists, and internal medicine.

Acknowledgments

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

References

Chua, A., Blankstein, R., and Ko, B. (2020, Aug 1). *"Coronary artery calcium in primary prevention"*: Australian Journal of General Practice.
https://www1.racgp.org.au/ajgp/home

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

Kumar, Ajitesh. (2022, Aug 1). *"Classification Problems Real-life Examples"*: Analytics Yogi. https://vitalflux.com/classification-problems-real-world-examples/

Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., ... & Webster, D. R. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature biomedical engineering, 2(3), 158-164.

Ryan, Steve. (2020, Aug 26). *"Understanding the EKG Signal"*: Atrial Fibrillation Resources for Patients. https://a-fib.com/treatments-for-atrial-fibrillation/diagnostic-tests-2/the-ekg-signal/

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

Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR. https://doi.org/10.48550/arXiv.1905.11946

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

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?. PloS one, 12(4), e0174944.