January 2023

Applying Deep Learning To Identify Imaging Biomarkers To Predict Cardiac Outcomes In Cancer Patients

Aishwarya Kishore Nene

Applying Deep Learning to Identify Imaging Biomarkers to Predict Cardiac Outcomes in Cancer Patients

A Thesis Submitted to the Yale University School of Medicine in Partial Fulfillment of the Requirements for the Degree of Doctor of Medicine

by

Aishwarya K. Nene, Class of 2023

Table of Contents

Abstract

APPLYING DEEP LEARNING TO IDENTIFY IMAGING BIOMARKERS TO PREDICT CARDIAC OUTCOMES IN CANCER PATIENTS

Aishwarya K. Nene, Crystal Cheung, Arman Avesta, Sajid Hossain, Harlan Krumholz, and Sanjay Aneja, Department of Therapeutic Radiology, Yale University, School of Medicine, New Haven, CT.

Cancer patients are a unique population with increased mortality from cardiovascular disease, however only half of high-risk patients are medically optimized. Physicians ascertain cardiovascular risk from several risk predictors using demographic information, family history, and imaging data. The Agatston score, a measure of total calcium burden in coronary arteries on CT scans, is the current best predictor for major adverse cardiac events (MACE). Yet, the score is limited as it does not provide information on atherosclerotic plaque characteristics or distribution. In this study, we use deep learning techniques to develop an imaging-based biomarker that can robustly predict MACE in lung cancer patients. We selected participants with screen-detected lung cancer from the National Lung Screening Trial (NLST) and used cardiovascular mortality as our primary outcome. We applied automated segmentation algorithms to low-dose chest CT scans from NLST participants to segment cardiac substructures. Following segmentation, we extracted radiomic features from selected cardiac structures. We then used this dataset to train a regression model to predict cardiovascular death. We used a pre-trained nnU-Net model to successfully segment large cardiac structures on CT

scans. These automated large cardiac structures had features that were predictive of MACE. We then successfully extract radiomic features from our areas of interest and use this high-dimensional dataset to train a regression model to predict MACE. We demonstrated that automated segmentation algorithms can result in low-cost non-invasive predictive biomarkers for MACE. We were able to demonstrate that radiomic feature extraction from segmented substructures can be used to develop a high-dimensional biomarker. We hope that such a scoring system can help physicians adequately determine cardiovascular risk and intervene, resulting in better patient outcomes.

Acknowledgments

This dissertation is the result of my own work and does not include work done in collaboration, except where specifically mentioned in the text. I would like to thank the following people for their guidance and support during my MD thesis project. Firstly, I wish to thank my supervisors Dr. Sanjay Aneja and Dr. Harlan Krumholz for giving me this opportunity and providing guidance and support. I am grateful for all the help Dr. Arman Avesta, Sajid Hossain, and Crystal Cheung have provided me from experimental technique training to data analysis on my project. I am thankful to the research team at Visage Imaging, especially Khaled Bousabarah for uploading our data to the Visage API and implementing the segmentation algorithms within the Visage API. Everyone in the Center for Outcomes Research and Evaluation has been incredibly welcoming and made this year very enjoyable. Lastly, I wish to acknowledge the Yale Fellowship for Medical Student Research for its generous financial support during my thesis project.

Chapter 1: Introduction

1.1 Cardiovascular disease

Mortality and healthcare burden from cardiovascular disease has been increasing globally. In 2020, ~19 million deaths were due to cardiovascular death globally, which represented an increase of 18.7% from 2010[1]. The total costs of cardiovascular disease (CVD) were estimated to be $378.0 billion in 2017-2018, as compared to $103.5 billion from 1996 to 1997. Furthermore, CVD was responsible for 12% of total US health expenditures from 2017-2018.

Cardiovascular disease is broadly defined as heart disease, including coronary heart disease, stroke, hypertensive disease, and other circulatory conditions, including the peripheral circulatory system. The dominant cause of cardiovascular disease is atherosclerosis. Atherosclerosis is an immunoinflammatory disease of medium and large arteries that is generally fueled by lipids, leading to the thickening or hardening of the arteries, with the most devastating consequences being stroke or myocardial infarction (MI)[2]. While cardiovascular disease remains a leading cause of morbidity and mortality in the United States[1], major adverse cardiac events (MACE) from atherosclerotic cardiovascular disease (ASCVD) are often undetected until their clinical presentation.

1.2 Cardiovascular disease and cancer

Cancer patients have an increased mortality risk from cardiovascular disease[3], yet only half of the high-risk patients are medically optimized with guideline-directed medical therapy[4]. While this increased risk might be explained

by common risk factors, it is important to note a high correlation between atherosclerosis and cancer even after adjusting for possible confounders such as gender, race, or smoking. A recent study using participants from the Multi-Ethnic Study of Atherosclerosis demonstrated higher atherosclerotic burden was associated with an increased risk of cancer (HR: 1.53, CI: 1.21-2.39)[5]. In patients with atherosclerosis, prostate cancer was the most common cancer diagnosed (21% of cases), followed by lung cancer (14% of cases).

Cancer patients have readily available imaging information, staging, and therapy-planning CT scans, which provide a unique opportunity to intervene based on cardiovascular risk. In this study, we are specifically interested in identifying cardiovascular risk for cancer patients in the National Lung Screening Trial (NLST). Lung cancer is the leading cause of cancer death in the United States and more than 80% of these deaths may be attributed to tobacco exposure[6]. Of note, smoking is also a well-known cardiovascular risk factor, contributing to the increased cardiovascular risk in lung cancer patients[7].

The NLST enrolled 53,454 people at high risk for lung cancer from August 2002 to April 2004[8]. Participants were randomized to two groups to either undergo three annual screenings with low-dose helical CT or single-view chest x-ray. Participants were followed through December 31, 2009, and data was collected on lung cancer diagnoses, cancer deaths, and deaths from other causes such as cardiovascular deaths. The study found that screening with the use of a low-dose CT scan for screening was significantly better than x-ray in improving all-cause

mortality[8]. More than 70,000 CT scans from the NLST are now publicly available for research.

1.3 Cardiovascular risk-stratification

Studies examining the prevention of CVD often focus on atherosclerotic cardiovascular disease risk or major adverse cardiac events. Clinical ASCVD is defined as a single established clinical event in any arterial bed, such as unstable angina requiring revascularization; while sub-clinical ASCVD is asymptomatic coronary or peripheral artery disease traditionally defined by imaging or risk factors such as cholesterol[9]. Oftentimes studies aimed at the prevention of cardiovascular events focus on a 10-year risk percentage, which is defined as the risk of having a cardiovascular event, such as a MI or stroke in 10 years. Other studies have now commonly used MACE as a composite endpoint in randomized control trials. A review identified peer-reviewed articles published from 2010-2020 and found that the most common definition for MACE included acute myocardial infarction and stroke (15.5%), and other studies often included all-cause death or cardiovascular death in the MACE outcome[10].

Many prospective cohort studies have tried to identify risk factors for ASCVD development using demographic and lab data starting with the Framingham Heart Study in 1948[11]. The study led to the identification of three risk factors associated with the development of coronary heart disease (CHD): elevated serum cholesterol levels, hypertension, and the presence of left ventricular hypertrophy on electrocardiogram[12]. By the 1970s, the risk factors thought to be most important expanded to include physical inactivity, diabetes,

glucose intolerance, serum triglycerides, high LDL cholesterol, low HDL cholesterol, atrial fibrillation, LVH, and heart failure[13]. These studies eventually led to the development of the new pooled ASCVD risk equations[14], which included cohorts of African-American and white participants with 12 years of follow-up data. This led to the creation of a 10-year risk score of ASCVD for African American and white men and women who were 40-79 years of age. This score uses age, sex, blood pressure, cholesterol, diabetes, smoking, and current hypertension treatment to sort patients into risk categories from low to very high-risk guide the medical management of ASCVD. This score was then validated in the community and found to overestimate 5-year cardiovascular risk[15]. While ASCVD is currently used to guide medical management, it may not be accurate for patients that are not African-American or white[15,16], or without comorbidities such as diabetes[15]. ASCD risk was also found to not account for socioeconomic status[17].

Genetic data has also been used for cardiovascular risk calculations. Individuals with familial hypercholesteremia have extreme ASCVD risk. Familial hypercholesteremia is due to mutations in the LDL receptor gene, apolipoprotein B gene, or PCSK9 genes causing elevated levels of LDL cholesterol usually >190 mg/dL. Homozygotes, about 1-6 per million people, usually begin to develop angiographically detectable coronary artery disease (CAD) by age 13[18,19]. Heterozygotes, about 1 in 500 people, demonstrate angiographically detectable disease by age 17-25[18,19]. Genetic FH is associated with a 22-fold increased risk for CAD[20]. Furthermore, genome-wide association studies have identified several loci associated with plaque development, as well as suggested a causal role

between genetic predisposition and ASCVD[21,22]. However, screening based on genetics does not seem to be effective for community screening since 7% of adults in the USA have severe hypercholesterolemia and the majority of these adults do not have a genetic cause[20].

Imaging is another modality that has been used to risk-stratify patients, including brachial artery reactivity testing, vascular compliance testing, aortic and carotid MRI, and US, and CT scans. B-mode carotid ultrasound is used to measure carotid intima-media thickness (IMT), which has been widely used as a surrogate for the presence of atherosclerosis[23]. A meta-analysis demonstrated that the relative risk of stroke was 1.32 per 1-standard deviation difference in carotid artery IMT, and similarly, the relative risk of myocardial infarction was 1.26 per 1-standard deviation difference in carotid artery IMT[24]. This led to the inclusion of carotid IMT in the assessment of the cardiovascular risk of CVD in 2010[25]. However, studies demonstrated that although the carotid IMT was associated with future cardiovascular events, carotid IMT did not outperform traditional risk factor calculations and did not significantly change risk calculation when added[23]. In 2013, routine IMT was no longer recommended for clinical risk assessment[14]. A few studies have shown the presence of plaques on ultrasonography might be more predictive for future cardiovascular events[26,27], however, its routine use remains controversial[23]. Perhaps the most well-defined risk score is CT-derived coronary artery calcium (CAC) scores, which quantify the burden of atherosclerotic plaques in coronary vessels[28] through coronary artery calcium (CAC) scores.

1.4 Agatston Score

Current guidelines primarily use CAC for cardiovascular risk assessment and suggest its use to determine if a borderline-risk patient should start statin therapy[29,30]. In the 1990s, Agatston et. al. demonstrated that coronary artery calcium burden could be calculated using CT scans[31]. The Agatston score (AS) was calculated as the sum of coronary artery calcium plaque volumes, assigning a weighted calcium peak CT attenuation factor (Hounsfield units) to the highest density of calcification in that plaque[31].

A variety of studies have since demonstrated that an increasing Agatston score is highly predictive of MACE and outperforms other risk factors in asymptomatic populations[32–34]. The NHLBI's MESA demonstrated that CAC was predictive of CHD events across ethnicities and ages[33,35]. Specifically, participants with a CAC score of 1-100 had a four times higher risk of coronary events, and those with a score > 100 were seven to ten times more likely to have a coronary event[33]. When analyzing the MESA results over a 12 year follow-up, CAC showed the best performance from 735 variables from non-invasive tests, questionnaires, and biomarker panels[34]. The Dallas Heart Study, a prospective study, also demonstrated the addition of CAC to the traditional risk factor model resulted in the increased predictive ability of coronary events[36]. In another landmark study, the Heinz Nixdorf Recall Study demonstrated CAC predicted stroke independently of traditional risk factors in low and intermediate risk populations[37]. A meta-analysis pooling data across 3 studies showed that the relative risk of stroke with CAC > 0 was 2.95[38].

The calcium score has also been studied in cancer patients. NLST participants[8] were evaluated by three methods of calcium scoring: visual assessment, vessel-specific scoring, and the Agatston score[39]. For overall visual assessment, radiologists described the calcium burden as none, mild, moderate, or heavy upon initial visual assessment. For vessel-specific scoring, radiologists calculated a calcium score for each of the 10 coronary vessel segments and summed these scores. Using multivariate cox regression for time to CHD death, they found that all three methods were comparable for risk assessment, with high calcium scores in each category corresponding to a hazard ratio of greater than 6 for death from CHD[39].

More recently, Zeleznik et. al. applied deep learning methods to automate calcium score calculations using the Agatston method[40]. In this proof-of-concept study, they evaluated over 20,00 individuals from the Framingham Heart Study[11] and NLST, using ASCVD mortality as the outcome. They developed a deep-learning-based coronary calcium measurement, which relied on heart segmentation of the CT scans followed by volumetric implementation of AS[31].

In another study, Atkins et. al. demonstrated that deep learning-based CAC was significantly predictive of MACE when applied to treatment planning CT scans for lung cancer patients. Another study in breast cancer patients demonstrated automated CAC scoring was strongly associated with coronary artery disease[41].

1.5 Limitations to the Agatston Score

While the current AS is predictive of MACE, the AS does not provide information on the location or distribution of the plaque. Multiple studies have

looked at CAC distribution amongst coronary vessels by manual segmentation. These studies demonstrated the number of coronary vessels with calcium burden and the presence of CAC in the proximal dominant coronary artery independently predict MACE and significantly improved the discriminatory capacity of AS to predict MACE[42,43].

Another study extracted previously defined radiomic features from segmented CAC and generated a radiomic score using 20 of these key features. They demonstrated that when the radiomic score was used with AS, it significantly improved MACE prediction[44]. Machine learning has been applied to CT angiography data to identify radiomic feature importance but has not yet been applied for feature extraction[45,46].

Another limitation of the current AS is that it measures calcium scores as a product of volume and density. However, previous studies have demonstrated that while the volume of calcium burden does correlate with increased risk of major cardiovascular events, density might be inversely correlated and protective of adverse events[47,48].

Given recent literature demonstrating the utility of applying deep learning methods to automate calcium scoring and the known limitations of the Agatston scoring method, we aim to use deep learning to develop a more robust imaging-based predictive biomarker of MACE. We hope this new biomarker would increase predictive power, enabling clinicians to appropriately risk-stratify and provide preventive care for patients.

1.6 Deep learning and cardiology

Physicians are routinely required to integrate data from multiple different modalities including but not limited to imaging, lab findings, and exam findings to make an actionable clinical judgment. The field of artificial intelligence was founded to describe human intelligence so precisely that a machine would be able to mimic it. Deep learning is a type of machine learning that relies on multi-layered neural networks to integrate multiple data streams and infer intelligently from that data[49]. A neural network consists of many connected neurons or processors that produce a sequence of activations, either influenced by the environment or other processors. The task of a neural network is to identify the weights or states that make a neural network exhibit the desired behavior, such as identifying a tumor. Given the advancements in computation, deep learning could help augment clinical decision-making in the future, thereby improving clinical outcomes[50].

Cardiovascular medicine is well suited to deep learning applications. For instance, clinical information is captured continuously through electrocardiograms or wearable devices. Deep learning has also been used to classify images from arrhythmia detection to image segmentation for cardiac MRIs, echocardiography, and CT scans[50]. Deep learning can further be subdivided into supervised and unsupervised learning. Supervised learning requires the input data to be labeled, i.e., this structure is the left atrium, providing direct feedback for training. Unsupervised learning is unbiased and asks for the machine to find hidden structures in the data, clustering all similar structures as left atriums. This approach allows for identifying novel patterns of features within the data[49].

1.7 Automated image segmentation

Deep learning approaches often involve isolating a region of interest, such as the total heart. Segmentation may be manual or deep learning based. Manual segmentations are time-consuming and subject to observer bias[51]. Deep-learning methods often require a trial and error process, adapting and training a neural network to the specific task, making decisions about data augmentation or post-processing[52]. Recently, nnU-Net was developed to overcome these challenges. The segmentation method nnU-Net performs automated configuration of pre-processing, network architecture, training, and post-processing for segmentation-based tasks[53]. This algorithm was then validated on 23 public datasets and outperformed other existing approaches. An ongoing study developed a hybrid algorithm[54] that applied the nnU-Net model to segment the whole heart and then applied a multi-atlas-based mapping of cardiac substructures[55]. Atlas-based segmentation relies on mapping labeled image structures onto the target image. Previously, a study looking at participants in the National Lung Screening Trial used multi-atlas-based mapping of coronary vessels and valves to calculate Agatston scores spatially using a slab approach[56]. They found that while slab-based Agatston scores outperform patient characteristic-based risk calculators, the total Agatston score performed similarly to the slab-based Agatston score. This may be a result of approximations used to map coronary features onto the heart.

While multi-atlas approaches have been able to accurately segment the whole heart and its chambers, there has not been much success with smaller structures such as coronary vessels[57,58]. The multi-atlas-based mapping of cardiac

substructures resulted in the mapping of 17 cardiac substructures on the dataset, including chambers and coronary vessels[55]. The accuracy of these automated segmentations was calculated using Dice scores, where 1 corresponds to a pixel-perfect match between the model output and the manually segmented image. It was found that whole heart segmentation had a Dice coefficient of 0.94, suggesting high accuracy, and Dice coefficients for the four chambers averaged around 0.85. However, the Dice scores for the coronary arteries were as low as 0.03, suggesting this multi-atlas approach might not be appropriate for smaller structures such as vessels[55].

1.8 Radiomic Features

Radiomics is a field focused on extracting quantitative data from medical images[59]. Extraction of these radiomic features often includes segmenting the region of interest, processing the image segment, and radiomic feature calculation. Image processing attempts to standardize uptake values as well as denoise the data. Radiomics relies on using advanced mathematical analysis of the spatial distribution of signal intensities and relationships between pixels to describe image features[51]. These features can further be divided into classes: statistical (histogram-based and texture-based, model-based, transform-based and shape-based)[60]. Transform-based features analyze the image matrix in a different space, following a Fourier, Gabor, or wavelet transformation. Due to different software implementations of these steps, radiomic features from published studies have been hard to reproduce[61–63]. Following these studies, the Image Biomarker Standardization Initiative set out to standardize a set of 169 radiomic features[60].

Importantly, the initiative used the most common image processing steps and provided reference values using these methods. These guidelines allow for standardized feature calculations to capture tissue properties such as shape or heterogeneity. These radiomic features then allow for high-dimensional information to be captured about medical images and has applications for biomarker development for MACE.

Chapter 2: Statement of purpose

Cancer patients provide a unique population for which there is increased risk of MACE and available CT data. We aim to use the NLST cohort as our training and validation cohort. The NLST lung screening trial[8] has low-dose non-contrast CT scans for 26,722 participants.

Aim 1: Previous studies have demonstrated successful segmentation of large cardiac substructures, yet there has been less success with segmenting smaller substructures. We will use deep learning methods to segment both large and small cardiac structures on CT scans.

- *Aim 1A:* We will apply the pre-trained nnU-Net model to low-dose non-gated CT scans from the NLST dataset. We will then validate if we correctly segmented large and small cardiac features.
- *Aim 1B:* For cardiac substructures that we were not able to segment correctly, we will train a deep learning model using nnU-Net on manually segmented features on the NLST CT scans.

Aim 2: We will seek to better understand if these automated cardiac segments are predictive of MACE as they provide a low-cost non-invasive way to predict cardiovascular risk.

- *Aim 2A:* Previous studies largely rely on image features contained within the coronary vasculature, and we wish to identify if there are any other features in CT scans including heart volume that may be predictive of MACE.

Aim 3: Because of the limitations of the Agatston score, we wish to better understand unique image features that are predictive of MACE.

- *Aim 3A:* We will extract radiomic features from our automatically segmented coronary structures to give us a high-dimensional dataset. We will use multivariate logistic regression to train a predictive model for MACE using these radiomic features.
- *Aim 3B:* We will compare our imaging-based biomarker to the Agatston score for its ability to predict MACE.

Chapter 3: Materials and Methods

3.1 Dataset

The National Lung Screening Trial (NLST)[8] enrolled 53,454 participants from August 2022 to April 2004 at 33 U.S. medical centers. The NLST was approved by the institutional review board at each of the 33 sites. Before randomization, participants provided written informed consent.

Participants were eligible for the trial if they were between 55 and 74 years of age, had a smoking history of at least 30 pack-years, and had quit within the previous 15 years. Demographic data were collected at the time of randomization through questionnaires. This demographic data included information on age, gender, ethnicity, education level, and smoking behavior[8]. Participants were excluded at the time of randomization if they had a diagnosis of lung cancer, had a chest CT in the previous 18 months, or had hemoptysis or unexplained weight loss of more than 15 lbs. Participants were then randomized to screenings by low-dose CT or x-ray.

For the 26,722 participants that underwent CT screening, they were invited to undergo three screenings (T0, T1, T2) at 1-year intervals. All low-dose CT scans were acquired with multidetector scanners with a minimum of 4 channels[64]. The average effective dose of the screening CT was selected to be 1.5 mSv, lower than the average of 8 mSv used for diagnostic chest CT. All of the scanning machines met the technical standards set by the American College of Radiology[64].

The NLST CT data is publicly available and was obtained with approval from the National Cancer Institute (NCI) through the NCI's Cancer Data Access System

(CDAS) on June 2, 2022, with access until June 2, 2027. When we obtained access to the NLST CT scans, the CT data had been previously split into training and testing datasets. For our initial data analysis, we selected participants with screen-detected lung cancer in the training dataset. Because of the strong association between cancer and poor cardiovascular outcomes, we sought to identify image-based biomarkers within this dataset. This results in 307 participants being selected for our initial analysis as shown in Figure 1. We further enhanced this dataset for preliminary analysis by selecting participants with deaths from cardiovascular causes (see 3.3 NLST Outcomes): 15 participants with cardiac deaths and 79 participants with no cardiac deaths during the follow-up period.

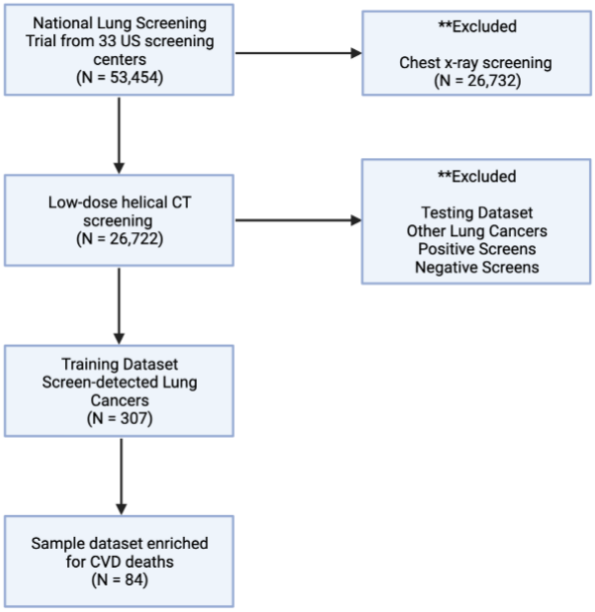


Figure 1. National Lung Screening Trial Flowchart. This flowchart details the total number of participants in the NLST trial and the selection criteria for our sample dataset.

3.2 Cardiac Segmentation

For the automated cardiac segmentation on CT scans, we implemented a model based on nnU-Net architecture. nnU-Net is a deep-learning segmentation

method that automates configuration, pre-processing, network architecture selection, training, and post-processing[53]. Given a new segmentation-based task, the algorithm extracts dataset properties to infer rule-based parameters for segmentation and combines these with fixed parameters. The nnU-Net then determines the optimal ensemble of these models. Finnegan et. al. developed a hybrid algorithm that implemented this nnU-Net model to perform heart segmentations on CT scans[54]. Following the identification of the heart structure, multi-atlas cardiac segmentation was implemented[55] to map 17 cardiac sub-structures including large and small coronary arteries, ventricles, atriums, heart valves, and conduction nodes. This code was available through the PlatiPy library, a processing library and toolkit for medical imaging in Python[65].

NLST CT scans were downloaded from the Cancer Data Access System and uploaded onto a secure Amazon Web Services S3 bucket. NLST CT scans were selected for the subset of 84 participants. Each patient had multiple scans each year following the time of randomization. To ensure accurate follow-up times, the T0 scan was used which was the CT scan the year the patient was randomized during the NLST trial. Scans were further selected to be axial images and have soft-tissue reconstruction.

The selected CT scans were processed and segmented on a secure Amazon Web Services EC2 instance (4 vCPUs, Deep Learning AMI GPU PyTorch 1.12.0 (Amazon Linux 2)). Using code from the PlatiPy library, selected scans were converted from DICOM to NiFTI formats, and the automated segmentation algorithm was applied by running a Python script on the EC2 instance.

Once we validated that the code ran successfully, we worked with the research team at Visage Imaging© to upload our images on their secure research API as well as integrate the segmentation code from PlatiPy. Khaled Bousabarah helped us to implement this code. We then batch-processed 84 scans to generate 18 cardiac substructures that we could visualize as structures within the Visage API.

We first focused on large cardiac substructures and coronary arteries. To test the accuracy of these segments, Arman Avesta, a trained radiologist, helped check the accuracy of select substructures for every 10th CT scan. For structures that had low accuracy, manual segmentation was performed. This was done by Aishwarya Nene and Crystal Cheung, following training provided by Arman Avesta. Manual segmentations were reviewed by Sanjay Aneja, Assistant Professor of Therapeutic Radiology. These structures will then be used to train a new nnU-Net model to correctly identify these cardiac substructures on our NLST CT dataset.

3.3 NLST Outcomes

NLST data has a follow-up time until 12/31/09, hence mortality, survival, and incidence analyses were all done by adjusting the follow-up time to this date for participants that were followed until a later timepoint. If a patient had not died, a fixed number of 58.2 days was subtracted from the follow-up time.

The NLST trial did not contain data for non-fatal MI or stroke, hence mortality from cardiovascular disease was used as the primary outcome for our study. Chiles et. al. previously used coronary heart disease death, which they defined by ICD codes I20-I25, I46, and I50; as well as all-cause mortality to assess

coronary artery calcification scoring methods in the NLST trial[39]. The World Health Organization defines mortality from cardiovascular diseases as including specifically diseases of the heart, essential hypertension and hypertensive renal disease, and cerebrovascular diseases[66]. However, the most inclusive definition restricts all deaths from cardiovascular disease as ICD codes I10-I79. We used these three different definitions of cardiac deaths (Table 1) across all NLST participants who received CT screening.

	Diseases	ICD Codes
Coronary heart disease	CHD	I20-I25, I46, I50
Specific CVD	CHD, diseases of the heart, Essential hypertension and hypertensive renal disease, Cerebrovascular diseases	I100-109, I11, I13, I20-51, I10, I12, I15, I60-169
All CVD	CHD, diseases of the heart, Essential hypertension and hypertensive renal disease, Cerebrovascular diseases, aneurysms, HHT, etc.	I10-I79

Table 1. Cardiovascular death definitions. Definitions for cardiovascular death are listed from most specific to most inclusive. The diseases associated with each label as well as their ICD-10 codes are listed.

For each definition of cardiovascular death, all participants in the NLST CT database were categorized by ICD-10 codes provided. About 19.5% of participants passed away during the follow-up period (N=26722). Coronary heart disease-related deaths were responsible for 14.4% of deaths, while deaths from any cardiovascular cause were responsible for 23.9% of deaths (Figure 2). To have discerning power in our model, we chose the definition that included deaths from

all cardiovascular causes. Furthermore, we selected cardiovascular death in our subset of 84 participants for which we initially trained our model.

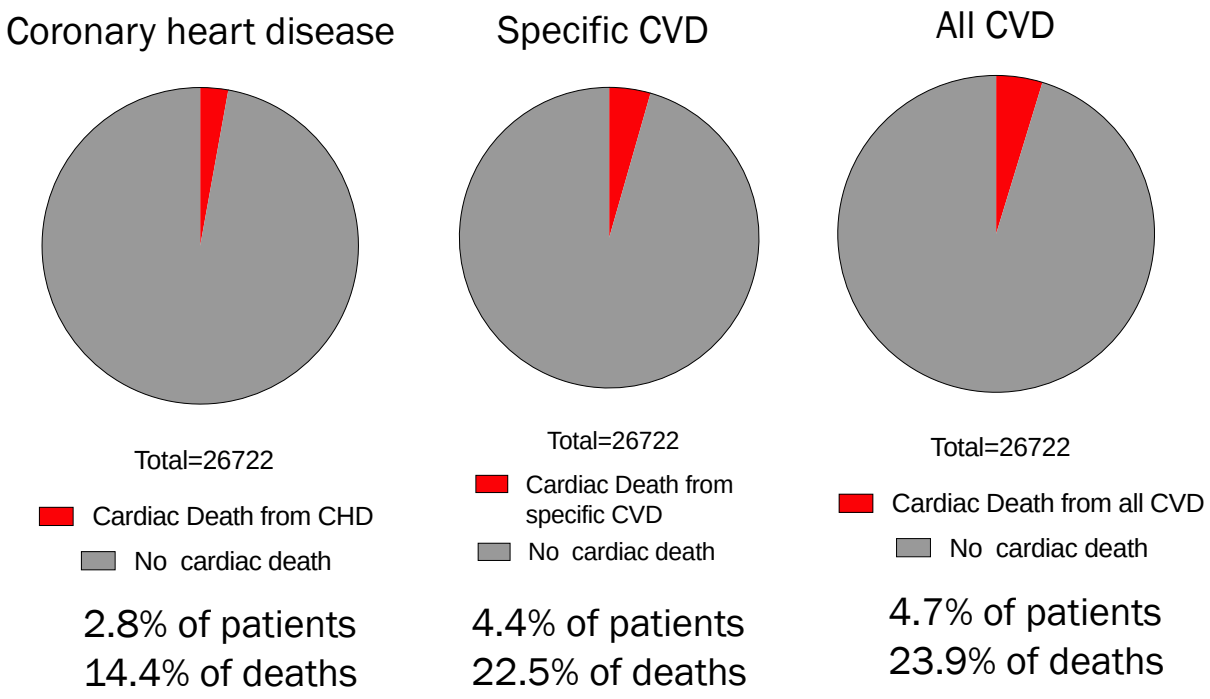


Figure 2. NLST cardiovascular outcomes for different cardiac diseases. Pie charts represent the fraction of participants that passed away due to a cardiac cause for each definition of cardiovascular death. The percentage of participants that passed away and the percentage of total deaths is listed below each figure.

3.4 Radiomic Feature Extraction

For radiomic feature extraction, we used *Pyradiomics*, a flexible open-source platform implemented in Python[67]. These features comply with the feature definitions described by the Imaging Biomarker Standardization Initiative[60]. Working with Visage Imaging© research team, we implemented the radiomic feature extraction code on the Visage API. We selected cardiac substructures of interest, extracting radiomic features from both the original image as well as the wavelet-filtered image. The wavelet filter yield 8 decompositions per level by applying a high or low pass filter across the three dimensions of the CT image[67].

This transformation allowed for a higher dimensional dataset for model training and development.

3.5 Feature Scaling and Selection

Before model training, it is necessary to appropriately scale the data and select features of high importance. Standardization of a dataset is commonly required for many machine learning techniques that perform assuming all features are centered around 0 and have a variance of that same order. Since radiomic features all have varying scales, we implemented a standard scaler from the scikit-learn library. This scaler removes the mean and scales each variable to unit variance. We first aggregated radiomic features from all selected cardiac substructures, scaled the dataset, and then selected features associated with cardiac death.

Following scaling, we used minimum redundancy maximum relevance (mRMR) feature selection to sort the different radiomic features. The mRMR algorithm was developed first using microarray gene expression data[68]. Commonly, filter methods for feature selection select the top X variables that are associated with the outcome. mRMR goes one step further to remove variables that are highly correlated amongst themselves. This allows for the selection of efficient features as well as broad features that are capturing a variety of different information about the dataset. Python was used to implement mRMR for selection on our dataset, for categorical feature selection of the categorical variable of cardiac death, as defined above.

3.5 Model Training

For our model development, we focused on developing a cardiac risk score that was based on CT-imaging based features. We focused on developing a score using three levels of information from the CT scan. Because the widely used Agatston score currently relies on calcifications from the coronary arteries, our first composite score relies on radiomic features from segmented coronary arteries alone. The other two composite scores we wished to discover were based on the whole heart and its major substructures, and a score based on imaging features present in the whole chest CT scan including the lungs.

For our model design, we used CT scans from the NLST with our primary outcome as death from all cardiovascular diseases. We used both automated and manual segmentations to identify cardiac substructures including the cardiac chambers and small coronary arteries. Following segmentation, radiomic features were extracted as detailed above for the selected cardiac substructures. Features were scaled and selected using mRMR with cardiac death as the outcome.

We used the top 40 radiomic features to train our logistic regression model. The dataset of 84 participants was split into training (N = 63) and validation cohorts (N = 21). For the training dataset, we first used univariate selection with a significance level of $P \leq 0.15$ to select variables that were significantly associated with our outcome variable. We then used backward elimination on a logistic regression model to iteratively remove features until all remaining features had a significance level of $P \leq 0.05$. We then validated the model on the validation cohort and reported the AUC.

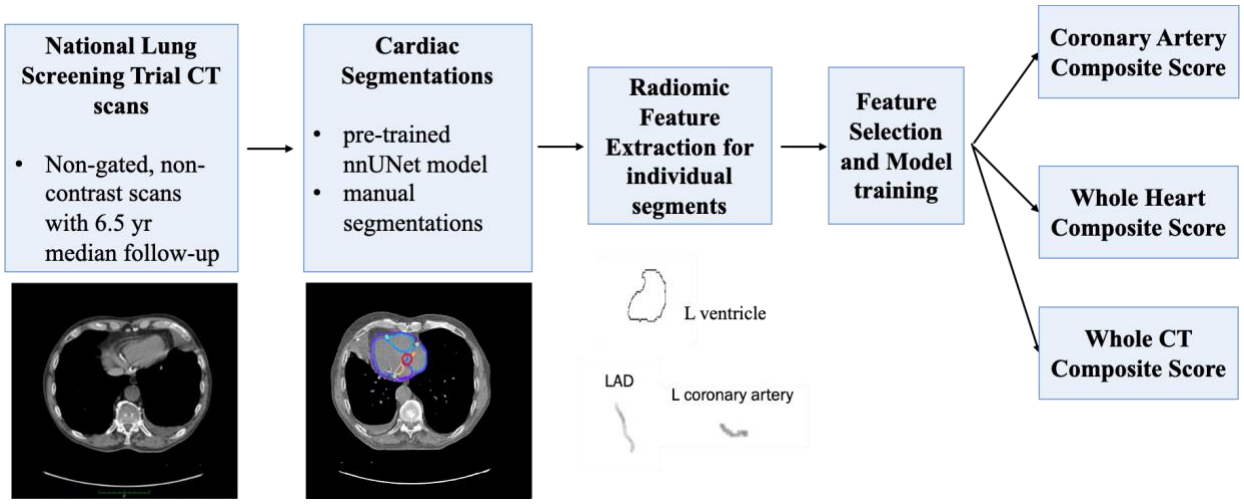


Figure 3. Experimental design flowchart. Starting with whole CT scans, we first segmented our images to areas of interest. We extracted radiomic features from these structures and trained our model to develop a composite score based on coronary arteries, whole heart, and whole CT.

3.6 Statistical Analysis

Prism and Matplotlib were used for the visualization of the data. To compare two groups of patient characteristics, unpaired parametric t-tests were performed to identify statistically significant ($P < 0.05$) differences.

For univariate feature selection, we ran univariate linear regression tests using the scikit-learn library, returning F-statistics and p-values. The p-values were used to select for features that would be used for the model training. Logistic regression was run using the scikit-learn library and p-values were reported for the coefficient of each variable, where the null hypothesis was that the coefficient is equal to zero (no effect).

Receiver operating characteristic curves were used to report the performance of our classification model. The area under the curve is reported to represent the probability the model score will be able to predict future cardiac death correctly.

3.7 Student Contributions

The application to obtain the NLST CT scans was written and submitted by the student. The student was responsible for designing the experiment, writing, and implementing the code, performing the data analysis, and creating the figures and writing the thesis under the supervision of Sanjay Aneja and Harlan Krumholz.

The student received help from the Visage research team to upload imaging data, implementing the pre-trained cardiac segmentation code into the Visage API, and batch processing radiomic extraction data. Arman Avesta assisted in validating automated segmentations and training Crystal Cheung and the student to perform manual segmentations of cardiac substructures. Sanjay Aneja validated the manual segmentations. Arman Avesta assisted in exporting the manual segmentations from the Visage API and running the nnU-Net model to retrain the coronary artery segmentations.

Chapter 4: Results

4.1 Patient Selection

Patient demographics for all NLST CT scan participants and our selected sample dataset are provided in Table 2. Participants in our sample dataset were significantly older (P <0.001) by 2 years compared to the total dataset, smoked 12 more pack-years (P <0.001), had shorter follow-up times (P < 0.001). This is likely due to our selection criteria that selected for participants with screen-detected lung cancer and a high proportion of cardiovascular deaths (17.9%). We saw a similar proportion of male and female participants in both datasets.

Characteristic	**All Participants (N = 26722)**	**Sample Dataset (N = 84)**
	***Mean* ± SD**	
Age (yrs)	61.4 ± 5.0	63.9 ± 5.3
Smoking history (Pack-years)	56.0 ± 24.0	68.2 ± 30
Follow-up time (yrs)	6.3 ± 1.1	5.2 ± 2.2
Sex	*no. (%)*	
Male	15769 (59.0)	54 (64.3)
Female	10953 (41.0)	30 (35.7)
CVD Death		
Yes	1266 (4.7)	15 (17.9)
No	25456 (95.3)	79 (94.0)

Table 2. National Lung Screening Trial Patient Characteristics. Selected self-reported patient characteristics are shown for NLST participants who were randomized to CT screening (N=26722) and our cardiovascular death enhanced subset (N=84).

4.2 Cardiac Segmentation

We implemented a hybrid algorithm[54] onto the NLST CT scans. The algorithm uses a nnU-Net model to automate heart segmentation and uses multi-atlas segmentation to map the other 17 cardiac substructures. The resulting axial, coronal and sagittal views are shown in Figure 4 below, with masks for the large chambers shown individually.

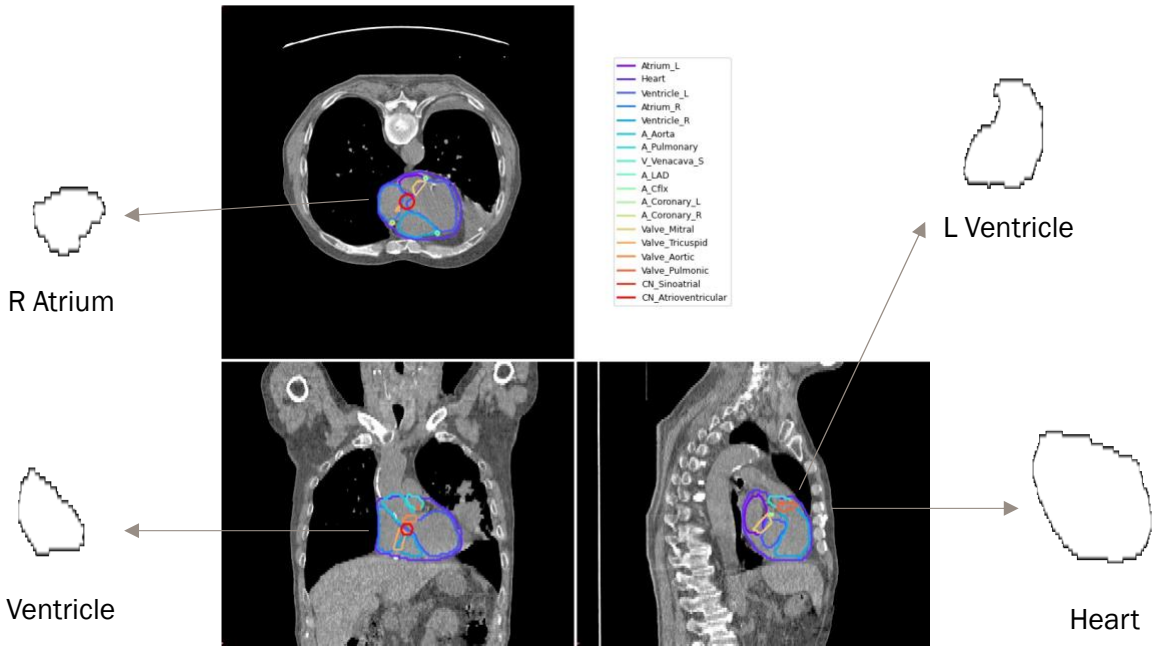


Figure 4. Pre-trained nnU-Net cardiac segmentation model identifies large cardiac substructures. The cardiac pre-trained model segments the heart and 17 other cardiac substructures. The cardiac chambers are highlighted in this image.

To validate if these automated segmentations were correct, we implemented this code in Visage and visualized it on the Visage API. A sample CT scan with cardiac segmentation in Visage is shown in Figure 5. We ran the segmentation algorithm on all 84 participants selected for the study.

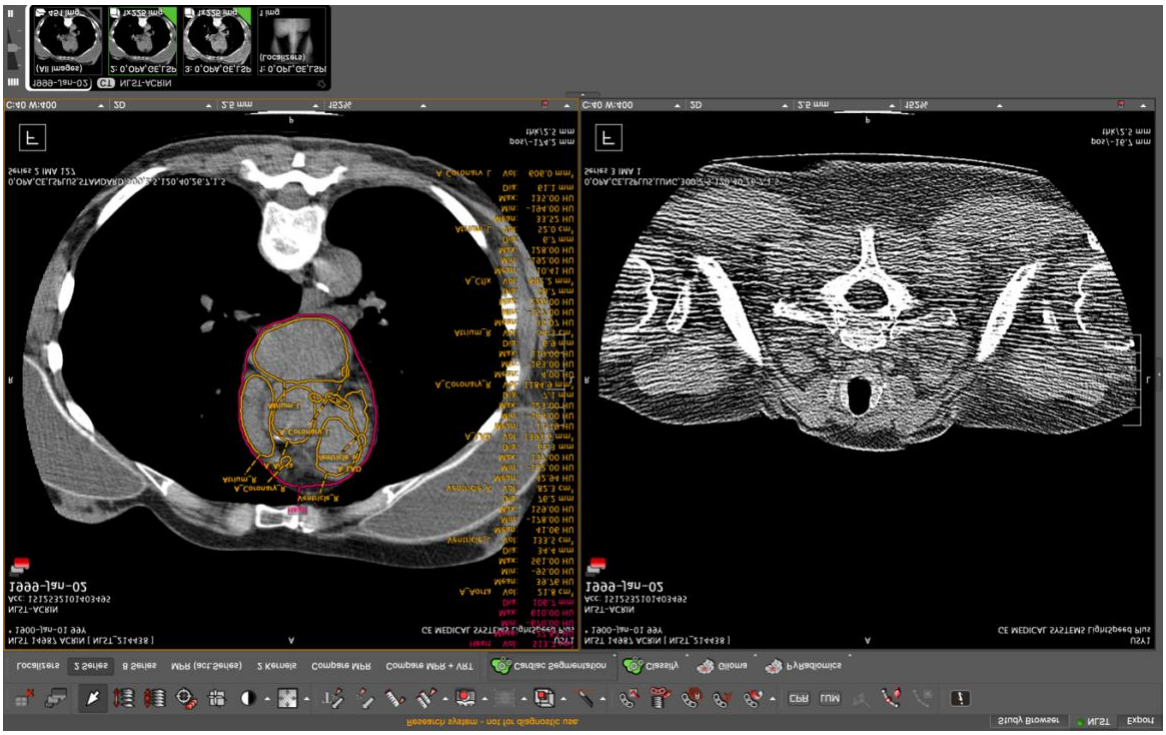


Figure 5. Segmentations visualized on Visage API. Cardiac Segmentation for CT scans on Visage API. The cardiac segmentation button applies the hybrid cardiac segmentation model to automatically segment 18 cardiac structures.

Visualizing the cardiac segmentations on the Visage API, we had a trained radiologist review the segmentations for the total heart, atriums, ventricles, and small coronary arteries. For every 10th sample, starting with sample 1, segments were validated for accuracy. Total heart segmentation was accurate for all 9 samples selected, and large structures such as the right atrium, right ventricle, and left ventricle had >75% accuracy. However, the left atrium was only accurate on 2/9 scans and the coronary arteries were never correctly identified in the sampled CT scans. Results for each sample are shown in Table 3. Figure 6 shows how the coronary arteries and left atrium were erroneously segmented.

Every 10th scan	Total Heart	R Atrium	L Atrium	L Ventricle	R Ventricle	R Main	L Main	LCflx	LAD
Sample 1	y	y	n	y	y	n	n	n	n
Sample 2	y	y	n	n	y	n	n	n	n
Sample 3	y	y	y	y	y	n	n	n	n
Sample 4	y	y	n	y	y	n	n	n	n
Sample 5	y	y	n	y	y	n	n	n	n
Sample 6	y	y	n	y	y	n	n	n	n
Sample 7	y	n	n	n	y	n	n	n	n
Sample 8	y	y	y	y	y	n	n	n	n
Sample 9	y	y	n	y	n	n	n	n	n

Table 3. Validating segmentations on Visage API. For each scan, a radiologist reviewed each segmentation for accuracy and determined if the segment was accurate (y) or not accurate (n).

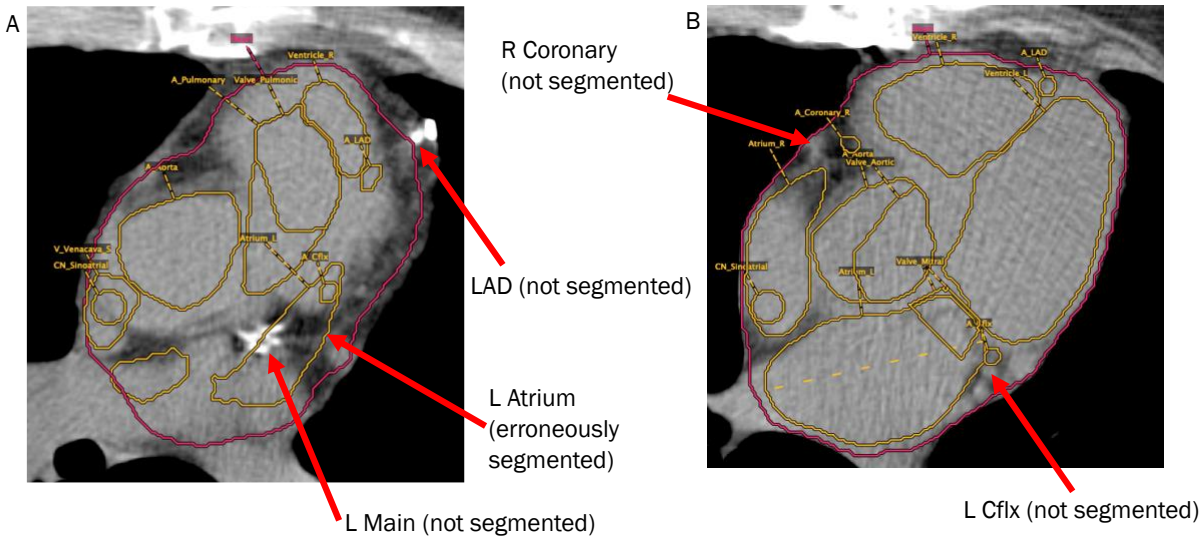


Figure 6. Pre-trained nnU-Net erroneously segments coronary arteries and L Atrium. (A) Calcifications appear as bright signals in the coronary arteries and are not correctly segmented as indicated by the lack of a yellow structure label. The left atrium, yellow label, is erroneously segmented and does not identify the correct dimensions or location. (B) The right coronary artery and left circumflex appear as round vessels on the CT scan, indicated by the red arrows, and are incorrectly identified by the segmentation algorithm, yellow labels, which are offset for both vessels.

For the incorrectly segmented coronary arteries and left atrium, we manually segmented 80 CT scans. Sample manual segmentations are shown for coronary arteries in Figure 7. These manual segmentations were validated by a radiologist. Following manual segmentation, we are currently in the process of re-

training the nnU-Net model to correctly identify the four coronary arteries and the left atrium.

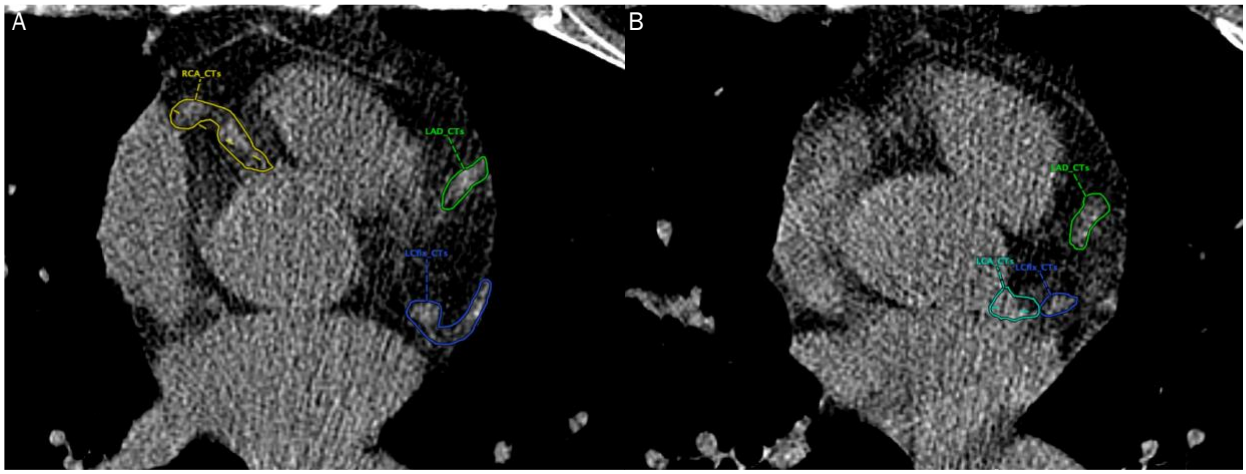


Figure 7. Coronary arteries are manually segmented on Visage API. (A) The right coronary, labeled RCA_CTs, is marked in yellow. The left anterior descending artery, LAD_CTs, is marked in green and is anterior to the blue-labeled left circumflex, LCflx_CTs. (B) The LAD and LCflx are seen branching off the left main artery, LCA_CTs, labeled in turquoise.

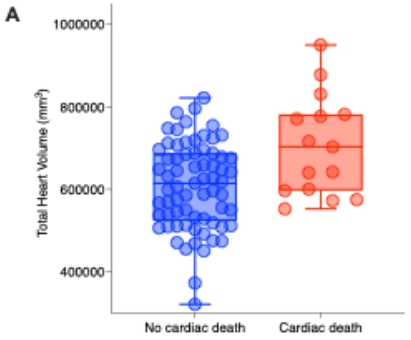
4.3 Cardiac substructure characteristics

Given the automated segmentations successfully identified the total heart and both ventricles, we aimed to understand if these characteristics of these features could predict MACE. Since left ventricular hypertrophy is a known risk factor of MACE[12], we evaluated whether ventricular or total heart volume was predictive of MACE. Volume was calculated using the mesh volume generated by *PyRadiomics*. Because these CT scans were non-gated, we wanted to ensure the volumes calculated for the heart chambers were within range. We compared the left ventricle volumes to reference ranges from cardiovascular magnetic resonance left ventricular end-diastolic volume (LVEDV). For NLST CT images, the left ventricle volume was 164.8 ± 29.29 mL for men and 140.5 ± 34.6 for women. Reference ranges for LVEDV were 155 ± 30 for men and 112 ± 21 for women[69]. Because LVEDV calculations did not include the ventricular wall, measurements

in the NLST CT scans were larger as they did include the ventricular wall and cavity. For both CT left ventricular volume and LVEDV, volume measurements were greater for men than women.

We further asked if automated volumetric measurements from non-gated CT scans were predictive of MACE. For this, we categorized participants according to whether they had a cardiac death during the NLST follow-up period. Total heart volume was significantly larger in the participants who had a cardiac death ($P = 0.002$, Figure 8A). Ventricular volumes also showed similar differences between participants who experienced a cardiac death and not. Left ventricular and right ventricular volumes were significantly increased for participants with cardiac death ($P = 0.001$, $P < 0.001$, Figure 8B).

The difference between total heart volume across cardiac death status becomes more apparent when looking at participants in quartiles. For the upper quartile of total heart volumes, 1 out of 3 participants had a cardiac death during the follow-up time. In contrast for the lowest quartile of total heart volumes, no participants had a cardiac death (Figure 8C). A Kaplan-Meier survival curve is shown in Figure 9 for participants with the upper quartile and lower quartile of total heart volume. The hazard ratio of having an upper quartile heart volume compared to a lower quartile heart volume was 6.68 (95% CI: 1.42 to 31.6).



C

Total Heart Volume Quartile	No cardiac death	Cardiac death
Lower Quartile	100%	0%
Upper Quartile	66.66%	33.33%

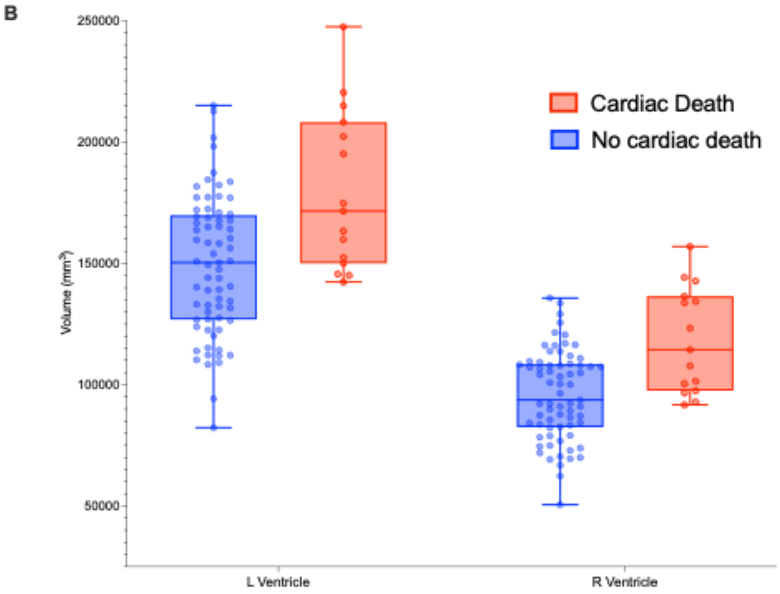


Figure 8. Heart volumes are increased in participants who had a cardiovascular death. (A) Participants who had a cardiac death had significantly increased total heart volume. (B) Ventricle volumes are larger for participants with cardiac death. (C) In the lower quartile of heart volume, participants did not have a cardiac death during the study follow-up period. However, 1/3 of participants with heart volume in the upper quartile had a cardiac death.

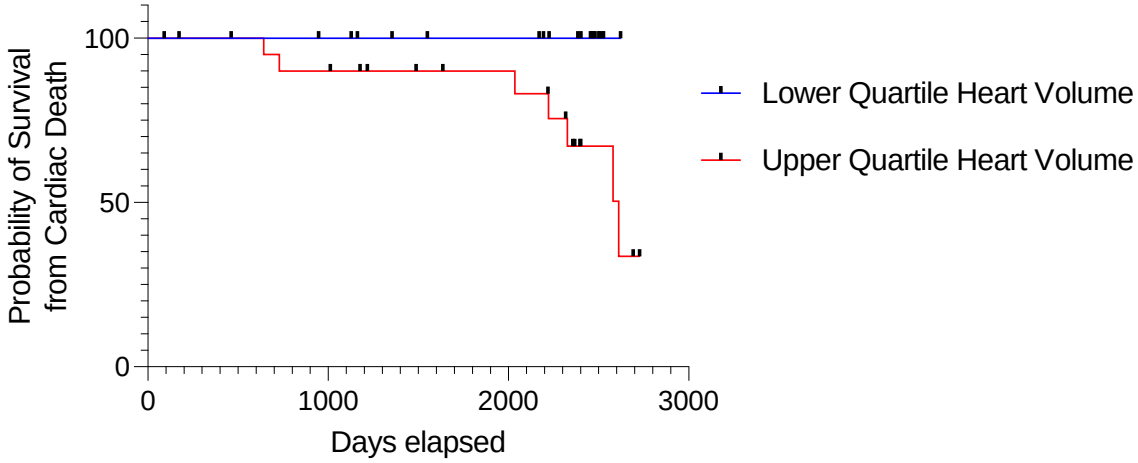


Figure 9. Total Heart is predictive of survival. Kaplan-Meier curve is shown for patients with a lower quartile of total heart volume and an upper quartile of total heart volume.

4.4 Model Training and Validation

We then extracted radiomic features from these automated segments to generate a high-dimensional dataset to train a predictive model for MACE. We successfully extracted 960 radiomic features for each automated segment. After aggregating and scaling the features for both ventricles and the total heart, we used a minimum redundancy maximum relevance feature selection algorithm to select for the top radiomic features. These features were selected for the ability to categorize participants who had a cardiac death and those who did not. The top ten features are shown in Figure 10.

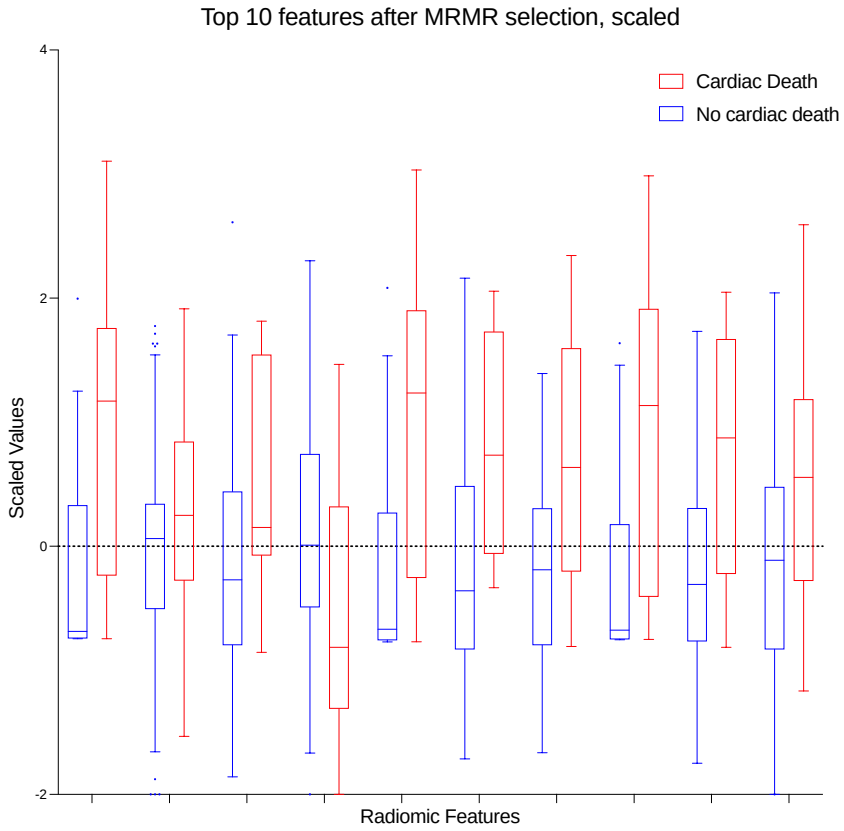


Figure 10. MRMR feature selection from the ventricle and total heart radiomic features. Features were selected for the ability to categorize cardiac death and have been scaled by subtracting the mean and scaling to unit variance.

After feature selection, we trained a model on the data from the 84 participants. We split participants into a training cohort of 75% of the participants (N = 63) and a validation cohort with 25% of the participants (N = 21). In the training cohort, the top 40 features were screened using univariate selection with $P \leq 0.15$. This resulted in 38 radiomic features being selected across all three cardiac segments. We then trained a logistic regression model with backward selection using these 38 radiomic features selecting for regression coefficients with $P \leq 0.05$. The resulting model had four radiomic features that significantly predicted cardiovascular death. These features included two radiomic features from the total heart segment and one from each ventricle. Three were measures of heterogeneity within the image. The resulting ROC curve is shown in Figure 11, and the model had an AUC of 0.62 in the validation cohort (N = 21). The accuracy of cardiac death predictions was 0.86.

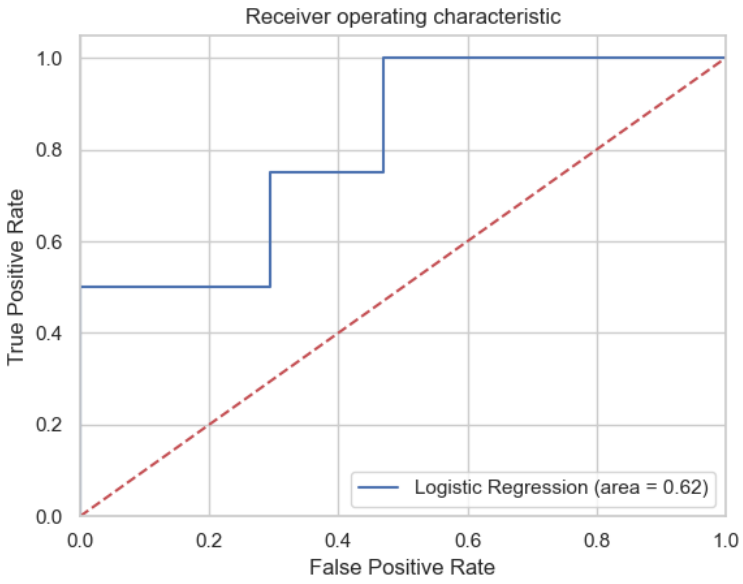


Figure 11. ROC Curve for radiomic feature model prediction in the validation cohort. This logistic model regression has been trained on a training cohort using radiomic features from the ventricles and total heart segments on CT scans.

Chapter 5: Discussion

5.1 Significance

We demonstrate that nnU-Net successfully segments the total heart. Multi-atlas-based segmentation onto this automated heart segmentation is also successful at identifying large cardiac substructures, such as the left and right ventricles. We also integrated this code into the Visage API, which is currently widely used across the Yale New Haven Health System, demonstrating the ability to automate segmentations on clinical CT scans in real-time. This automation of cardiac segmentation has tremendous potential to help expedite Agatston score calculations and allow for efficient imaging-based biomarker development.

The automated cardiac substructures had features that were predictive of MACE, offering a low-cost non-invasive predictive biomarker. Both ventricular and total heart volumes were significantly larger in patients that had a cardiac death during the NLST follow-up time. The hazard ratio for cardiac death was over 6 between patients who had the largest quartile of heart volume versus those in the smallest quartile. This hazard ratio is comparable to that found in previous studies looking at MACE predictive ability of calcium scores[39].

We were able to extract high-dimensional radiomic features from each of these cardiac substructures. This high-dimensional dataset allowed us to capture information about image texture, shape, and heterogeneity[60]. We found that heterogeneity-based radiomic features were predictive of MACE with an accuracy of 0.86.

Low-dose CT (LDCT) scans have been well-established as life-saving for cancer screening[8]. This study suggests that CT scans may also have utility in improving cardiovascular outcomes. This is a unique study that utilizes readily available imaging studies for cancer patients to ascertain cardiovascular risk. Our findings have the potential to be applied to other cancer patients who receive chest CT scans for screening or staging.

5.2 Challenges and Limitations

Our chosen dataset is artificially selected for increased cardiovascular death outcomes. As a result, the dataset is significantly different in comparison to the total National Lung Screening Trial study. Patients have more extensive smoking histories and more mortality due to lung cancer and cardiovascular disease. Since smoking and lung cancer are known risk factors for developing cardiovascular disease[30], our model's predictive ability may be limited to this very sick population. While these experimental design decisions were aimed at improving the model signal, it will be important to stress test this model against different variables, including smoking history, demographic data, and mortality from lung cancer. It will also be important to test this model in patients who do not develop cancer for external validity.

Consistent with previous studies[57,58], multi-atlas-based segmentations were not successful in identifying small structures such as coronary arteries. This imperfect contouring could be due to cardiac motion in the non-gated CT scans which were not isolated to systole or diastole, resulting in variability of vessel architecture. It will be important to see whether a nnU-Net approach trained on

manually segmented small structures will be able to overcome these limitations and correctly delineate these small structures.

Although the NLST CT scans were non-gated, automated segmentations of the cardiac substructures had calculated volumes that were appropriately within range compared to reference values from cardiac MRIs[69]. It will be important to validate these automated volumes with calculated volumes from manual segmentations on NLST CT scans.

While our composite predictive biomarker did show some predictive ability for MACE, our study was underpowered and only focused on large cardiac substructures. Our validation cohort consisted of 21 patients and only had 4 cardiac deaths, out of which 3 patients were correctly classified, leading to a poor AUC of 0.62. In future studies, it will be important to expand the number of patients. Furthermore, following the successful automation of smaller cardiac substructures, it will be important to include radiomic features from these segments in the model training and validation.

5.3 Future Directions

We will apply a nnU-Net approach to segment the left atrium and smaller coronary arteries successfully. To validate successful segmentation, for all automated structures, we will use a paired t-test to compare volume calculations for these automated segments with volume calculations from radiologist-validated manual segmentations. We will increase our sample size by applying this automated approach to all NLST participants.

Once we have successful segmentations for all cardiac substructures, we will work towards developing a three-tiered biomarker. The biomarker we have found in this study only utilizes large cardiac substructures. We wish to develop a biomarker based on the coronary arteries, the total heart and its substructures, and the whole CT scan. For the whole CT scan, we wish to use radiomic features from the segmentations of the thoracic aorta and lungs.

Once we have trained our radiomic feature-based biomarker on our NLST patients at each level, we wish to compare it against standard-of-care Agatston score calculations for our NLST participants. Previous literature demonstrates calcium scoring varies across demographic data (sex, race, age) and imaging parameters[9]. We hope to also stress test our automated biomarker against demographic and imaging parameters to understand how our model performs in different populations.

Our ultimate goal is to develop a low-cost non-invasive imaging-based biomarker to accurately predict a person's risk of developing MACE. We wish to integrate our biomarker with current imaging application software so clinicians will be able to calculate risk in real-time each time a patient receives a CT scan.

Dissemination

This study has currently not been made available to the public through oral presentations or peer-reviewed publications. The manuscript is being prepared and will be submitted for review at a peer-reviewed publication soon.

References

1. Virani SS, Alonso A, Benjamin EJ, et al. *Heart Disease and Stroke Statistics—2020 Update a Report from the American Heart Association*. Vol 141.; 2020. doi:10.1161/CIR.0000000000000757
2. Falk E. Pathogenesis of Atherosclerosis. *J Am Coll Cardiol*. 2006;47(8 SUPPL.):0-5. doi:10.1016/j.jacc.2005.09.068
3. Sturgeon KM, Deng L, Bluethmann SM, et al. A population-based study of cardiovascular disease mortality risk in US cancer patients. *Eur Heart J*. 2019;40(48):3889-3897. doi:10.1093/eurheartj/ehz766
4. Al-Kindi SG, Oliveira GH. Prevalence of Preexisting Cardiovascular Disease in Patients with Different Types of Cancer the Unmet Need for Onco-Cardiology. *Mayo Clin Proc*. 2016;91(1):81-83. doi:10.1016/j.mayocp.2015.09.009
5. Handy CE, Desai CS, Dardari ZA et al. The association of coronary artery calcium with noncardiovascular disease from the Multi-Ethnic Study of Atherosclerosis. *JACC Cardiovasc Imaging*. 2016;9(5):568-576. doi:10.1016/j.jcmg.2015.09.020
6. United States Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress A Report of the Surgeon General. *A Rep Surg Gen*. 2014:1081.
7. Critchley JA, Capewell S. Smoking cessation for the secondary prevention of coronary heart disease. *Cochrane Database Syst Rev*. 2012. doi:10.1002/14651858.cd003041.pub3
8. The National Lung Screening Trial Research Team *. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med*. 2011;365(5):395-409.
9. Rosenblit PD. Extreme Atherosclerotic Cardiovascular Disease (ASCVD) Risk Recognition. *Curr Diab Rep*. 2019;19(8). doi:10.1007/s11892-019-1178-6
10. Bosco E, Hsueh L, McConeghy KW, Gravenstein S, Saade E. Major adverse cardiovascular event definitions used in observational analysis of administrative databases: a systematic review. *BMC Med Res Methodol*. 2021;21(1):1-18. doi:10.1186/s12874-021-01440-5
11. DAWBER TR, MEADORS GF, MOORE FE. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health*. 1951;41(3):279-281. doi:10.2105/ajph.41.3.279
12. KANNEL WB, DAWBER TR, KAGAN A, REVOTSKIE N, STOKES J. Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study. *Ann Intern Med*. 1961;55:33-50. doi:10.7326/0003-4819-55-1-33
13. Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The Framingham study. *Am J Med*. 1977;62(5):707-714. doi:10.1016/0002-9343(77)90874-9
14. Goff DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the American college of cardiology/American heart association task force on practice guidelines. *Circulation*. 2014;129(25 SUPPL. 1):49-73. doi:10.1161/01.cir.0000437741.48606.98
15. Rana JS, Tabada GH, Solomon MD, et al. Accuracy of the Atherosclerotic Cardiovascular Risk Equation in a Large Contemporary, Multiethnic Population. *J*

Am Coll Cardiol. 2016;67(18):2118-2130. doi:10.1016/j.jacc.2016.02.055
16. Y.C. C, H.M. L, S.M. C. Does use of pooled cohort risk score overestimate the use of statin?: a retrospective cohort study in a primary care setting. *BMC Fam Pract*. 2014;15:172. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L609766475%0Ahttp://dx.doi.org/10.1186/s12875-014-0172-y.
17. Henderson KH, Kaufman BG, Stearns S, et al. Validation of the Atherosclerotic Cardiovascular Disease (Ascvd) Pooled Cohort Risk Equations By Education Level: the Atherosclerosis Risk in Communities (Aric) Study. *J Am Coll Cardiol*. 2016;67(13):1842. doi:10.1016/s0735-1097(16)31843-5
18. Cuchel M, Bruckert E, Ginsberg HN, et al. Homozygous familial hypercholesterolaemia: New insights and guidance for clinicians to improve detection and clinical management. A position paper fromthe Consensus Panel on Familial Hypercholesterolaemia of the European Atherosclerosis Society. *Eur Heart J*. 2014;35(32):2146-2157. doi:10.1093/eurheartj/ehu274
19. Mabuchi H, Koizumi J, Shimizu M, Takeda R. Development of coronary heart disease in familial hypercholesterolemia. *Circulation*. 1989;79(2):225-332. doi:10.1161/01.cir.79.2.225
20. Khera A V., Won HH, Peloso GM, et al. Diagnostic Yield and Clinical Utility of Sequencing Familial Hypercholesterolemia Genes in Patients With Severe Hypercholesterolemia. *J Am Coll Cardiol*. 2016;67(22):2578-2589. doi:10.1016/j.jacc.2016.03.520
21. Franceschini N, Giambartolomei C, de Vries PS, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat Commun*. 2018;9(1):1-14. doi:10.1038/s41467-018-07340-5
22. Gan W, Bragg F, Walters RG, et al. Genetic predisposition to type 2 diabetes and risk of subclinical atherosclerosis and cardiovascular diseases among 160,000 Chinese adults. *Diabetes*. 2019;68(11):2155-2164. doi:10.2337/db19-0224
23. Nezu T, Hosomi N, Aoki S, Matsumoto M. Carotid Intima-media thickness for atherosclerosis. *J Atheroscler Thromb*. 2016;23(1):18-31. doi:10.5551/jat.31989
24. Lorenz MW, Markus HS, Bots ML, Rosvall M, Sitzer M. Prediction of clinical cardiovascular events with carotid intima-media thickness: A systematic review and meta-analysis. *Circulation*. 2007;115(4):459-467. doi:10.1161/CIRCULATIONAHA.106.628875
25. Greenland P, Alpert JS, Beller GA, et al. 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: Executive summary: A report of the American College of cardiology foundation/American Heart association task force on practice guidelines. *Circulation*. 2010;122(25):2748-2764. doi:10.1161/CIR.0b013e3182051bab
26. Störk S, Van Den Beld AW, Von Schacky C, et al. Carotid artery plaque burden, stiffness, and mortality risk in elderly men: A prospective, population-based cohort study. *Circulation*. 2004;110(3):344-348. doi:10.1161/01.CIR.0000134966.10793.C9
27. Belcaro G, Nicolaides AN, Ramaswami G, et al. Carotid and femoral ultrasound morphology screening and cardiovascular events in low risk subjects: A 10-year follow-up study (the CAFES-CAVE study). *Atherosclerosis*. 2001;156(2):379-387. doi:10.1016/S0021-9150(00)00665-1
28. Nakahara T, Dweck MR, Narula N, Pisapia D, Narula J, Strauss HW. Coronary Artery Calcification: From Mechanism to Molecular Imaging. *JACC Cardiovasc Imaging*. 2017;10(5):582-593. doi:10.1016/j.jcmg.2017.03.005

29. Grundy SM, Stone NJ, Bailey AL, et al. *2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines*. Vol 139.; 2019. doi:10.1161/CIR.0000000000000625
30. Arnett DK, Blumenthal RS, Albert MA, et al. *2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines*. Vol 140.; 2019. doi:10.1161/CIR.0000000000000678
31. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol*. 1990;15(4):827-832. doi:10.1016/0735-1097(90)90282-T
32. Hoffmann U, Massaro JM, D'Agostino RB, Kathiresan S, Fox CS, O'Donnell CJ. Cardiovascular Event Prediction and Risk Reclassification by Coronary, Aortic, and Valvular Calcification in the Framingham Heart Study. *J Am Heart Assoc*. 2016;5(2):1-11. doi:10.1161/JAHA.115.003144
33. Detrano R, Guerci AD, Carr JJ, et al. Coronary Calcium as a Predictor of Coronary Events in Four Racial or Ethnic Groups. *N Engl J Med*. 2008;358(13):1336-1345. doi:10.1056/nejmoa072100
34. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res*. 2017;121(9):1092-1101. doi:10.1161/CIRCRESAHA.117.311312
35. Tota-Maharaj R, Blaha MJ, Blankstein R, et al. Association of coronary artery calcium and coronary heart disease events in young and elderly participants in the multi-ethnic study of atherosclerosis: A secondary analysis of a prospective, population-based cohort. *Mayo Clin Proc*. 2014;89(10):1350-1359. doi:10.1016/j.mayocp.2014.05.017
36. Paixao ARM, Ayers CR, El Sabbagh A, et al. Coronary artery calcium improves risk classification in younger populations. *JACC Cardiovasc Imaging*. 2015;8(11):1285-1293. doi:10.1016/j.jcmg.2015.06.015
37. Hermann DM, Gronewold J, Lehmann N, et al. Coronary artery calcification is an independent stroke predictor in the general population. *Stroke*. 2013;44(4):1008-1013. doi:10.1161/STROKEAHA.111.678078
38. Chaikriangkrai K, Jhun HY, Palamaner Subash Shantha G, et al. Coronary artery calcium score as a predictor for incident stroke: Systematic review and meta-analysis. *Int J Cardiol*. 2017;236:473-477. doi:10.1016/j.ijcard.2017.01.132
39. Chiles C, Duan F, Gladish GW, Ravenel JG. Association of Coronary Artery Calcification and Mortality in the National Lung Screening Trial. 2016;276(1):82-90. doi:10.1148/radiol.15142062.Association
40. Zeleznik R, Foldyna B, Eslami P, et al. Deep convolutional neural networks to predict cardiovascular risk from computed tomography. *Nat Commun*. 2021;12(1). doi:10.1038/s41467-021-20966-2
41. Gal R, Van Velzen SGM, Hooning MJ, et al. Identification of Risk of Cardiovascular Disease by Automatic Quantification of Coronary Artery Calcifications on Radiotherapy Planning CT Scans in Patients with Breast Cancer. *JAMA Oncol*. 2021;7(7):1024-1032. doi:10.1001/jamaoncol.2021.1144
42. Blaha MJ, Budoff MJ, Tota-Maharaj R, et al. Improving the CAC Score by Addition of Regional Measures of Calcium Distribution: Multi-Ethnic Study of Atherosclerosis. *JACC Cardiovasc Imaging*. 2016;9(12):1407-1416. doi:10.1016/j.jcmg.2016.03.001
43. Ferencik M, Pencina KM, Liu T, et al. Coronary Artery Calcium Distribution Is an

Independent Predictor of Incident Major Coronary Heart Disease Events: Results from the Framingham Heart Study. *Circ Cardiovasc Imaging*. 2017;10(10):1-9. doi:10.1161/CIRCIMAGING.117.006592

44. Eslami P, Foldy B, Scholtz JE, et al. Radiomics of coronary artery calcium in the framingham heart study. *Radiol Cardiothorac Imaging*. 2020;2(1). doi:10.1148/ryct.2020190119
45. Yang S, Koo BK, Hoshino M, et al. CT Angiographic and Plaque Predictors of Functionally Significant Coronary Disease and Outcome Using Machine Learning. *JACC Cardiovasc Imaging*. 2021;14(3):629-641. doi:10.1016/j.jcmg.2020.08.025
46. Tesche C, Bauer MJ, Baquet M, et al. Improved long-term prognostic value of coronary CT angiography-derived plaque measures and clinical parameters on adverse cardiac outcome using machine learning. *Eur Radiol*. 2021;31(1):486-493. doi:10.1007/s00330-020-07083-2
47. Criqui MH, Denenberg JO, Ix JH, et al. Calcium density of coronary artery plaque and risk of incident cardiovascular events. *Jama*. 2014;311(3):271-278. doi:10.1001/jama.2013.282535
48. Criqui MH, Knox JB, Denenberg JO, et al. Coronary Artery Calcium Volume and Density: Potential Interactions and Overall Predictive Value: The Multi-Ethnic Study of Atherosclerosis. *JACC Cardiovasc Imaging*. 2017;10(8):845-854. doi:10.1016/j.jcmg.2017.04.018
49. Schmidhuber J. Deep Learning in neural networks: An overview. *Neural Networks*. 2015;61:85-117. doi:10.1016/j.neunet.2014.09.003
50. Krittanawong C, Johnson KW, Rosenson RS, et al. Deep learning for cardiovascularmedicine: A practical primer. *Eur Heart J*. 2019;40(25):2058-2069C. doi:10.1093/eurheartj/ehz056
51. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights Imaging*. 2020;11(1). doi:10.1186/s13244-020-00887-2
52. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42(December 2012):60-88. doi:10.1016/j.media.2017.07.005
53. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z
54. Finnegan R, Chin V, Chlap P et al. Open-source, fully-automated hybrid cardiac substructure segmentation: development and optimisation. *Phys Med Biol*. 2022;Under Revi.
55. Finnegan R, Dowling J, Koh ES, et al. Feasibility of multi-atlas cardiac segmentation from thoracic planning CT in a probabilistic framework. *Phys Med Biol*. 2019;64(8):0-20. doi:10.1088/1361-6560/ab0ea6
56. de Vos BD, Lessmann N, de Jong PA, Išgum I. Deep learning–quantified calcium scores for automatic cardiovascular mortality prediction at lung screening low-dose ct. *Radiol Cardiothorac Imaging*. 2021;3(2). doi:10.1148/ryct.2021190219
57. Zhou R, Liao Z, Pan T, et al. Cardiac atlas development and validation for automatic segmentation of cardiac substructures. *Radiother Oncol*. 2017;122(1):66-71. doi:10.1016/j.radonc.2016.11.016
58. Zhuang X, Rhode K, Arridge S, et al. An atlas-based segmentation propagation framework using locally affine registration - Application to automatic whole heart segmentation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2008;5242 LNCS(PART 2):425-433. doi:10.1007/978-3-540-85990-1_51

59. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to radiomics. *J Nucl Med*. 2020;61(4):488-495. doi:10.2967/JNUMED.118.222893
60. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-338. doi:10.1148/radiol.2020191145
61. Berenguer R, Del Rosario Pastor-Juan M, Canales-Vázquez J, et al. Radiomics of CT features may be nonreproducible and redundant: Influence of CT acquisition parameters. *Radiology*. 2018;288(2):407-415. doi:10.1148/radiol.2018172361
62. Welch ML, McIntosh C, Haibe-Kains B, et al. Vulnerabilities of radiomic signature development: The need for safeguards. *Radiother Oncol*. 2019;130:2-9. doi:10.1016/j.radonc.2018.10.027
63. Meyer M, Ronald J, Vernuccio F, et al. Reproducibility of CT radiomic features within the same patient: Influence of radiation dose and CT reconstruction settings. *Radiology*. 2019;293(3):583-591. doi:10.1148/radiol.2019190928
64. * NLSTRT. The national lung screening trial: Overview and study design. *Radiology*. 2011;258(1):243-253. doi:10.1148/radiol.10091808
65. Chlap P, Finnegan R. Platipy documentation. https://pyplati.github.io/platipy/. Published 2022.
66. Organization WH, Geneva. *ICD-10: International Statistical Classification of Diseases and Related Health Problems*. Vol 2004. Second.
67. Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104-e107. doi:10.1158/0008-5472.CAN-17-0339
68. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *Proc 2003 IEEE Bioinforma Conf CSB 2003*. 2003;3(2):523-528. doi:10.1109/CSB.2003.1227396
69. Kawel-Boehm N, Hetzel SJ, Ambale-Venkatesh B, et al. *Reference Ranges ("Normal Values") for Cardiovascular Magnetic Resonance (CMR) in Adults and Children: 2020 Update*. Vol 22. BioMed Central; 2020. doi:10.1186/s12968-020-00683-3