

# PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://SPIDigitalLibrary.org/conference-proceedings-of-spie)

## Cardiovascular disease and all-cause mortality risk prediction from abdominal CT using deep learning

Elton, Daniel, Chen, Andy, Pickhardt, Perry, Summers, Ronald

Daniel C. Elton, Andy Chen, Perry J. Pickhardt, Ronald M. Summers, "Cardiovascular disease and all-cause mortality risk prediction from abdominal CT using deep learning," Proc. SPIE 12033, Medical Imaging 2022: Computer-Aided Diagnosis, 120332N (4 April 2022); doi: 10.1117/12.2612620

**SPIE.**

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

# Cardiovascular disease and all-cause mortality risk prediction from abdominal CT using deep learning

Daniel C. Elton<sup>a</sup>, Andy Chen<sup>a</sup>, Perry J. Pickhardt<sup>\*b</sup>, and Ronald M. Summers<sup>\*a</sup>

<sup>a</sup>Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD 20892-1182, USA

<sup>b</sup>School of Medicine and Public Health, University of Wisconsin, Madison, WI 53726, USA

## ABSTRACT

Cardiovascular disease is the number one cause of mortality worldwide. Risk prediction can help incentivize lifestyle changes and inform targeted preventative treatment. In this work we explore utilizing a convolutional neural network (CNN) to predict cardiovascular disease risk from abdominal CT scans taken for routine CT colonography in otherwise healthy patients aged 50-65. We find that adding a variational autoencoder (VAE) to the CNN classifier improves its accuracy for five year survival prediction (AUC 0.787 vs. 0.768). In four-fold cross validation we obtain an average AUC of 0.787 for predicting five year survival and an AUC of 0.767 for predicting cardiovascular disease. For five year survival prediction our model is significantly better than the Framingham Risk Score (AUC 0.688) and of nearly equivalent performance to method demonstrated in Pickhardt et al. (AUC 0.789) which utilized a combination of five CT derived biomarkers.

**Keywords:** cardiovascular disease, machine learning, deep learning, longevity, artificial intelligence, risk, abdominal CT, all-cause mortality

## 1. PURPOSE

Globally, cardiovascular disease (CVD) remains the number one cause of mortality, causing 17.9 million deaths in 2019, including 38% of all premature deaths from noncommunicable diseases.<sup>1</sup> As standards of living increase globally, people have more time and resources to invest in preventative measures to protect their long term health. Since the 1960s, the rate of deaths from CVD in the United States has been reduced by about 50% through a mixture of lifestyle changes and improved treatment. It has been argued that up to 90% of CVD is preventable if lifestyle changes and other interventions are implemented early enough.<sup>2</sup> Informing a patient that they have higher than average CVD risk can help encourage positive lifestyle changes. Additionally, several primary prevention treatments are under investigation to reduce CVD risk including metformin,<sup>3,4</sup> low-dose aspirin,<sup>5</sup> and gene therapy.<sup>6</sup> Given the costs and risks associated with these treatments, targeting them to the most at-risk patients will be important.

Genetic screening can help with targeting but completely neglects the effects of environmental exposures. A popular risk scoring system for CVD risk used currently is the Framingham risk score (FRS) which factors in age, sex, cholesterol level, and blood pressure.<sup>7</sup> In the Framingham Heart Study 8,491 patients were followed for 12 years and it was found that the FRS achieved a C-statistic (equivalent to AUC for censored data) for predicting CVD of 0.763 (95% confidence interval (CI) 0.746 - 0.780) in men and 0.793 (95% CI, 0.772 - 0.814) in women.<sup>8</sup> Previously we showed that the FRS yields an AUC of 0.688 (95% CI 0.650–0.727) for predicting five year survival for the patient cohort we will study.<sup>9</sup> A 2019 meta-analysis compared three popular risk models for predicting 10-year CVD risk - the FRS, Framingham Adult Treatment Panel III model, and pooled cohort model.<sup>10</sup> They found modest C-statistics ranging between 0.68-0.74.<sup>10</sup>

Imaging biomarkers hold much promise for improving CVD risk scoring. In particular, coronary artery calcification (CAC) scoring has been heavily studied.<sup>11,12</sup> CAC scores are typically obtained using a specialized ECG gated non-contrast cardiac CT scan. The addition of CAC score has been shown to improve risk prediction

---

\* Co-senior authors. Further author information: send correspondence to Daniel Elton ([delton@mgh.harvard.edu](mailto:delton@mgh.harvard.edu)) or Ronald Summers ([rms@nih.gov](mailto:rms@nih.gov))

when added to more traditional risk factors such as age, gender, blood biomarkers, and family history.<sup>13-15</sup> In the Multi-Ethnic Study of Atherosclerosis (MESA) dataset the FRS yielded a C-statistic for CVD risk (median follow up time 7.6 years) of 0.62 but that rose to 0.78 when CAC score was included.<sup>13</sup> Work by Mets et al. obtained a C-statistic of 0.71 (95% CI 0.67-0.76) for three year survival prediction using the National Lung Screening Trial dataset with a multivariate model that combined CAC score and clinical parameters (age, smoking status, etc.)<sup>14</sup> McClelland et al. obtained a C-statistic of 0.80 for 10 year prediction of coronary heart disease when combining CAC score with blood biomarkers, family history, age, and gender.<sup>15</sup> Recently several works have shown how deep learning systems can automate CAC scoring.<sup>16-20</sup> Interestingly, Chao et al. have demonstrated that feeding a box around the heart from a chest CT into a multi-view convolutional neural net (CNN) architecture can lead to improved CVD and mortality prediction over the state of the art for automated CAC scoring or manual CAC grading.<sup>20</sup> This suggests that CNNs can extract additional features beyond the coronary artery calcification which help inform the risk prediction. Partially by inspired that work, in this work we study feeding an entire abdominal CT scan into a CNN to perform risk scoring rather than extracting custom biomarkers such as aortic calcification score and visceral/subcutaneous fat ratio like we did in a prior work.<sup>9</sup>

The vast majority of prior work on CVD or all-cause mortality risk prediction has looked at either chest CT<sup>19-24</sup> or chest X-ray.<sup>25,26</sup> Among the works that utilize chest CT, the publicly available National Lung Screening Trial (NLST) dataset has been heavily utilized<sup>14,19,20,23,24,26</sup> in addition to the Dutch-Belgian lung cancer screening trial,<sup>21</sup> or in-house chest CT datasets.<sup>22</sup> NLST participants have a history of smoking, putting them at higher than average risk for CVD. In contrast, our dataset consists of scans from an otherwise healthy patient cohort. In this work we focus on risk scoring from abdominal CT scans, which has only been studied previously in a handful of works.<sup>9,27,28</sup> Abdominal CT are rich with biomarkers known to be relevant to cardiovascular risk such as visceral fat and aortic plaque.<sup>29,30</sup> This is the first work to feed the entire abdominal CT scan into a deep learning model.

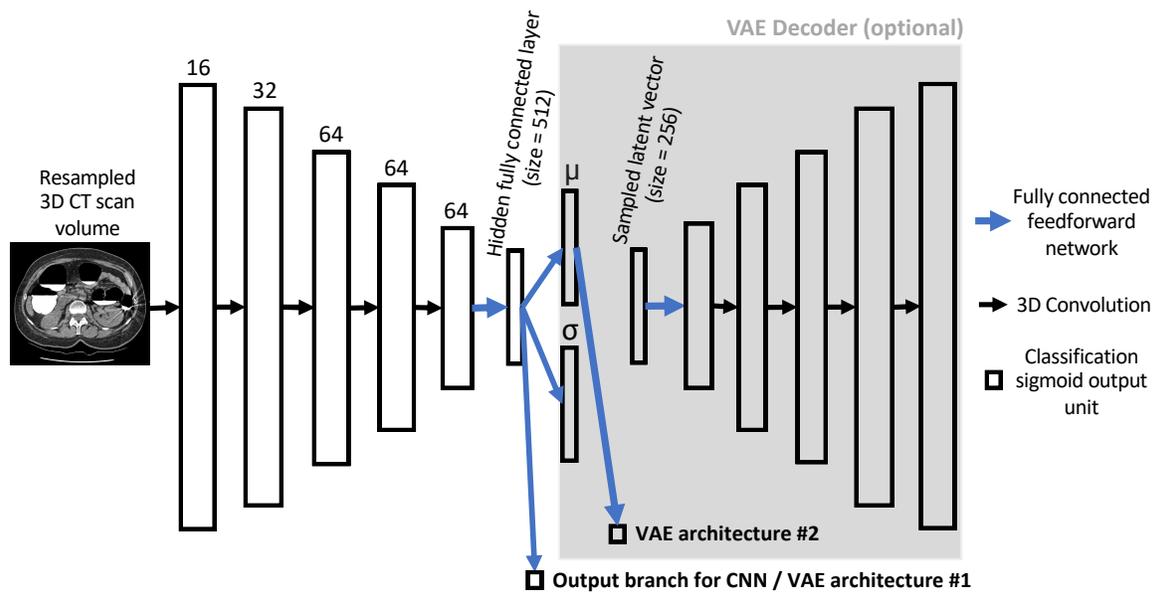


Figure 1. The architectures employed. The CNN architecture consists of five CNN layers followed by two fully connected layers. Optionally a VAE decoder may be attached to the CNN to encourage the CNN to obtain good high level features. If the VAE is attached, one can experiment with attaching the classification output to the vector of latent means (architecture 2)

## 2. METHODS

The dataset we utilize consists of CT colonography (CTC) scans from the University of Wisconsin Medical Center.<sup>27</sup> 9,223 people underwent CTC scans between April 2004 and December 2016. Further details on

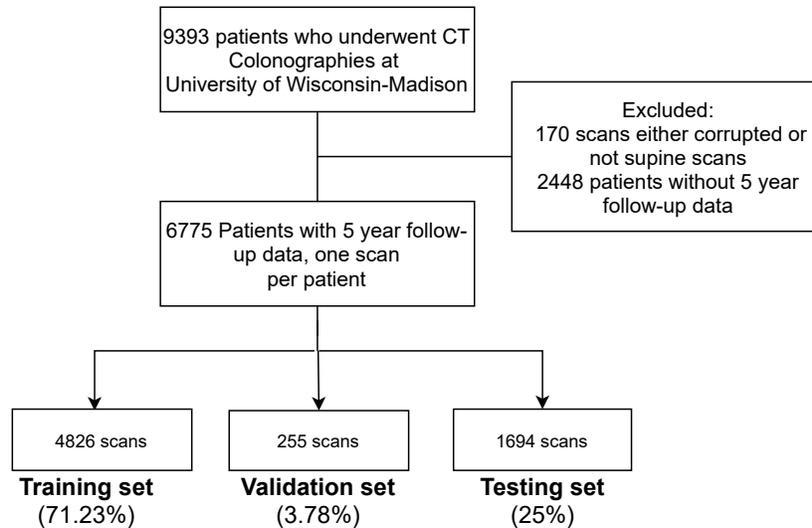


Figure 2. Patient flow chart showing how the patient cohort was selected and split into train, validation, and test sets for five year survival prediction.

the cohort are provided in Pickhardt et al, 2020.<sup>9</sup> As in prior works<sup>9,27</sup> cardiovascular disease was defined as myocardial infarction, cerebrovascular accident, or development of congestive heart failure. These reflect the endpoints considered by the FRS for cardiovascular disease.

We found 6775 patients had five year follow up data indicating if they survived for five years and out of those 216 (3%) died within five years. Similarly, 7008 patients had follow up data indicating whether or not they were diagnosed with CVD within five years, and out of those 399 developed CVD (5.6%). We used roughly 71% of the data for training, 5% for validation, and 25% for test. The data flow and number of patients in the training, validation, and testing folds for 5 year survival prediction is summarized in figure 2.

The architectures for the CNN and VAE are shown in figure 1. The two architectures were identical apart from the presence of the VAE decoder. To reduce the number of parameters in the network the first convolutional layer contains a large 7x7x7 kernel applied with a stride of 2 in each direction and padding of 3 on the edges. The rest of the convolutional layers use 4x4x4 kernels with a stride of 2 and padding of 1. Group normalization and dropout (dropout rate 0.5) were used after each convolutional layer. In summary, the CNN encoder contains 5 convolutional layers and two feedforward networks and has a total of 5,414,558 parameters. The VAE architecture contains 3 additional fully connected layers and 5 additional upsampling layers and has a total of 21,060,930 parameters. We use a relatively small latent vector size of  $D = 256$  to encourage the network to find efficient high level features.

We also attempt to add the patient age and gender information to the model. As is standard practice in machine learning, we encode the age and gender information in one-hot vectors. Ages were encoded in a vector of length 60 corresponding to ages between 30-90 years old (any patient with age less than 30 is set to 30 and any above 90 is set to 90). The one hot vectors are concatenated to the hidden layer.

The images are resampled to a size of 192x192x128 using trilinear interpolation, clipped to  $[-500, 500]$ , and then rescaled to between  $[0, 1]$ . To deal with the data imbalance (only  $\approx 3\%$  of patients died with 5 years) we reweight the data during training so the ratio of survived/died is 50/50 during training. The RAdam optimizer<sup>31</sup> was used with a learning rate of 0.001 and batch size of 4. The data augmentations employed were limited to random rotations around a random axis by  $\pm 0-15$  degrees, random 0-20% cropping in the cranial-caudal direction, and random left-right flipping.

Most existing works using a VAE, such as van Velzen et al.,<sup>23</sup> train an autoencoder first for image reconstruction and then train a separate classification model that utilizes either the latent vector or a hidden layer vector taken from the “bottleneck” portion of the VAE. In this work we perform joint training, training our

VAE model for both image reconstruction and classification at the same time. This procedure is inspired by previous work where a VAE decoder was attached to a 3D V-Net.<sup>32</sup> The author claims that adding the VAE helps regularize the 3D V-Net and prevent overfitting.<sup>32</sup> In a similar way we hypothesize that joint training with the VAE attached will improve the generalization performance of the classifier by encouraging the network to learn robust high level representations. Joint training is more challenging to implement since it requires carefully weighting the loss terms, otherwise one or more loss terms may be ignored during gradient descent training.

The loss function applied at iteration  $i$  for the VAE architecture was:

$$L_i = \beta L_{CE} + L_{L1} + \Gamma_i \gamma L_{KL} \tag{1}$$

where

$$L_{L1} = \|I_{input} - I_{pred}\| \quad L_{KL} = \frac{1}{N} \sum \mu^2 + \sigma^2 - \log \sigma^2 - 1 \tag{2}$$

The loss function for the classifier was the standard cross-entropy loss weighted by a factor  $\beta$ . For the VAE reconstruction loss we used an L1 loss function. As has been adopted in prior works,<sup>33</sup> a weighting factor  $\Gamma_i$  was applied to Kullback-Leibler loss term of the VAE and linearly increased from 0 to 1 over the course of the first 5,000 iterations. This practice prevents the optimizer from only focusing on minimizing the KL loss in the early stages of training. The factors  $\beta$  and  $\gamma$  had to be determined empirically. After experimenting with weightings of  $\beta \in \{1, 10, 100\}$  and  $\gamma \in \{0.001, 0.0001\}$  we used  $\beta = 10$  and  $\gamma = 0.001$  as this was the only combination where all three loss functions decreased during training. Further tuning of these hyperparameters was not attempted. The F1 score and AUC in the validation set were monitored during training. All models were trained for exactly four epochs, which was found to be sufficient for the validation F1 and AUC to converge.

### 3. RESULTS

method	AUC 5-year mortality	AUC 5-year CVD or mortality	AUC 5-year CVD
CNN only	0.768±0.026		
CNN+VAE (arch #1)	0.787±0.030		0.767 ± 0.036
CNN+VAE (arch #1) + age+sex	0.777±0.031		
CNN+VAE (arch #2)	0.770±0.034		
FRS <sup>9</sup>	0.688	0.695	
BMI <sup>9</sup>	0.499	0.552	
best CT derived model from Pickhardt et al. 2020 <sup>9</sup>	0.789	0.742	
best multivariate model (CT biomarkers+FRS) <sup>9</sup>	0.796	0.751	

Table 1. Summary of average AUCs in four-fold cross validation and standard deviation over the four folds.

box size	validation AUC	test AUC
128x128x128	0.75	0.74
192x192x128	0.76	0.76
192x192x192	0.78	0.79
256x256x128	0.78	0.76

Table 2. Hyperparameter optimization experiments testing different box sizes for five year survival prediction using the CNN-only architecture. Validation AUC is an average of the validation AUCs obtained in the the last epoch of training and the test AUC is the AUC on the hold-out test set for the final model after 4 epochs of training.

A summary of the average AUCs obtained in four-fold cross validation is provided in table 3 and select ROC curves are shown in figure 4. In line with our hypothesis, we found that the VAE obtained a higher AUC than the CNN only (0.787 vs 0.768), although we note that due to the limited number of folds employed means the result is not statistically significant (Welch's  $t$ -test  $p = 0.34$ ). VAE architecture option #2, where the classifier is attached to the latent vector means, performed slightly worse than VAE architecture option #1. The concatenation of age and sex information to the final layer of the classifier did not improve the AUC. This may partially be due

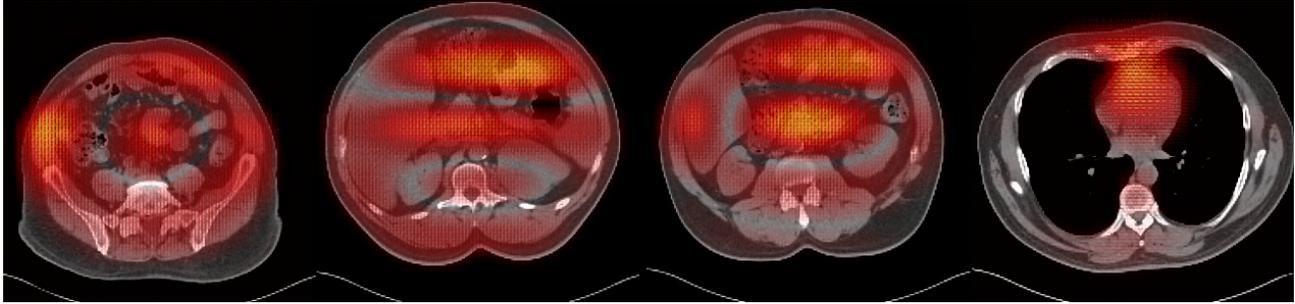


Figure 3. Visualization of a case from the publicly available Kidney Tumor Segmentation Challenge (KiTS) dataset which contains aortic plaque. The patterns observed here were found to be remarkably consistent across five different cases we looked at. This method is likely showing only where the network is looking generally in the first few layers and it is hard to draw many firm conclusions about how the model functions internally on the basis of this sort of visualization.

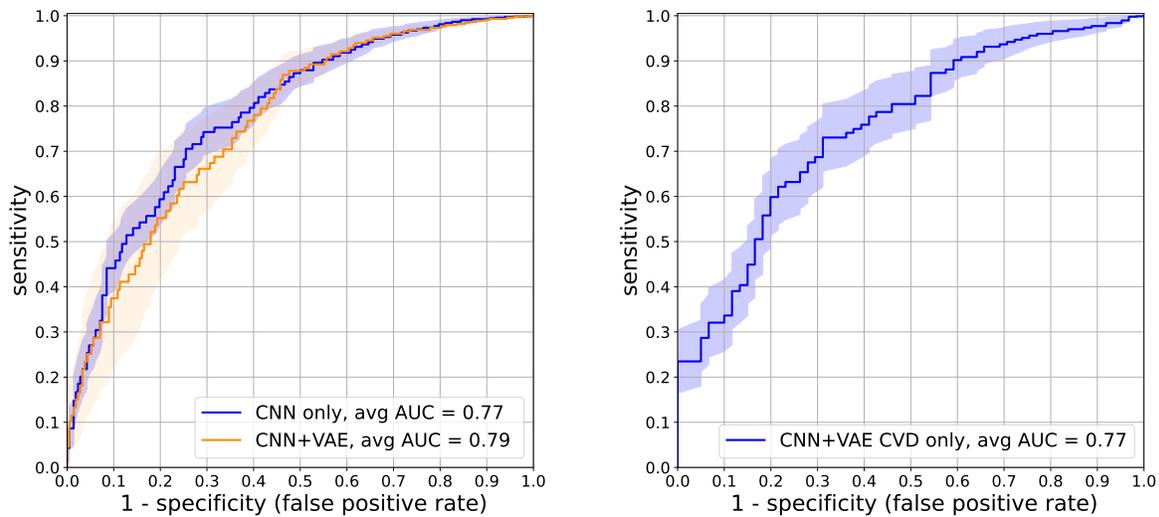


Figure 4. Select average ROC curves under 4-fold cross validation for five year survival prediction (left) and five year CVD prediction (right). The shaded regions show the standard deviation over the four folds.

to the fact that the ages of the patients did not vary that much. It may be possible to better incorporate the age and gender information, eg. by adding an additional fully connected layer.

The results in table 3 were obtained using a box size of 192x192x128. After obtaining those results we ran some hyperparameter optimization experiments to investigate the effect of box size, the results of which are shown in table 3). The results show a slightly lower AUC is found when using a smaller box of size 128x128x128 and a slightly higher AUC is obtained when the box size is increased to 192x192x192, at least for the one fold under consideration. The results suggest that a larger box of 192x192x192 may be beneficial but there is little benefit to increasing the box size beyond that.

We experimented with visualizing model using guided backpropagation<sup>34</sup> using a publicly available code-base.<sup>35</sup> An example of the visualization obtained can be found in figure 3. As shown by Adebayo et al.,<sup>36</sup> “saliency” methods like guided backprop are only sensitive to what is going on in the first few layers. Such methods show roughly where the network is looking, but do not provide much insight into what the network is doing and should be interpreted with caution.<sup>37,38</sup>

## 4. NEW OR BREAKTHROUGH WORK

For the first time we have shown how a deep learning model can predict all-cause mortality risk directly from an abdominal CT using a novel jointly trained VAE architecture. We found that the use of the VAE decoder improves the model over using just a plain CNN architecture. The VAE based model is significantly better than using the Framingham Risk Score at predicting five year mortality and is of equivalent accuracy to the best system for abdominal CT demonstrated in Pickhardt et al.<sup>9</sup> (AUC 0.787 vs 0.789). Our model has the advantage of being much simpler than the prior system, which requires the use of five independent software tools for plaque, bone mineral density, visceral/subcutaneous fat, liver fat, and muscle quantification. On a modern workstation with a NVIDIA Quadro RTX 8000 GPU our model runs in about four seconds and that time reduces to less than two seconds when our model is preloaded into GPU memory. Many avenues are open for improving our model further.

## ACKNOWLEDGMENTS

This research was supported in part by the Intramural Research Program of the National Institutes of Health Clinical Center.

## REFERENCES

- [1] “Cardiovascular diseases (CVDs) [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)),” (June 2021).
- [2] McGill, H. C., McMahan, C. A., and Gidding, S. S., “Preventing heart disease in the 21st century,” *Circulation* **117**, 1216–1227 (Mar. 2008).
- [3] Han, Y., Xie, H., Liu, Y., Gao, P., Yang, X., and Shen, Z., “Effect of metformin on all-cause and cardiovascular mortality in patients with coronary artery diseases: a systematic review and an updated meta-analysis,” *Cardiovascular Diabetology* **18** (July 2019).
- [4] Kulkarni, A. S., Gubbi, S., and Barzilai, N., “Benefits of metformin in attenuating the hallmarks of aging,” *Cell Metabolism* **32**, 15–30 (July 2020).
- [5] Patrono, C. and Baigent, C., “Role of aspirin in primary prevention of cardiovascular disease,” *Nature Reviews Cardiology* **16**, 675–686 (June 2019).
- [6] King, A., “A CRISPR edit for heart disease,” *Nature* **555**, S23–S25 (Mar. 2018).
- [7] Wilson, P. W. F., D’Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., and Kannel, W. B., “Prediction of coronary heart disease using risk factor categories,” *Circulation* **97**, 1837–1847 (May 1998).
- [8] D’Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., and Kannel, W. B., “General cardiovascular risk profile for use in primary care,” *Circulation* **117**, 743–753 (Feb. 2008).
- [9] Pickhardt, P. J., Graffy, P. M., Zea, R., Lee, S. J., Liu, J., Sandfort, V., and Summers, R. M., “Automated CT biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: a retrospective cohort study,” *The Lancet Digital Health* **2**, e192–e200 (Apr. 2020).
- [10] Damen, J. A., Pajouheshnia, R., Heus, P., Moons, K. G. M., Reitsma, J. B., Scholten, R. J. P. M., Hooft, L., and Debray, T. P. A., “Performance of the framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis,” *BMC Medicine* **17** (June 2019).
- [11] McClelland, R. L., Nasir, K., Budoff, M., Blumenthal, R. S., and Kronmal, R. A., “Arterial age as a function of coronary artery calcium (from the multi-ethnic study of atherosclerosis [MESA]),” *The American Journal of Cardiology* **103**, 59–63 (Jan. 2009).
- [12] Chiles, C., Duan, F., Gladish, G. W., Ravenel, J. G., Baginski, S. G., Snyder, B. S., DeMello, S., Desjardins, S. S., and and, R. F. M., “Association of coronary artery calcification and mortality in the national lung screening trial: A comparison of three scoring methods,” *Radiology* **276**, 82–90 (July 2015).
- [13] Yeboah, J., McClelland, R. L., Polonsky, T. S., Burke, G. L., Sibley, C. T., O’Leary, D., Carr, J. J., Goff, D. C., Greenland, P., and Herrington, D. M., “Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals,” *JAMA* **308**, 788 (Aug. 2012).

- [14] Mets, O. M., Vliegenthart, R., Gondrie, M. J., Viergever, M. A., Oudkerk, M., de Koning, H. J., Mali, W. P., Prokop, M., van Klaveren, R. J., van der Graaf, Y., Buckens, C. F., Zanen, P., Lammers, J.-W. J., Groen, H. J., Işgum, I., and de Jong, P. A., “Lung cancer screening CT-based prediction of cardiovascular events,” *JACC: Cardiovascular Imaging* **6**, 899–907 (Aug. 2013).
- [15] McClelland, R. L., Jorgensen, N. W., Budoff, M., Blaha, M. J., Post, W. S., Kronmal, R. A., Bild, D. E., Shea, S., Liu, K., Watson, K. E., Folsom, A. R., Khera, A., Ayers, C., Mahabadi, A.-A., Lehmann, N., Jöckel, K.-H., Moebus, S., Carr, J. J., Erbel, R., and Burke, G. L., “10-year coronary heart disease risk prediction using coronary artery calcium and traditional risk factors,” *Journal of the American College of Cardiology* **66**, 1643–1653 (Oct. 2015).
- [16] González, G., Washko, G. R., Estépar, R. S. J., Cazorla, M., and Espinosa, C. C., “Automated agatston score computation in non-ECG gated CT scans using deep learning,” in [*Medical Imaging 2018: Image Processing*], Angelini, E. D. and Landman, B. A., eds., SPIE (Mar. 2018).
- [17] Commandeur, F., Slomka, P. J., Goeller, M., Chen, X., Cadet, S., Razipour, A., McElhinney, P., Gransar, H., Cantu, S., Miller, R. J. H., Rozanski, A., Achenbach, S., Tamarappoo, B. K., Berman, D. S., and Dey, D., “Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study,” *Cardiovascular Research* **116**, 2216–2225 (Dec. 2019).
- [18] Lee, H., Martin, S., Burt, J. R., Bagherzadeh, P. S., Rapaka, S., Gray, H. N., Leonard, T. J., Schwemmer, C., and Schoepf, U. J., “Machine learning and coronary artery calcium scoring,” *Current Cardiology Reports* **22** (July 2020).
- [19] Zeleznik, R., Foldyna, B., Eslami, P., Weiss, J., Alexander, I., Taron, J., Parmar, C., Alvi, R. M., Banerji, D., Uno, M., Kikuchi, Y., Karady, J., Zhang, L., Scholtz, J.-E., Mayrhofer, T., Lyass, A., Mahoney, T. F., Massaro, J. M., Vasan, R. S., Douglas, P. S., Hoffmann, U., Lu, M. T., and Aerts, H. J. W. L., “Deep convolutional neural networks to predict cardiovascular risk from computed tomography,” *Nature Communications* **12** (Jan. 2021).
- [20] Chao, H., Shan, H., Homayounieh, F., Singh, R., Khera, R. D., Guo, H., Su, T., Wang, G., Kalra, M. K., and Yan, P., “Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography,” *Nature Communications* **12** (May 2021).
- [21] de Vos, B. D., de Jong, P. A., Wolterink, J. M., Vliegenthart, R., Wielingen, G. V., Viergever, M. A., and Işgum, I., “Automatic machine learning based prediction of cardiovascular events in lung cancer screening data,” in [*Medical Imaging 2015: Computer-Aided Diagnosis*], Hadjiiski, L. M. and Tourassi, G. D., eds., SPIE (Mar. 2015).
- [22] Oakden-Rayner, L., Carneiro, G., Bessen, T., Nascimento, J. C., Bradley, A. P., and Palmer, L. J., “Precision radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework,” *Scientific Reports* **7** (May 2017).
- [23] van Velzen, S., Zreik, M., Lessmann, N., Viergever, M. A., de Jong, P. A., Verkooijen, H. M., and Işgum, I., “Direct prediction of cardiovascular mortality from low-dose chest CT using deep learning,” in [*Medical Imaging 2019: Image Processing*], Angelini, E. D. and Landman, B. A., eds., SPIE (Mar. 2019).
- [24] Guo, H., Kruger, U., Wang, G., Kalra, M. K., and Yan, P., “Knowledge-based analysis for mortality prediction from CT images,” *IEEE Journal of Biomedical and Health Informatics* **24**, 457–464 (Feb. 2020).
- [25] Karargyris, A., Kashyap, S., Wu, J. T., Sharma, A., Moradi, M., and Syeda-Mahmood, T., “Age prediction using a large chest x-ray dataset,” in [*Medical Imaging 2019: Computer-Aided Diagnosis*], Hahn, H. K. and Mori, K., eds., SPIE (Mar. 2019).
- [26] Raghu, V. K., Weiss, J., Hoffmann, U., Aerts, H. J., and Lu, M. T., “Deep learning to estimate biological age from chest radiographs,” *JACC: Cardiovascular Imaging* (Mar. 2021).
- [27] O’Connor, S. D., Graffy, P. M., Zea, R., and Pickhardt, P. J., “Does nonenhanced CT-based quantification of abdominal aortic calcification outperform the framingham risk score in predicting cardiovascular events in asymptomatic adults?,” *Radiology* **290**, 108–115 (Jan. 2019).
- [28] Zambrano Chaves, J. M., Chaudhari, A. S., Wentland, A. L., Desai, A. D., Banerjee, I., Boutin, R. D., Maron, D. J., Rodriguez, F., Sandhu, A. T., Jeffrey, R. B., Rubin, D., and Patel, B., “Opportunistic assessment of ischemic heart disease risk using abdominopelvic computed tomography and medical record data: a multimodal explainable artificial intelligence approach,” *medRxiv* (2021).

- [29] Sethi, A., Taylor, L., Ruby, J. G., Venkataraman, J., Sorokin, E., Cule, M., and Melamud, E., “Calcification of abdominal aorta is a high risk underappreciated cardiovascular disease factor in a general population,” *medRxiv* (2020).
- [30] Pickhardt, P. J., Graffy, P. M., Perez, A. A., Lubner, M. G., Elton, D. C., and Summers, R. M., “Opportunistic screening at abdominal CT: Use of automated body composition biomarkers for added cardiometabolic value,” *RadioGraphics* **41**, 524–542 (Mar. 2021).
- [31] Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., and Han, J., “On the variance of the adaptive learning rate and beyond,” in [*Proceedings of the 8th International Conference on Learning Representations (ICLR)*], (2020).
- [32] Myronenko, A., “3D MRI brain tumor segmentation using autoencoder regularization,” *arXiv e-prints*, arXiv:1810.11654 (Oct. 2018).
- [33] Sandfort, V., Yan, K., Graffy, P. M., Pickhardt, P. J., and Summers, R. M., “Use of variational autoencoders with unsupervised learning to detect incorrect organ segmentations at CT,” *Radiology: Artificial Intelligence* **3**, e200218 (July 2021).
- [34] Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. A., “Striving for simplicity: The all convolutional net,” in [*3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings*], Bengio, Y. and LeCun, Y., eds. (2015).
- [35] Böhle, M., Eitel, F., Weygandt, M., and Ritter, K., “Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based alzheimer's disease classification,” *Frontiers in Aging Neuroscience* **11** (July 2019).
- [36] Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., and Kim, B., “Sanity checks for saliency maps,” in [*Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18*, 9525–9536, Curran Associates Inc., Red Hook, NY, USA (2018)].
- [37] Rudin, C., “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence* **1**, 206–215 (May 2019).
- [38] Ghassemi, M., Oakden-Rayner, L., and Beam, A. L., “The false hope of current approaches to explainable artificial intelligence in health care,” *The Lancet Digital Health* **3**, e745–e750 (Nov. 2021).