# Project Proposal: Feature Extraction on MESA CT Images

Timothy Burt, Luben Popov, Yuan Zi

## 

## Introduction

When doctors or researchers conduct quantitative analysis, real-time monitoring and treatment planning for a particular internal tissue or organ, they usually need to know detailed information about the tissue or organ in order to make correct treatment decisions. Biomedical imaging has become an indispensable part in the diagnosis and treatment of diseases. We will apply computer vision methods to extract features from cardiac Computed Tomography (CT) images and combine Big Data with ethnic information to determine the importance of ethnicity in diagnosing Atherosclerosis.

Atherosclerosis, or Arteriosclerotic vascular disease (ASVD), is a progressive disease caused by a buildup of plaque within arteries [7]. In later stages, it can lead to coronary heart disease (CHD), the leading cause of death in the world [8], as well as other potentially fatal conditions which place a substantial burden on hospital care and healthcare costs. At the same time, it is one of the most preventable diseases, with a reduction of up to 90% of CHD symptoms by diet, exercise, and not smoking. There are a cornucopia of risk factors for the disease, and how each of them is coupled to each other in various conditions/states is an active area of research.

The Multi-Ethnic Study of Atherosclerosis (MESA) is a clinical study that seeks to understand the factors that indicate the development of cardiovascular disease while the condition is still subclinical [2]. The study uses data from “a diverse, population-based sample of 6,814 asymptomatic men and women aged 45-84. Approximately 38 percent of the recruited participants are white, 28 percent African-American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent” [2].

CT scans combine multiple X-ray images taken from different angles to form cross-sectional images (or “slices”) of a patient’s internal organs [5]. The MESA study yielded paired cardiac CT scans for each patient involved in the study [6]. These CT scans provide an abundance of useful cardiac image data that, if effectively mined and utilized in a deep learning system, could allow for fast automated detection of important features that it would take a human being much longer to spot, or be easily missed. Such a tool would benefit both medical professionals and cardiac patients, allowing for easier and potentially more efficient detection of early cardiac disease.

## Goal and Objectives

### Goal

The goal of this project is to use the MESA CT images to create convolutional neural network (CNN) based deep learner that can extract important features from cardiac CT images. If the learner is successfully completed and works effectively, it will be used to develop a tool that allows a user with minimal radiology training (i.e. lab technician) to extract useful features from cardiac CT scans of patients.

The final project will consist of the deliverables, which include documentation and the code. The deliverables should be able to easily write up as a MICCAI conference paper.

### Objectives

1. Simple to use (GUI interface)
2. Works with common CT scanner brands and imaging modalities (w/wo contrast dyes)
3. Input: DICOM .dcm images, user-provided clicks (4) on top/bottom of heart slices, possibly other dimensions as inputs for the CNN (age, ethnicity, atherosclerosis +/-, blood test results, etc.)
4. Output: Relatively fast numerical analysis (<1 min. on a desktop computer), prediction of heart attacks (BETA feature), visualization of the features via some applied filters

## Methods

### Development Tools

All software development for this project will be done in the Python programming language. Since this language is the most popular choice for deep learning and digital image processing, it is an obvious choice for a computer vision project. The TensorFlow library will be used for the deep learning part of the project, as it is one of the most popular and well documented machine learning libraries available for Python. The OpenCV library will be used for the digital image processing part of the project, again due to its popularity and abundance of documentation. The frontend part of the project will be implemented using the PyQT framework, as QT is one of the most efficient and powerful UI frameworks available. GitHub will be used for version control.

### Ontology: Deep Learning

The learner for this project will use a convolutional neural network (CNN) handled by the TensorFlow library. The learner will be trained on the MESA CT images when we get access to them.

To train and build our tool before obtaining the MESA CT images, we can utilize the The Cancer Imaging Archive (TCIA) Public Access datasets. Eight of these datasets match the type of data that we’re trying to train the learner on (cardiac CT scans) [Appendix A], and all of them are fully open for public use; anyone can sign up for an account on the TCIA website and access the datasets freely, provided that the NBIA Data Retriever is installed on the user’s machine. This data can then be viewed using the free Horos DICOM viewer tool.

## Evaluation

The main factor in evaluating the success of this project is the effectiveness of our learner. To assess how effectively the learner trains and processes data, we will use 10-fold cross validation [3], where we train using 90% of our data and validate using 10% of our data a total of ten times, each time using a different 10% of the data for validation and the rest as the training data. We then take the average validation metrics (precision, accuracy, f-measure) of the ten validation cycles and use those values as our final validation metrics. This approach will protect us from overfitting and from getting lucky with a training set that happens to work well with a particular validation set. Our official validation metrics must be based on the MESA CT images, and only those images (not third party images we downloaded early in the project).

Performance is also a major consideration in evaluating our work. We must record total run times for the 10-fold cross validation using the MESA CT images, and also the run time of feature selection on a single new image introduced to an already trained learner. As stated earlier, we are targeting a runtime of less than one minute for processing a single image on suitable hardware (a desktop computer with a relatively high end GPU that has parallel compute units such as nVidia CUDA cores or AMD stream processors). Performance metrics must be consistently taken on the same machine, as changes in hardware across multiple performance tests can result in misleading data.

## Timeline

See Monday.com board timeline (can be exported as an Excel file)

## References

1. <https://wiki.cancerimagingarchive.net/display/Public/>
2. <https://www.mesa-nhlbi.org>
3. <https://en.wikipedia.org/wiki/Cross-validation_(statistics)>
4. <https://www.mayoclinic.org/tests-procedures/ct-scan/about/pac-20393675>
5. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2943871/>
6. <https://www.nhlbi.nih.gov/health-topics/atherosclerosis>
7. Califf, R. and Topol, E. (2007). *Textbook of cardiovascular medicine*. Philadelphia: Lippincott Williams & Wilkins.
8. Tesař, L., Shimizu, A., Smutek, D., Kobatake, H., & Nawano, S. (2008). Medical image analysis of 3D CT images based on extension of Haralick texture features. Computerized Medical Imaging and Graphics, 32(6), 513–520. doi:10.1016/j.compmedimag.2008.05.005
9. McKie, S. J., Hardwick, D. J., Reid, J. H., & Murchison, J. T. (2005). Features of cardiac disease demonstrated on CT pulmonary angiography. Clinical Radiology, 60(1), 31–38. doi:10.1016/j.crad.2004.07.005
10. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., & Comaniciu, D. (2008). Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features. IEEE Transactions on Medical Imaging, 27(11), 1668–1681. doi:10.1109/tmi.2008.2004421
11. Wolterink, J. M., Leiner, T., de Vos, B. D., van Hamersvelt, R. W., Viergever, M. A., & Išgum, I. (2016). Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. Medical Image Analysis, 34, 123–136. doi:10.1016/j.media.2016.04.004
12. Avci, E., & Turkoglu, I. (2009). An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases. Expert Systems with Applications, 36(2), 2873–2878. doi:10.1016/j.eswa.2008.01.030
13. Kubo, M., Niki, N., Nakagawa, S., Eguchi, K., Kaneko, M., Moriyama, N., … Yamaguchi, N. (1999). Extraction algorithm of pulmonary fissures from thin-section CT images based on linear feature detector method. IEEE Transactions on Nuclear Science, 46(6), 2128–2133. doi:10.1109/23.819294
14. Philip, K. P., Dove, E. L., McPherson, D. D., Gotteiner, N. L., Stanford, W., & Chandran, K. B. (1994). The fuzzy Hough transform-feature extraction in medical images. IEEE Transactions on Medical Imaging, 13(2), 235–240. doi:10.1109/42.293916
15. Laguna, P., Jane, R., & Caminal, P. (n.d.). Adaptive feature extraction for QRS classification and ectopic beat detection. [1991] Proceedings Computers in Cardiology. doi:10.1109/cic.1991.168986
16. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., & Comaniciu, D. (2007). Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal Space Learning and Steerable Features. 2007 IEEE 11th International Conference on Computer Vision. doi:10.1109/iccv.2007.4408925
17. Owen, D. R. J., Lindsay, A. C., Choudhury, R. P., & Fayad, Z. A. (2011). Imaging of Atherosclerosis. Annual Review of Medicine, 62(1), 25–40. doi:10.1146/annurev-med-041709-133809
18. Bild, D. E. (2002). Multi-Ethnic Study of Atherosclerosis: Objectives and Design. American Journal of Epidemiology, 156(9), 871–881. doi:10.1093/aje/kwf113
19. Achenbach, S., & Raggi, P. (2010). Imaging of coronary atherosclerosis by computed tomography. European Heart Journal, 31(12), 1442–1448. doi:10.1093/eurheartj/ehq150
20. Jayagopal, A., Linton, M. F., Fazio, S., & Haselton, F. R. (2010). Insights into Atherosclerosis Using Nanotechnology. Current Atherosclerosis Reports, 12(3), 209–215. doi:10.1007/s11883-010-0106-7
21. Von Eckardstein, A. (Ed.). (2005). Atherosclerosis: Diet and Drugs. Handbook of Experimental Pharmacology. doi:10.1007/3-540-27661-
22. Saba, L., Sanches, J. M., Pedro, L. M., & Suri, J. S. (Eds.). (2014). Multi-Modality Atherosclerosis Imaging and Diagnosis. doi:10.1007/978-1-4614-7425-8
23. Ross, R. (1999). Atherosclerosis — An Inflammatory Disease. New England Journal of Medicine, 340(2), 115–126. doi:10.1056/nejm199901143400207

## Appendix A: Open source datasets

* 8 TCIA datasets match the criteria (CT, HEART) and are open-source
  + NSCLC Radiogenomics, subject IDs (AMC-015, AMC-017)
    - Total of 2136 images, around 1 GB total (DICOM .dcm format)
    - Series spreadsheet and manifest files from TCIA webtool:
      * [NBIA-series-data-1567452924992.csv](https://cdn.discordapp.com/attachments/616839316458700822/618169531017330708/NBIA-series-data-1567452924992.csv)
      * [NBIA-manifest-1567452819121.tcia](https://cdn.discordapp.com/attachments/616839316458700822/618169538651095040/NBIA-manifest-1567452819121.tcia)