Quantification of abdominal fat from computed tomography using deep learning and its association with electronic health records in an academic biobank

Matthew T. MacLean, B.S.[1,2], Qasim Jehangir, M.D.[3], Marijana Vujkovic, Ph.D.[3], Yi-An Ko, Ph.D., Harold Litt, M.D.-Ph.D.[1], Arijitt Borthakur, Ph.D.[1], Hersh Sagraiya, M.D.[1], Mark Rosen, M.D., Ph.D.[1], David A. Mankoff, M.D.[1], Mitchell D. Schnall, M.D.-Ph.D.[2], Haochang Shou Ph.D.[4], Julio Chirinos M.D.,Ph.D.[3], Scott M. Damrauer, M.D., Ph.D.[5], Drew A. Torigian, M.D., M.A.[1], Rotonya Carr, M.D.[3], Daniel J. Rader, M.D.[2,3,*], Walter R. Witschey, Ph.D.[1,*]

[1]Department of Radiology
[2]Department of Genetics,
[3]Department of Medicine,
[4]Department of Biostatistics, Epidemiology and Informatics,
[5]Department of Surgery,
Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

*These authors contributed equally to this work.

Address correspondence to:
Walter R. Witschey
Department of Radiology,
Perelman Center for Advanced Medicine
3400 Civic Center Boulevard
Philadelphia, PA 19104
Email: witschey@pennmedicine.upenn.edu

Keywords: Deep learning, subcutaneous fat, visceral fat, body mass index
Word count:

Abstract

Objective

The objective was to develop a fully automated algorithm for abdominal fat segmentation and deploy this method at scale and associated with diagnoses in an academic biobank.

Materials and Methods

We built a fully automated image curation and labeling technique using deep learning and distributive computing to identify subcutaneous and visceral abdominal fat compartments from 47,587 CT scans in 13,422 patients in the Penn Medicine Biobank (PMBB). A classification network identified the inferior and superior borders of the abdomen, and a segmentation network differentiated visceral and subcutaneous fat. Following technical evaluation of our method, we conducted studies to validate known relationships with adiposity.

Results

When compared with 100 manually annotated cases, the classification network was on average within one 5 mm slice for both the superior (0.3±0.6 slices) and inferior (0.7±0.6 slices) borders. The segmentation network also demonstrated excellent performance with interclass correlation coefficients of 0.99 ($p<2e-16$) for subcutaneous and 0.99 ($p<2e-16$) for visceral fat on 100 testing cases. We performed integrative analyses of abdominal fat with the phenome extracted from the electronic health record and found highly significant associations with diabetes mellitus, hypertension, renal failure, among other phenotypes.

Conclusion

This work presents a fully automated and highly accurate method for the quantification of abdominal fat that can be applied to routine clinical imaging studies to fuel translation scientific discovery.

Introduction

Advanced imaging is a critical component of healthcare. It is utilized extensively in modern medicine and is a major component of the electronic health record (EHR). Medical centers collect enormous quantities of imaging data that could be extremely valuable for translational science, but quantitative traits are not systematically extracted for use by either researchers or clinicians. One of these quantitative traits is abdominal adipose mass, which is highly relevant to human health and disease and can be quantified from medical images such as CT scans. Obesity, a condition of increased adiposity, has been associated with numerous diseases including cardiovascular disease, diabetes, stroke, and cancer [1 2]. However, obesity is diagnosed by body mass index which is a poor measure of adipose as it is calculated using only weight and height and does not account for variations in body composition [3].

Furthermore, not all fat is equal. Visceral fat, which is located within the abdominal cavity adjacent to vital organs, portends an even greater risk for many pathologies including cardiovascular disease, insulin resistance, and certain cancers [4]. Multiple studies have investigated the unique metabolic properties and neuroendocrine signaling mechanisms of visceral fat to understand its pathogenicity [4-6]. However, the complex relationships between visceral adipose and components of metabolic syndrome are not yet understood. Fundamental to the continued study of this topic is the ability to accurately quantify and distinguish visceral and subcutaneous fat. Many studies rely on waist circumference (WC) as a proxy for visceral adipose [7-9]. However, while WC has demonstrated clinical significance beyond BMI, it correlates more strongly with subcutaneous than visceral adipose [10-12]. Computed tomography (CT) scans offer a solution as they are routinely performed and provide a cross sectional view of anatomy that is often used to measure abdominal fat.

Robust and automatic techniques for extracting fat imaging traits from CT may improve consistency of diagnosis between patients with similar fat profiles and pathology and aid caregivers. Likewise, fully automated segmentations are needed to develop large-scale investigations of fat and integrate imaging biomarkers with other EHR data. Such large-scale repositories are becoming available in health biobanks [13]. However, fat quantification is time consuming, and it is not feasible to perform manual or semi-automated analysis on thousands of clinical studies. Methods that could rapidly and accurately derive fat measures would be of great benefit to this field of population-based clinical health.

Many methods that have been proposed for automatic fat quantification have limitations that prevent their application to the noisy data of a large clinical cohort. These methods often rely on expected anatomic attenuation profiles to apply shape models, active contours, or other

statistical approaches [14-17]. These can work well on a curated dataset but easily fail when artifacts or unexpected pathology is encountered. Deep learning offers a data-driven approach to this problem and has been applied to a range of biomedical segmentation and classification tasks with impressive results [18-21]. Specific to abdominal fat quantification, deep learning has been utilized in multiple studies [22-25]. However, many of these methods are only applicable to single-slice quantification and do not address identifying the slices of interest. Additionally, further work is needed to evaluate the utility of applying these deep learning approaches at scale on a diverse dataset. Altogether, these clinical and translational applications are increasingly motivating a need for automated methods to extract fat biomarkers from CT.

We built a fully automated abdomen and pelvis image curation and fat labeling technique using deep learning and applied it to CT scans to identify subcutaneous and visceral fat compartments. Two deep learning networks were trained and applied to previously unseen imaging data: 1) given a complete CT scan of the abdomen and pelvis, a classification network was trained to label a subset of axial CT images that showed the abdomen and 2) a segmentation network was trained to label pixels belonging to the abdominal compartment. After both networks were applied, subcutaneous and visceral fat compartments were segmented using intensity-based thresholding of the interior and exterior abdominal compartments. The technique was validated using an independent set of abdomen and pelvis CT images. It was subsequently applied it to 47,587 CT scans from 13,422 patients enrolled in the Penn Medicine Biobank (PMBB), a centralized resource of annotated blood and tissue samples linked with clinical EHR and genetic data. As additional validation of the methodology, we performed integrative analyses of the imaging traits with other phenotypic data extracted from the EHR including blood biomarkers, body mass index, and diagnoses (ICD-9) codes.

Materials and Methods

Penn Medicine Biobank

This study used data collected from participants in the Penn Medicine Biobank (PMBB). The PMBB is a resource for advanced imaging, genetics, blood biomarkers and other electronic health record data at Penn Medicine, a multi-hospital health system headquartered in Philadelphia, PA. It is a detailed long-term, prospective, epidemiological study of over 50,000 volunteers containing approximately 27,485 unique diagnostic codes (ICD9 and ICD10) and 1,025,963 radiology studies. All patients provided informed consent to participate in the PMBB and to utilization of electronic health record and image data.

Study Cohort

At the time this study was completed (January 2020), the biobank had enrolled 52,441 patients. Participants of the PMBB were included in this study if they had an abdominal and pelvis CT scan as indicated by an Current Procedural Terminology (CPT) code 74176 (n=10,093), 74177 (n=19,431), and 74178 (n=1,923). CPT codes were queried directly from image (dicom) metadata for PMBB participants. A detailed flowchart showing the number of participants in this study, imaging studies, and number of imaging series (multiple series are collected per scan) is shown in Figure 1. Imaging series were excluded if they were less than 125 millimeters (mm) in the superior-inferior dimension, as they would be unlikely to include the entire abdominal field. Studies were also excluded if the abdominal compartment was determined to be less than 100 mm in the superior-inferior dimension; these studies were excluded after the classification network had been applied. Details of the demographic and clinical summary statistics are available in Table 1.

Image Analysis

A schematic of the overall approach to subcutaneous and visceral fat segmentation using deep learning is shown in Figure 2. A detailed description of the implementation, training and prediction steps are discussed in this section. Briefly, two deep learning neural networks were used to 1) classify 2D images of the abdomen or pelvis as showing the abdomen and then 2) segment these images for the abdominal compartment. In a final step, intensity-based thresholding criteria were applied to visceral and subcutaneous compartments to label fat in these regions.

Ground truth segmentations

Ground truth segmentations were provided to the classification and segmentation networks for training. All training data was manually generated by a trained technician under the supervision of a board-certified abdominal radiologist. As discussed in more detail below, the two-category classification network identified axial CT images located in the abdomen (e.g. 0 = not in abdomen, 1 in abdomen). Training data consisted of images and their categories as provided by the technician and radiologist. The superior border of the abdomen was defined as the axial view that the lungs were no longer visible. The inferior border was defined as the first axial view that the bottom of the L5 vertebra was visible. These borders were manually labeled on 468 training scans of which 375 scans (35,305 slices) were used for training and 93 scans (8,775) for validation.

The segmentation network was trained to label pixels inside the abdomen. The network was trained on a total of 62 scans with 50 scans (2,059 slices) randomly selected for training and 12 scans (498 slices) for validation. Training data was selected iteratively when the model underperformed on a scan. Of note, the network was not trained to directly label subcutaneous and visceral fat. Rather, the segmentation consisted of two steps, where 1) the abdominal compartment was segmented, as shown in Figure 2, and 2) fat voxels were located by thresholding for HU between -190 and -30. All fat voxels within the abdominal compartment segmentation are visceral fat and all those outside the segmentation are subcutaneous fat. Image augmentation included variations in zoom, rotation, horizontal flipping, image shear, brightness, as well as width and height shifting.

Hardware Specifications and Data Processing

All architectures were implemented in Python using the Tensorflow package. Training was conducted using an NVIDIA P100 graphical processing unit (GPU) and inference utilized NVIDIA K80 GPUs. The network inputs were 2D axial slices with size of 256x256 pixels. A window width of 1800 HU and level of 400 HU was used for the abdominal compartment delineation network and a window width of 400 HU and a level of 50 HU was used for the abdominal contour segmentation network. The architecture of the neural networks is depicted in Supplementary Figure 1. Processing was performed in approximately 4 hours (300 milliseconds/scan) by utilizing distributive cloud computing across 50 virtual instances each equipped with an NVIDIA K80 GPU.

Classification Model Architecture – Identification of Images Showing Abdominal Anatomy

The first network labeled 2D slices as belonging to the abdominal cavity. Images were shown to convolutional layers which flattened into fully connected layers modeled after the VGG-16 classification network (Supplementary Figure 1A) [26]. This network standardized the quantification of subcutaneous and visceral fat across patients by identifying defined anatomical endpoints for the abdominal cavity. The network was trained to output a probability between 0 and 1 indicating the likelihood that the slice is within the boundary of the abdomen. To extract the boundaries from this array of probabilities we first subtracted 0.50 from every value such that such that values are in the range [-0.5, 0.5]. Next, we apply Kadane's algorithm to find the contiguous sub-list of these values that gives the maximum sum [27]. Image augmentation was utilized to expand the training set and included variations in zoom, rotation, as well as width and height shifting. Performance of this network was evaluated on a testing set of 100 abdomen/pelvis CT scans selected at random from the PMBB.

The key differences between the implementation shown here and the implementation originally described in Simonyan for the ImageNet 2014 challenge were the following:

(1) The proposed network may best resemble the network B described in [28], which has two convolutional layers before pooling each layer. There were three fully connected intermediate layers in each. In our network, the final fully connected output layer had two output categories instead of 1000 categories used for image classification; this layer was activated using a softmax function.

(2) There were differences in the input image matrix size and number of filter channels per layer. These differences were mainly to accommodate the larger image size of the CT images.

Segmentation Model Architecture – Labeling of subcutaneous and visceral fat pixels

An additional network was trained to help delineate subcutaneous and visceral fat from axial 2D slices modeled after a U-Net architecture (Supplementary Figure 1B) [29]. The networks output a probability for each voxel indicating the probability that it belongs to the foreground, and probabilities greater than or equal to 0.50 are attributed to the foreground. Additional post-processing steps include only preserving the largest connected component and filling any holes for each slice.

The key differences between the segmentation CNN and the original implementation described in Ronneberger, et al. [29] for electron microscopy are the following:

(1) Our input image matrix size was 256x256. The reasons for this smaller image size were practical: CT images have somewhat smaller matrix size than the electron microscopy

slides. Also, smaller matrix size allowed for the U-net to require less computational resources (RAM and CPU).

(2) We had 14 instead of 18 convolutional layers. Again, this is mainly due to the practical need for a smaller network, given that the input image had a smaller matrix size. For the same reasons, the network used 3 skip connections instead of 4.

(3) We included a dropout step after the final convolution layer prior label prediction to reduce overfitting. 20% of data was removed during dropout. Details of the dropout are described elsewhere [30].

To evaluate segmentation performance, two approaches were utilized. First, 100 abdomen/pelvis CT scans were randomly selected from the PMBB and visceral/subcutaneous fat was calculated using a manual contour on a single slice between L3 and L4. These values were compared to automatically derived measurements. Second, an additional testing set of 10 scans was randomly selected and both manual as well as automated segmentations for the abdominal contour were produced on every abdominal slice. DICE coefficients were calculated to measure agreement between manual and automatic segmentations for the abdominal contour as well as visceral and subcutaneous fat.

Association Studies

BMI Correlation Analysis

The CNN derived metrics for subcutaneous and visceral fat were further validated by comparing these values to clinically assessed BMI values. The most recent BMI measurement was attributed to each scan, and measurements acquired greater than 365 days from the scan were excluded. Pearson's correlation coefficient was computed to measure the degree of association. In all association analyses, we utilized the average metric area of subcutaneous or visceral fat across all slices. To associate a single fat value to each patient, we took the median (or mean if exactly two) value for studies with multiple scans or for patients with multiple studies. This approach for attributing a single image derived phenotype to a patient was utilized in all association studies.

Phenome-wide Association Study

A phenome-wide association study (PheWAS) was performed to investigate the phenotypic associations of having a higher visceral to subcutaneous fat ratio (VSR). ICD10 codes were first mapped to ICD9 codes using the 2017 general equivalency mapping (GEM).

Next, ICD9 codes were aggregated into Phecodes using the PheWAS R package to create 1,816 phecodes. Patients with at least 2 occurrences of a PheCode are considered cases, those with none are controls, and those with 1 are treated as missing. Phecodes with less than 100 cases were excluded. Logistic regression was then performed with each phecode as the outcome and VSR as a predictor. Regression was performed controlling for the covariates of age, sex, and race. Bonferroni multiple comparison correction was used to determine the level of significance.

Relationship between Lab Values and VSR

To investigate the relationship between VSR and common clinical laboratory studies, VSR values were organized into quartiles. Since VSR is associated with both sex and race, patients were first stratified into four categories based on sex and race. Patients with the highest and lowest 25% of values within each of these categories were identified and the distribution of laboratory measures between the high and low groups was compared. Lab values acquired greater than 90 days and BMI measurements acquired greater than 365 days from the scan-date were excluded. To test for difference between the distributions, A Wilcox rank-sum test was then performed.

Results

Patient Cohort Acquisition and Analysis

After applying exclusion criteria, 31,442 studies containing a total of 52,875 scans corresponding to 13,508 patients were analyzed for visceral and subcutaneous fat. Following the exclusion of additional scans based on the size of abdominal window and volume of fat detected, 31,203 studies containing 47,587 scans corresponding to 13,422 patients were utilized in association studies (Figure 1).

Image Analysis – Technical Validation

Classification Network

The performance of the classification CNN to identify the abdominal borders was assessed by randomly selecting 100 CTs of the abdomen and pelvis and comparing the CNN and manually derived results. The automated method was on average within one 5 mm slice from the manually selected slice for both the superior (0.3±0.6 slices) and inferior (0.7±0.6 slices) borders. For the superior border, the prediction ranged from 3 slices below to 2 slices above the manual label. For the inferior border they ranged from 8 slices below (one vertebral level) to 3 slices above the manual label. Before application of the maximum sub-list algorithm, the localizer network had an accuracy of 0.99±0.01, sensitivity of 0.99±0.03, specificity of 0.99±0.01 in classifying slices. After application of Kaden's algorithm, the accuracy, sensitivity, and specificity was unchanged.

Segmentation Network

Performance of the segmentation CNN in labeling visceral and subcutaneous fat was found by comparing predictions with manual contours. Representative segmentation results are shown in Figure 3. Scatterplots and Bland-Altman plots showing evaluation results for the single slice segmentations on 100 different patients is shown in Figure 4 (Panels A-D). Interclass correlation coefficients of 0.9999 ($p<2e-16$) and 0.9998 ($p<2e-16$) were achieved for the prediction of subcutaneous and visceral fat, respectively. There was a significant bias of 1.5±1.6 cm^2 ($p=3.0e-15$) for subcutaneous and -1.5±1.6 cm^2 ($p=3.0e-15$) for visceral fat. For this analysis, the average areas were 301.8±172.0 cm^2 for subcutaneous and 154.14±109.28 for visceral fat. Region-of-interest area overlap ratios (Dice coefficients) were calculated for the additional testing set of 10 randomly selected CTs for which the entire abdominal compartment

was manually contoured. This evaluation resulted in high Dice values for the abdominal compartment (0.983±0.005), subcutaneous fat (0.998±0.003), and visceral fat (0.990±0.005).

Association Studies

Association studies were conducted investigating the relationship between the CNN derived fat values and BMI, clinical lab values, and billing codes. The relationships between the CNN derived fat values and BMI are shown in Figure 4 (Panels E and F). There was a significant correlation between BMI and both subcutaneous ($r=0.877$, $p<2e-16$) and visceral ($r=0.524$, $p<2e-16$) fat.

Next, we compared the distribution of lab values and BMI measurements for patients in the bottom quartile of VSR values to those in the top quartile. For the high VSR group there was a significant increase in triglycerides ($p=3.7e-10$), glycated hemoglobin (A1C; $p=1.3e-4$), blood urea nitrogen (BUN; $p=3.5e-12$), creatinine ($p=1.5e-53$) and BMI ($p=1.6e-4$). There was also a significant decrease in high-density lipoprotein (HDL; $p=0.0020$). Density plots showing the distribution of values for the two groups are shown in Figure 5.

A phenome-wide association study of VSR revealed significant associations with several pathologies. The plot is shown in Figure 6. The strongest association was with diabetes mellitus ($p=1.4e-22$). There were multiple hits for other endocrine disorders related to diabetes or lipid metabolism. Within the circulatory system, the strongest signal was with hypertension ($p=1.2e-18$) followed by hypertension-related complications. In the renal system, there were seven significant hits, including chronic kidney disease ($p=5.0e-17$), renal failure ($p=1.7e-14$), and renal transplant ($p=3.8e-12$).

Discussion

In this paper, we present a fully automated approach to accurately quantify abdominal fat from clinical CT scans. By utilizing distributive processing in a cloud computing environment, we quantified full abdominal volumes of visceral and subcutaneous fat from 52,875 scans in 4 hours (~216 scans per minute). This demonstrates the efficiency of a fully trained deep learning network to label abdominal CT images at a high rate which is necessary to perform large-scale research applications or to support automatic, quantitative reporting of abdominal fat by radiologists. The segmentation algorithm builds on recent developments in deep learning [19 28]. This new application used a classification network, which labels images between the inferior aspect of the lung and L5 vertebrae, and a segmentation network, which labels visceral and subcutaneous fat pixels. The algorithm showed excellent agreement with manual segmentation of visceral and subcutaneous fat compartments. To our knowledge, this is the first automatic technique for CT abdominal fat segmentation to be associated at scale to electronic health record data from an academic biobank.

Computational methods are needed to consistently identify abdominal anatomy for fat quantification. Several studies used the spine as a fiducial marker to identify a 2D slice [31-33], yet the accuracy of single slice measurements of fat area as an indirect marker for total abdominal fat is uncertain. Computational methods that measure total fat within the abdomen from 3D imaging [14 25] are expected to show higher accuracy across a range of body habitus. In our study, the outstanding agreement between manual and automatic techniques was likely attributed to the 3D high resolution input images, increased layer density of the VGG-based model and diversity of the training data. Kadane's algorithm, which found landmarks directly from the output of the neural network, improves accuracy when compared with a less sophisticated approach that simply identifies the largest contiguous positive sub-list.

Following identification of abdominal boundaries, we used a CNN to segment the abdominal contour to differentiate subcutaneous and visceral fat. Our results showed outstanding agreement between manual and automated measurements of fat. There was a significant bias of 1.45 cm^2 for both subcutaneous and visceral fat. However, compared to the average areas of 301.82 cm^2 for subcutaneous and 154 cm^2 for visceral fat in our testing set, this bias represented a small error (<1%). Previous studies that utilized radial projections or shape analysis were not tested on patients with extreme body habitus [14 34]. Figure 3 highlights the fault-tolerance of the algorithm to variations in patient body habitus and BMI.

We found that the association between subcutaneous or visceral fat and BMI was strong and similar compared with previous studies that found associations of r=0.73-0.93 for

subcutaneous and r=0.61-0.77 for visceral fat [35-37]. In agreement with these studies, we found a stronger association between BMI and subcutaneous fat than with visceral fat. When comparing lab values between patients with high and low values of VSR, the high VSR group had significantly higher triglycerides and lower HDL but no significant change in LDL. This was consistent with previous findings that VSR was positively associated with triglycerides, negatively associated with HDL, and showed no significant association with LDL [17]. Glycated hemoglobin (A1C) levels were significantly elevated in our analysis which aligns with a known association with diabetes [4]. While an association between VSR and liver enzymes has been reported [38], our analysis did not find significant associations for AST or ALT. Our finding of increased creatinine and BUN in the higher VSR group is consistent with previous findings that visceral fat is associated with decreased renal function and progression to end-stage renal disease, particularly in diabetic kidney disease [39 40].

By investigating association between VSR and phecodes, we were able to interrogate disease associations in an unbiased manner. As expected, we found numerous associations with diabetic, hypertensive, and kidney disease pathologies. There were some unexpected associations with transplants and human immunodeficiency virus (HIV) infections. The association with HIV is likely due to lipodystrophy which is commonly seen in these patients [41]. Similarly, the association with transplants is likely secondary to the use of corticosteroids following transplant which is known at high doses to increase visceral adiposity [42]. The negative association with bariatric surgery is likely because patients who get bariatric surgery have significant depots of subcutaneous fat.

There are several limitations to this study. While the attenuation range for fat has been defined in literature, CT scans acquired over several decades may contain artifacts or utilize reconstruction algorithms that distort attenuation outside of the range. Given our automated approach, it is possible that scans do not reach the inferior or superior borders of the abdomen, and this could skew the resulting fat values depending on which part of the abdomen is visualized. We attempted to correct for this by eliminating scans with fewer than 100 mm of abdominal slices. This study also derived disease phenotypes from EHR billing codes which are often incomplete. As this was a retrospective cohort, there could be significant selection bias for sicker patients or certain diseases based on biobank recruitment methods. Further work should be performed to investigate the associated phenotypes to refine our understanding and identify any causative relationships. Additionally, there is great potential in the utilization of this method in the context of a biobank such as the PMBB where genetic information is available. This may

provide greater insight into the mechanism of pathogenicity for visceral fat, which would be of great interest to the scientific community.

In conclusion, this study presents a fully automated method for the quantification of abdominal fat that functions with high accuracy and can be applied efficiently in a cloud computing environment. This integration further highlights how autonomous image trait quantification can facilitate translational research especially in the context of an academic biobank.

References

1. Frezza EE, Wachtel MS, Chiriva-Internati M. Influence of obesity on the risk of developing colon cancer. Gut 2006;**55**(2):285-91 doi: 10.1136/gut.2005.073163[published Online First: Epub Date]|.
2. Van Gaal LF, Mertens IL, De Block CE. Mechanisms linking obesity with cardiovascular disease. Nature 2006;**444**(7121):875-80 doi: 10.1038/nature05487[published Online First: Epub Date]|.
3. Burkhauser RV, Cawley J. Beyond BMI: the value of more accurate measures of fatness and obesity in social science research. J Health Econ 2008;**27**(2):519-29 doi: 10.1016/j.jhealeco.2007.05.005[published Online First: Epub Date]|.
4. Bergman RN, Kim SP, Catalano KJ, et al. Why visceral fat is bad: mechanisms of the metabolic syndrome. Obesity (Silver Spring) 2006;**14 Suppl 1**:16s-19s doi: 10.1038/oby.2006.277[published Online First: Epub Date]|.
5. Funahashi T, Nakamura T, Shimomura I, et al. Role of adipocytokines on the pathogenesis of atherosclerosis in visceral obesity. Intern Med 1999;**38**(2):202-6 doi: 10.2169/internalmedicine.38.202[published Online First: Epub Date]|.
6. Matsuzawa Y, Funahashi T, Nakamura T. The concept of metabolic syndrome: contribution of visceral fat accumulation and its molecular mechanism. J Atheroscler Thromb 2011;**18**(8):629-39 doi: 10.5551/jat.7922[published Online First: Epub Date]|.
7. Janssen I, Katzmarzyk PT, Ross R. Body mass index, waist circumference, and health risk: evidence in support of current National Institutes of Health guidelines. Arch Intern Med 2002;**162**(18):2074-9 doi: 10.1001/archinte.162.18.2074[published Online First: Epub Date]|.
8. Pouliot MC, Despres JP, Lemieux S, et al. Waist circumference and abdominal sagittal diameter: best simple anthropometric indexes of abdominal visceral adipose tissue accumulation and related cardiovascular risk in men and women. Am J Cardiol 1994;**73**(7):460-8 doi: 10.1016/0002-9149(94)90676-9[published Online First: Epub Date]|.
9. Lofgren I, Herron K, Zern T, et al. Waist circumference is a better predictor than body mass index of coronary heart disease risk in overweight premenopausal women. J Nutr 2004;**134**(5):1071-6 doi: 10.1093/jn/134.5.1071[published Online First: Epub Date]|.
10. Staiano AE, Reeder BA, Elliott S, et al. Body mass index versus waist circumference as predictors of mortality in Canadian adults. Int J Obes (Lond) 2012;**36**(11):1450-4 doi: 10.1038/ijo.2011.268[published Online First: Epub Date]|.
11. Bosy-Westphal A, Booke CA, Blocker T, et al. Measurement site for waist circumference affects its accuracy as an index of visceral and abdominal subcutaneous fat in a Caucasian population. J Nutr 2010;**140**(5):954-61 doi: 10.3945/jn.109.118737[published Online First: Epub Date]|.
12. Camhi SM, Bray GA, Bouchard C, et al. The relationship of waist circumference and BMI to visceral, subcutaneous, and total body fat: sex and race differences. Obesity (Silver Spring) 2011;**19**(2):402-8 doi: 10.1038/oby.2010.248[published Online First: Epub Date]|.
13. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;**562**(7726):203-09 doi: 10.1038/s41586-018-0579-z[published Online First: Epub Date]|.
14. Zhao B, Colville J, Kalaigian J, et al. Automated quantification of body fat distribution on volumetric computed tomography. J Comput Assist Tomogr 2006;**30**(5):777-83 doi: 10.1097/01.rct.0000228164.08968.e8[published Online First: Epub Date]|.
15. Automatic segmentation of abdominal fat from CT data. 2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1; 2005. IEEE.

16. Kim YJ, Lee SH, Kim TY, Park JY, Choi SH, Kim KG. Body fat assessment method using CT images with separation mask algorithm. J Digit Imaging 2013;**26**(2):155-62 doi: 10.1007/s10278-012-9488-0[published Online First: Epub Date]|.
17. Cha EDK, Veturi Y, Agarwal C, Patel A, Arbabshirani MR, Pendergrass SA. Using Adipose Measures from Health Care Provider-Based Imaging Data for Discovery. J Obes 2018;**2018**:3253096 doi: 10.1155/2018/3253096[published Online First: Epub Date]|.
18. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2016. Springer.
19. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;**542**(7639):115-18 doi: 10.1038/nature21056[published Online First: Epub Date]|.
20. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. Jama 2016;**316**(22):2402-10 doi: 10.1001/jama.2016.17216[published Online First: Epub Date]|.
21. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. Jama 2017;**318**(22):2199-210 doi: 10.1001/jama.2017.14585[published Online First: Epub Date]|.
22. Jiang F, Li H, Hou X, et al. Abdominal adipose tissues extraction using multi-scale deep neural network. Neurocomputing 2017;**229**:23-33
23. Weston AD, Korfiatis P, Kline TL, et al. Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. Radiology 2019;**290**(3):669-79 doi: 10.1148/radiol.2018181432[published Online First: Epub Date]|.
24. Park HJ, Shin Y, Park J, et al. Development and Validation of a Deep Learning System for Segmentation of Abdominal Muscle and Fat on Computed Tomography. Korean J Radiol 2020;**21**(1):88-100 doi: 10.3348/kjr.2019.0470[published Online First: Epub Date]|.
25. Wang Y, Qiu Y, Thai T, Moore K, Liu H, Zheng B. A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images. Comput Methods Programs Biomed 2017;**144**:97-104 doi: 10.1016/j.cmpb.2017.03.017[published Online First: Epub Date]|.
26. Hassan M. VGG16: Convolutional Network for Classification and Detection, 2019.
27. Takaoka T. Efficient algorithms for the maximum subarray problem by distance matrix multiplication. Electronic Notes in Theoretical Computer Science 2002;**61**:191-200
28. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv: 1409.1556 2015
29. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv: 1505.04597 2015
30. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;**15**:1929-58
31. Kullberg J, Hedstrom A, Brandberg J, et al. Automated analysis of liver fat, muscle and adipose tissue distribution from CT suitable for large-scale studies. Sci Rep 2017;**7**(1):10425 doi: 10.1038/s41598-017-08925-8[published Online First: Epub Date]|.
32. Chung SJ, Kim D, Park MJ, et al. Metabolic syndrome and visceral obesity as risk factors for reflux oesophagitis: a cross-sectional case-control study of 7078 Koreans undergoing health check-ups. Gut 2008;**57**(10):1360-5 doi: 10.1136/gut.2007.147090[published Online First: Epub Date]|.
33. Kim S, Cho B, Lee H, et al. Distribution of abdominal visceral and subcutaneous adipose tissue and metabolic syndrome in a Korean population. Diabetes Care 2011;**34**(2):504-6 doi: 10.2337/dc10-1364[published Online First: Epub Date]|.

34. Chung H, Cobzas D, Birdsell L, Lieffers J, Baracos V. Automated segmentation of muscle and adipose tissue on CT images for human body composition analysis. SPIE Medical Imaging 2009;**7261**
35. Janssen I, Heymsfield SB, Allison DB, Kotler DP, Ross R. Body mass index and waist circumference independently contribute to the prediction of nonabdominal, abdominal subcutaneous, and visceral fat. The American journal of clinical nutrition 2002;**75**(4):683-88
36. Kim JH, Doo SW, Cho KS, et al. Which anthropometric measurements including visceral fat, subcutaneous fat, body mass index, and waist circumference could predict the urinary stone composition most? BMC urology 2015;**15**(1):17
37. Camhi SM, Bray GA, Bouchard C, et al. The relationship of waist circumference and BMI to visceral, subcutaneous, and total body fat: sex and race differences. Obesity 2011;**19**(2):402-08
38. Jung C-H, Rhee E-J, Kwon H, Chang Y, Ryu S, Lee W-Y. Visceral-to-Subcutaneous Abdominal Fat Ratio Is Associated with Nonalcoholic Fatty Liver Disease and Liver Fibrosis. Endocrinology and Metabolism 2020;**35**(1):165-76
39. Kim SR, Yoo JH, Song HC, et al. Relationship of visceral and subcutaneous adiposity with renal function in people with type 2 diabetes mellitus. Nephrology Dialysis Transplantation 2011;**26**(11):3550-55
40. Wang Y, Chen F, Wang J, et al. The Relationship Between Increased Ratio of Visceral-to-Subcutaneous Fat Area and Renal Outcome in Chinese Adults With Type 2 Diabetes and Diabetic Kidney Disease. Canadian journal of diabetes 2019;**43**(6):415-20
41. Carr A. HIV lipodystrophy: risk factors, pathogenesis, diagnosis and management. Aids 2003;**17**:S141-S48
42. Rockall A, Sohaib S, Evans D, et al. Computed tomography assessment of fat distribution in male and female patients with Cushing's syndrome. European journal of endocrinology 2003;**149**(6):561-68

Acknowledgements

This work was supported by the Sarnoff Cardiovascular Research Foundation (MM), NIH NCATS UL1TR001878, NIH/NHLBI R01 HL137984, R01 AA026302-02, P30 DK0503060 (RC), and the Penn Center for Precision Medicine.

Table 1. Population characteristics for cohort

Characteristic	Cohort (n=13,422)	
Demographics	*mean*	*std. dev.*
Age (years)	57.9	15.2
Sex	*n*	*%*
Male	6,747	50.3
Female	6,662	49.7
Ancestry		
European	8,410	67.2
African	3,490	27.9
Asian	212	1.7
Other/Unknown	1310	9.8
Clinical Metrics	*mean*	*std. dev.*
Body mass index, kg/m^2	28.8	7.1
Systolic Blood Pressure, mmHg	126.5	12.0
Diastolic Blood Pressure, mmHg	74.8	7.3
Diagnoses	*N*	*%*
Hypertension		
Yes	7,106	63.3
No	5,415	36.7
Diabetes		
Yes	3,649	35.0
No	8,873	65.0
Heart Failure		
Yes	2,496	22.6
No	10,026	77.4
Ischemic Heart Disease		
Yes	3,110	27.1
No	9,412	72.9
Renal Failure		
Yes	3,673	35.5
No	8,849	64.5
Lab Values	*mean*	*std. dev.*
Cholesterol, mg/dL	176.0	39.1
HDL cholesterol, mg/dL	50.8	15.5
LDL cholesterol, mg/dL	98.6	32.2
Triglycerides, mg/dL	124.8	64.2
HbA1c, %	6.4	2.5
BUN, mg/dL	17.9	9.8
Creatinine, mg/dL	1.0	0.4

Figure Legends

Figure 1: Detailed description of the Penn Medicine Biobank participants included in this study and exclusion criteria applied at each step in analysis. An imaging study was defined as a single visit and one or more CT scans of the abdomen or pelvis. An imaging scan (or series) was defined as a set of CT images with or without contrast; multiple scans were performed per imaging study. CNN_1 is the classification deep learning network used to identify axial CT views that show the abdomen. CNN_2 is the segmentation network which delineates the abdominal contour. After applying exclusion criteria, association studies were performed using body mass index, blood biomarkers and diagnostic codes on 13,422 patients. If a patient had multiple studies or scans, the median value across studies or scans was reported.

Figure 2: Automated extraction of abdominal fat from CT scans. A classification network (CNN_1) classified 2D image slices as showing views of the abdomen or not. In a second step, a segmentation network (CNN_2) delineated the border of the inner abdominal contour. Fat voxels were identified based on CT attenuation and those within the contour represent visceral and those outside represent subcutaneous fat. CT=Computed Tomography, CNN=Convolutional Neural Network.

Figure 3: Representative segmentation results from five different patients. First column shows CT slice with yellow line indicating the CNN predicted abdominal contour. From this contour, our algorithm identifies subcutaneous fat (second column) and visceral fat (third column).

Figure 4: Scatterplots and Bland-Altman plots comparing CNN to manually derived area for subcutaneous (A, B) and visceral (C, D) fat where fat was derived from a single slice on 100 randomly selected scans. Panels E and F show scatterplots of CNN derived area versus BMI for subcutaneous (E) and visceral (F) fat where fat was derived from all available biobank scans. CNN=Convolutional Neural Network, ICC=Interclass Correlation Coefficient, BMI=Body Mass Index.

Figure 5: Biomarker distributions dichotomized by VSR values. HDL = High density lipoprotein, LDL = low density lipoprotein, A1C = glycated hemoglobin, AST = aspartate aminotransferase, ALT = alanine aminotransferase, BUN = blood urea nitrogen, BMI = body mass index.

Figure 6: Phenome-Wide Association Study of VSR. Blue line indicates level of significance with Bonferroni multiple-comparison correction. Upward-facing triangles indicate a positive association and downward-facing triangles indicates a negative association.

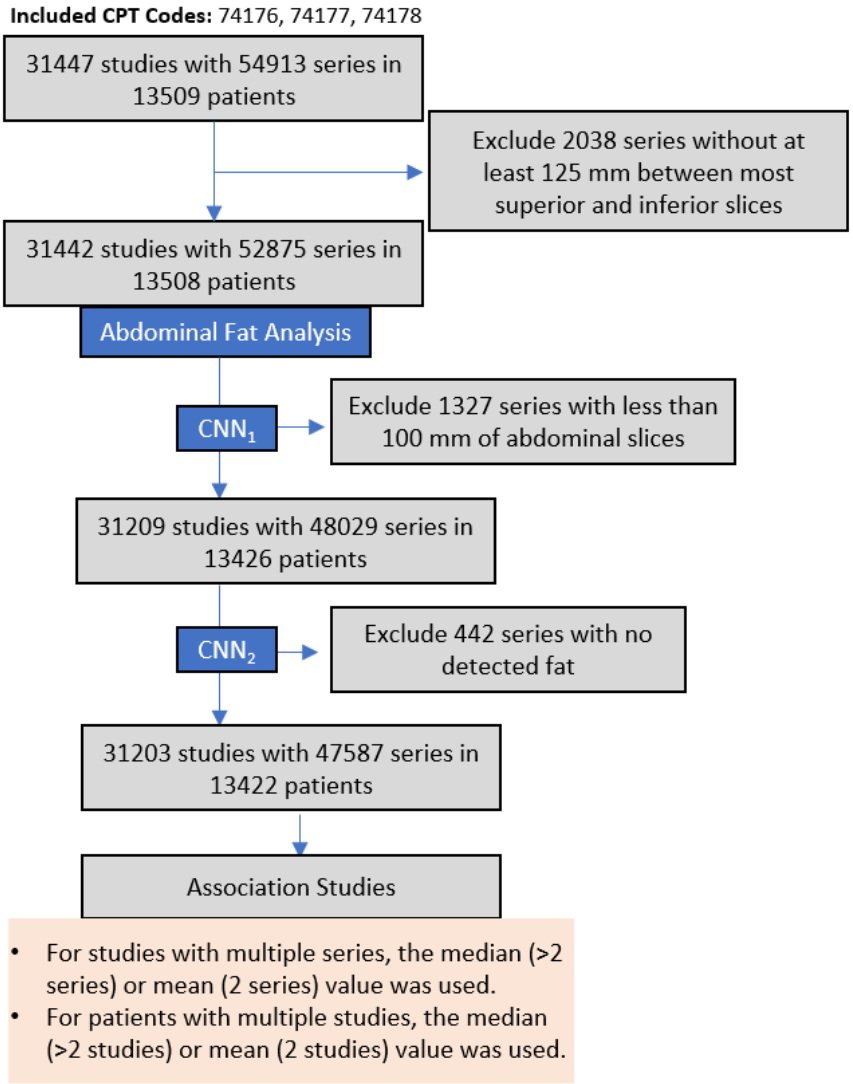


Figure 1: Detailed description of the Penn Medicine Biobank participants included in this study and exclusion criteria applied at each step in analysis. An imaging study was defined as a single visit and one or more CT scans of the abdomen or pelvis. An imaging scan (or series) was defined as a set of CT images with or without contrast; multiple scans were performed per imaging study. CNN_1 is the classification deep learning network used to identify axial CT views that show the abdomen. CNN_2 is the segmentation network which delineates the abdominal contour. After applying exclusion criteria, association studies were performed using body mass index, blood biomarkers and diagnostic codes on 13,422 patients. If a patient had multiple studies or scans, the median value across studies or scans was reported.

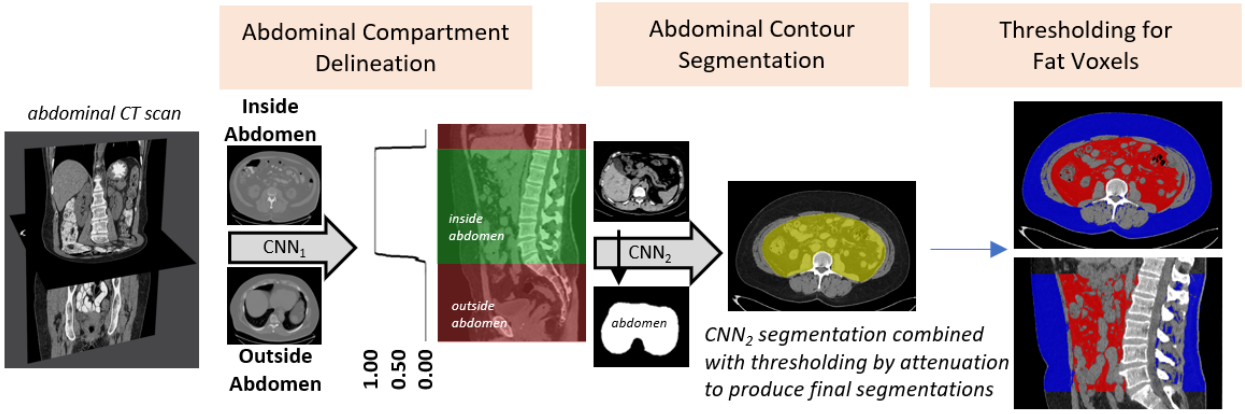


Figure 2: Automated extraction of abdominal fat from CT scans. A classification network (CNN_1) classified 2D image slices as showing views of the abdomen or not. In a second step, a segmentation network (CNN_2) delineated the border of the inner abdominal contour. The segmentation network was not trained to directly label subcutaneous and visceral fat. Rather, the segmentation consisted of two steps, where 1) the abdominal compartment was segmented (yellow) and 2) fat voxels were located by thresholding for HU between -190 and -30 (red and blue). Fat voxels within the contour represent visceral and those outside represent subcutaneous fat. CT=Computed Tomography, CNN=Convolutional Neural Network.

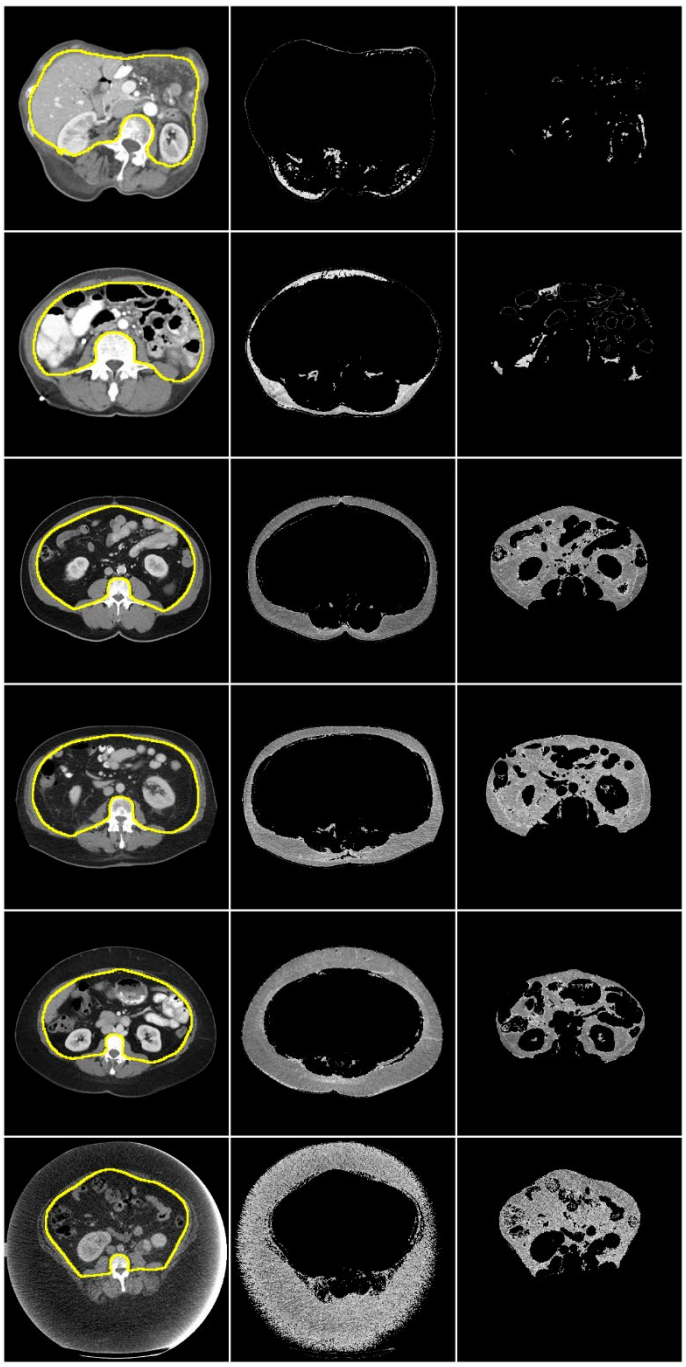


Figure 3: Representative segmentation results from five different patients. First column shows CT slice with yellow line indicating the CNN predicted abdominal contour. From this contour, our algorithm identifies subcutaneous fat (second column) and visceral fat (third column).

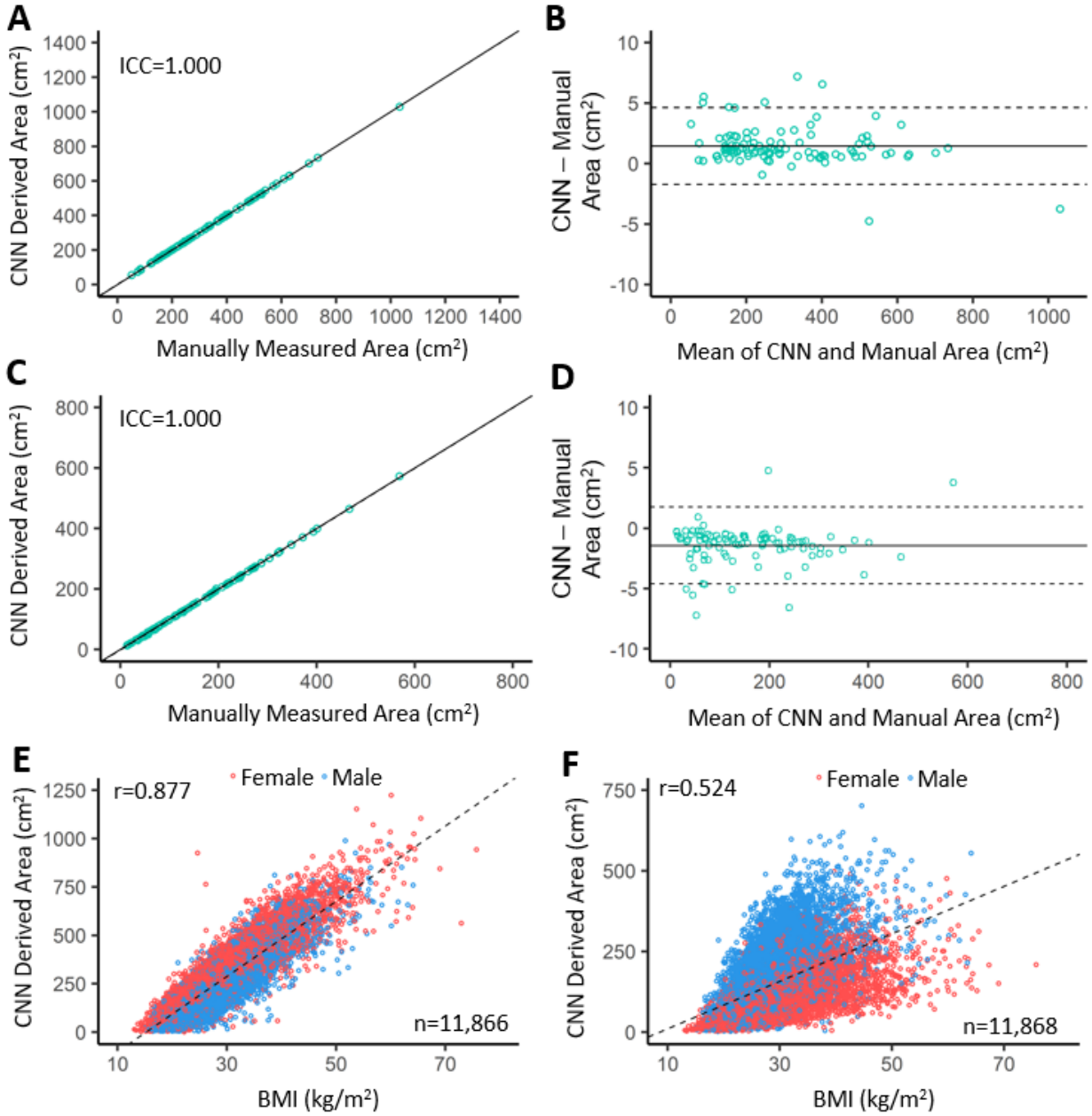


Figure 4: Scatterplots and Bland-Altman plots comparing CNN to manually derived area for subcutaneous (A, B) and visceral (C, D) fat where fat was derived from a single slice on 100 randomly selected scans. Panels E and F show scatterplots of CNN derived area versus BMI for subcutaneous (E) and visceral (F) fat where fat was derived from all available biobank scans. CNN=Convolutional Neural Network, ICC=Interclass Correlation Coefficient, BMI=Body Mass Index.

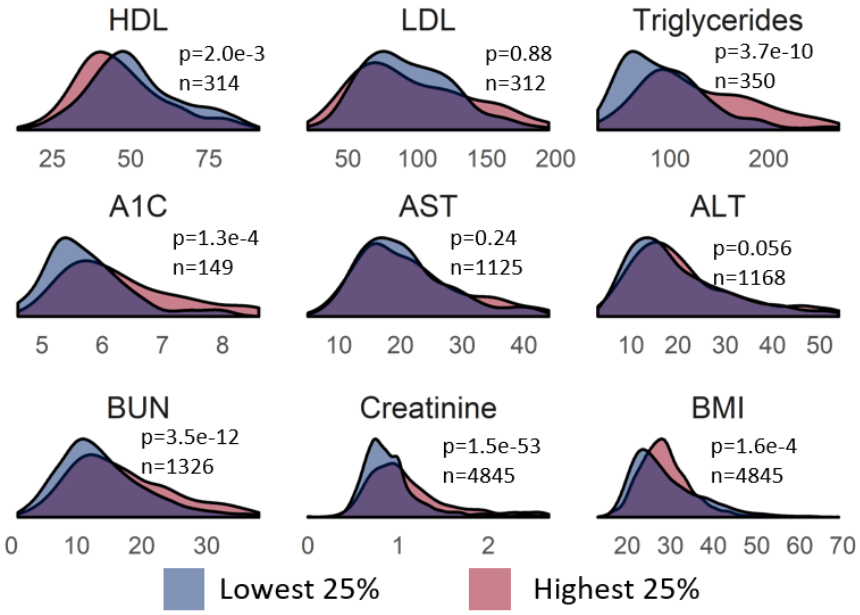


Figure 5: Biomarker distributions dichotomized by VSR values. HDL = High density lipoprotein, LDL = low density lipoprotein, A1C = glycated hemoglobin, AST = aspartate aminotransferase, ALT = alanine aminotransferase, BUN = blood urea nitrogen, BMI = body mass index.

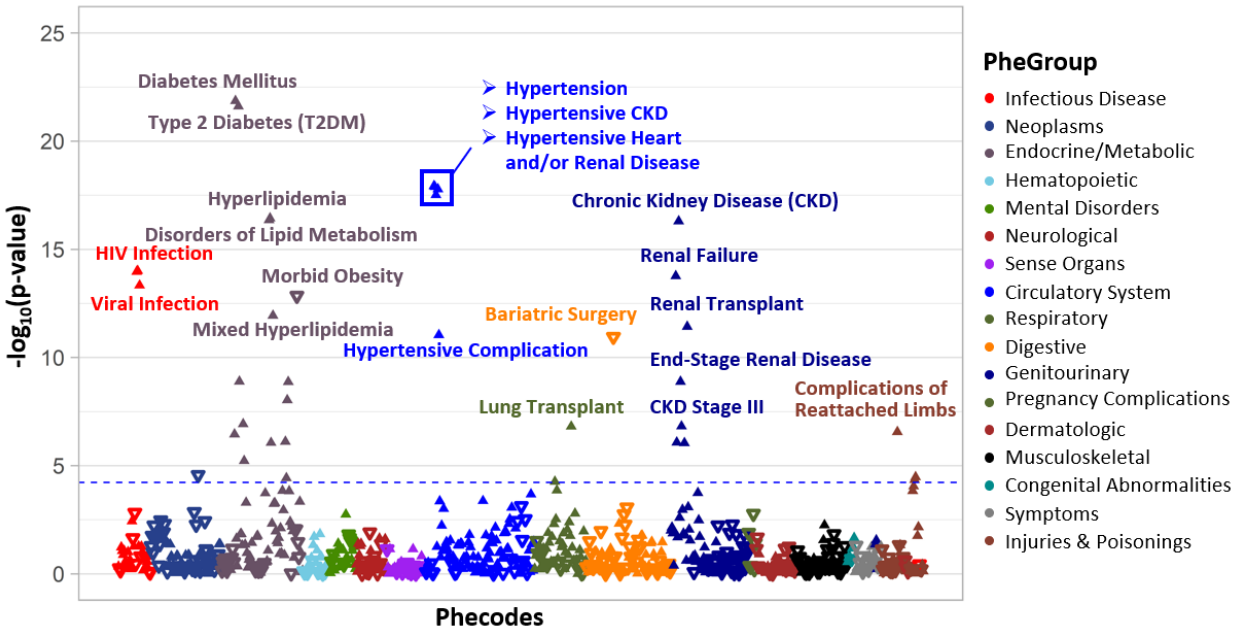
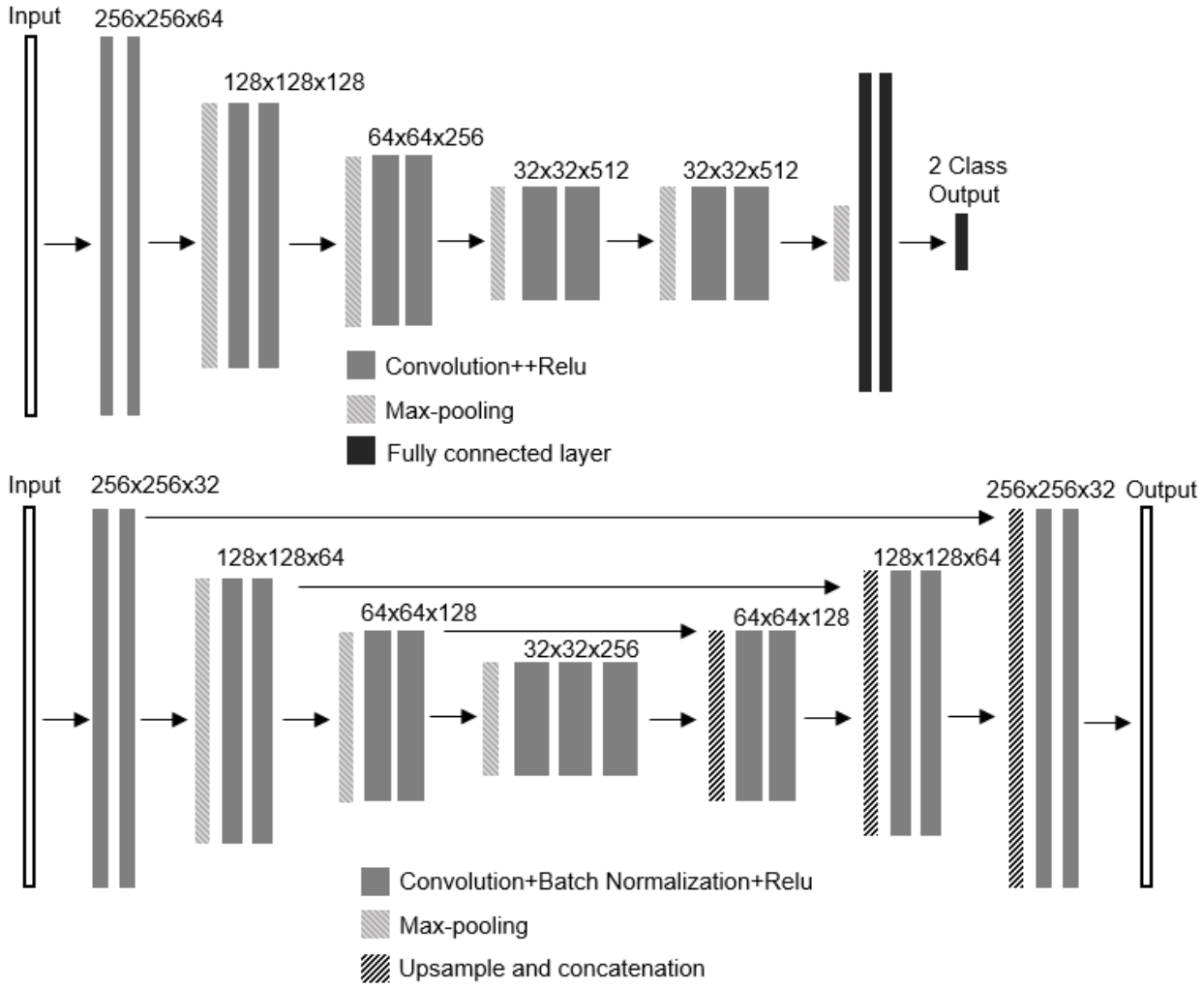


Figure 6: Phenome-Wide Association Study of VSR. Blue line indicates level of significance with Bonferroni multiple-comparison correction. Upward-facing triangles indicate a positive association and downward-facing triangles indicates a negative association.



Supplementary Figure 1: Convolutional neural networks used for A, classifying abdominal slices and B, pixel-level labeling of abdominal contour (segmentation). The neural network in A is modelled after VGG-16 and shows 5 stages of CNNs paired with two dense layers and a 2 - category output layer. The CNN in B is modelled after U-Net with 21 total layers in contracting and expanding paths with skip connections between layers of the same size. Localized feature information from the contracting path is combined with contextual information from the expanding path. The final output layer is a probability map for each pixel belonging to the quantitative imaging trait.