**Article 25fa pilot End User Agreement**

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) ‘Article 25fa implementation’ pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication other than authorised under this licence or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library
Radboud University
Copyright Information Point
PO Box 9100
6500 HA Nijmegen

You will be contacted as soon as possible.
Automatic Coronary Calcium Scoring in Low-Dose Chest Computed Tomography

Ivana Išgum*, Mathias Prokop, Meindert Niemeijer, Max A. Viergever, and Bram van Ginneken

Abstract—The calcium burden as estimated from non-ECG-synchronized computed tomography (CT) exams acquired in screening of heavy smokers has been shown to be a strong predictor of cardiovascular events. We present a method for automatic coronary calcium scoring with low-dose, non-contrast-enhanced, non-ECG-synchronized chest CT.

First, a probabilistic coronary calcium map was created using multi-atlas segmentation. This map assigned an a priori probability for the presence of coronary calcifications at every location in a scan. Subsequently, a statistical pattern recognition system was designed to identify coronary calcifications by texture, size, and spatial features; the spatial features were computed using the coronary calcium map. The detected calcifications were quantified in terms of volume and Agatston score.

The best results were obtained by merging the results of three different supervised classification systems, namely direct classification with a nearest neighbor classifier, and two-stage classification with nearest neighbor and support vector machine classifiers. We used a total of 231 test scans containing 45 674 mm$^3$ of coronary calcifications. The presented method detected on average 157/198 mm$^3$ (sensitivity 79.2%) of coronary calcium volume with on average 4 mm$^3$ false positive volume.

Calcium scoring can be performed automatically in low-dose, noncontrast enhanced, non-ECG-synchronized chest CT in screening of heavy smokers to identify subjects who might benefit from preventive treatment.

Index Terms—Automatic coronary calcium scoring, cardiovascular risk assessment, chest computed tomography (CT), lung cancer screening.

I. INTRODUCTION

ATHEROSCLEROSIS is the leading cause of death in the developed countries and it is also rising in the developing world. One manifestation of atherosclerosis is appearance of calcified plaques in the arterial wall. Arterial calcifications can be visualized and quantified with computed tomography (CT) and their amount is expressed in terms of a calcium score (Agatston, volume, or mass). Based on the Agatston coronary calcium score, a risk of a cardiovascular disease is determined, and the score has been shown to be an independent and strong predictor of cardiovascular events [1]–[6]. Such scoring is usually performed in ECG-synchronized cardiac CT scans without contrast enhancement. However, coronary calcification can be quantified with any CT scan that visualizes the heart, for example with chest CT scans that are obtained in screening of heavy smokers. It has been shown that participants in screening trials of heavy smokers are affected by fatal and nonfatal cardiovascular events as much as by lung cancer [4], [7]–[10].

In such a screening of heavy smokers, instead of only detecting lung cancer as a single disease, both lung cancer and atherosclerosis could be detected at an early stage [11]. This might especially be interesting when the screening images would allow automatic detection of cardiovascular disease. However, CT scans made in the setting of screening of heavy smokers are not acquired with protocols dedicated for calcium scoring. They are instead acquired with low radiation dose and without ECG-synchronization.

Fig. 1 shows examples of coronary calcifications in low-dose non-ECG synchronized chest scans from the Dutch–Belgian lung cancer screening trial (NELSON) [12]. The examples demonstrate that coronary calcifications in chest CT are strongly affected by cardiac motion. To illustrate the difference between scoring in cardiac and chest CT scans, Fig. 2 shows examples of the same coronary calcifications in retrospectively gated cardiac CT and in low-dose non-ECG-synchronized chest CT from the NELSON trial in two subjects. Image noise and cardiac motion may make precise quantification of coronary calcium in chest CT practically impossible [13]. Nevertheless, studies of Wu et al. [14], Budoff et al. [15], and Kim et al. [16] investigated the relation between
coronary calcium scores obtained in ungated low-dose chest CT and ECG-gated cardiac CT. All studies showed that low-dose non-ECG-synchronized chest CT scans, as acquired in screenings of heavy smokers allow quantification of coronary calcium scores that correlates well with Agatston scores \((r = 0.96\) in [15] and \(r = 0.89\) in [16]). Moreover, the study of Jacobs et al. [4] showed that coronary calcium scores obtained in scans acquired in the NELSON trial are an independent predictor of all-cause mortality and cardiovascular events.

In clinical practice calcifications are scored manually. Commercial software packages show areas of high intensity, and a human operator needs to manually identify each area that represents calcification. A threshold of 130 Hounsfield units (HU) is commonly used. Calcium scores are then automatically computed by the software. Some vendors have started providing multi-atlas based segmentation and it provides an automatic probabilistic coronary calcium map. This map has been created with the coronary calcium scoring system is designed which employs size, texture, and spatial information, were crucial for accurate coronary calcification identification. The disadvantage of this approach is that it requires aortic and cardiac segmentation. Kurkure et al. [21] detected coronary calcifications with a classification system employing a heart-centered coordinate system which was used to extract spatial features. Brunner et al. [22] detected zones of the coronary arteries using the coordinate system described in [21], and subsequently utilized an equidistant division of the coronary artery zones. Calcifications were then detected by a classification that employed spatial features derived from these coronary artery zones and regions.

In contrast to previously presented methods for coronary calcium scoring, this work presents a system for automatic coronary calcium scoring in low-dose, noncontrast-enhanced, non-ECG-synchronized chest CT scans. The key contribution of the proposed algorithm is that the approximate positions of the coronary calcifications are inferred using a probabilistic coronary calcium map. This map has been created with multi-atlas based segmentation and it provides an a priori probability for appearance of coronary calcifications at every position in the scan. Subsequently, a statistical pattern recognition system is designed which employs size, texture, and spatial features to detect coronary calcifications. The spatial features are computed from the coronary calcium map. The method has been validated in a screening cohort of heavy smokers. A preliminary version of this system is described in [27].

The paper is organized as follows. Section II describes the probabilistic coronary calcium map and Section III a pattern recognition system for calcification detection. Section IV describes the data used for evaluation. Experiments and results are provided in Section V. Discussion and conclusion are presented in Sections VI and VII.
II. CORONARY CALCIUM MAP

In noncontrast enhanced scans, coronary arteries are not visible unless they are embedded in fat or calcified. Consequently, automatic segmentation of coronary arteries in these scans is not a feasible first step for detecting coronary calcifications. Yet, earlier work [19], [21], [22], has shown that the spatial position of candidate calcifications is an important feature in determining whether the candidate calcification is truly a calcification. This can be done by locating the coronary calcifications in the scan relative to other anatomical structures, e.g., heart or aorta [19], [21]. A more promising approach utilizes the property that coronary calcifications appear at typical locations in the scan. They are most often found in the main branches of the coronary arteries (left main, left anterior descending, left circumflex, and right coronary arteries), frequently in their distal parts, and less often proximally. Therefore, using examples of such calcifications, a statistical map describing their (typical) locations and their variations can be designed.

Therefore, the coronary calcium map was directly created from the scans using multi-atlas based segmentation [28]. $M$ scans $A_i$, where $i = 1, \ldots, M$, were selected and registered to a single atlas image $U$. In other words, a spatial correspondence between each image $A_i$ and image $U$ was determined, so that the similarity between the images is maximized with respect to the transformation $u$

$$u = \arg \min_u C[u; U(p), A_i(p)]$$  \hspace{1cm} (1)

where $u$ is the optimal transformation making $A_i(u(p))$ spatially aligned to $U(p)$, $p = (x, y, z)$ denotes a voxel in the image, and $C$ is an appropriate cost function.

An affine transformation was performed to achieve coarse alignment between the images, followed by elastic registration to achieve accurate alignment. For the optimization of the cost function (negative mutual information) an iterative stochastic gradient descent optimizer was used [29]. To avoid local minima in the cost function, a multi-resolution Gaussian pyramid was employed, using a sub-sampling factor of two in each dimension. Also, a multi-grid approach was used for the nonrigid registration. For the affine registration five resolutions were used, in each of which 512 iterations of the stochastic gradient descent optimizer were performed. The derivative of the mutual information was calculated based on 4096 image samples, randomly chosen at every iteration. For the nonrigid B-spline registration, four resolutions were used. The B-spline grid spacing used in these resolutions was 64, 32, 16, and 8 voxels, respectively. The optimizer performed 512 iterations at each resolution. To estimate the derivative of the mutual information, 4096 image samples were used that were randomly chosen at every iteration. For both affine and nonrigid registration 32 histogram bins were used. Registrations were performed using elastix (http://elastix.isi.uu.nl) [30]. A detailed description of the registration approach and parameter settings is provided in [31].

Let $S_{A_i}$, $i = 1, \ldots, M$ be the manual (binary) segmentation of coronary calcifications in the scan $A_i$, for all $i$. The probabilistic coronary calcium map $S_p$ was obtained by averaging the transformed binary segmentations of coronary calcifications $S_i(u)$

$$S_p(p) = \frac{1}{M} \sum_{i=1}^{M} S_i(u_i(p)).$$  \hspace{1cm} (2)

Subsequently, the coronary calcium map was blurred to yield a probabilistic segmentation of the coronary calcifications in the atlas image

$$S_{\sigma}(p) = S_p(p) * g_{\sigma}$$  \hspace{1cm} (3)

where $g_{\sigma}$ is Gaussian blurring with a kernel width of three voxels. Fig. 3 shows three sections from this map. The map provides an a priori probability for spatial appearance of coronary calcifications at every location in the image.

III. PATTERN RECOGNITION SYSTEM

A statistical pattern recognition system was designed which identified all potential calcifications in a scan, and subsequently determined their texture, size, and spatial characteristics. Spatial characteristics were computed using the coronary calcium map. Based on these features true coronary calcifications were identified using supervised classification.
A. Candidate Extraction

As in clinical practice, potential calcifications (candidates) were extracted by thresholding (130 HU) and 3-D component labeling [4]. The candidates included coronary calcifications, aortic calcifications, other types of calcifications (e.g., aortic valve calcifications, calcifications around the trachea), bony structures (e.g., ribs, sternum, spine), and noise.

B. Feature Computation

Each candidate was described using size, texture, and spatial features.

1) Size Feature: As size feature, the volume of each candidate was calculated.

2) Texture Features: To obtain texture features, the following Gaussian filters were computed.

Let $I$ be an image with $N$ extracted candidates for coronary calcifications $C_i$, $i = 1, \ldots, N$. For each center of gravity $c_i$ of a candidate $C_i$ the following features were calculated:

\[
L_k(c_i) = I(c_i) * g_{\sigma_k}(c_i)
\]

\[
L_{(u,k)}(c_i) = I(c_i) * \frac{\partial g_{\sigma_k}(c_i)}{\partial u}
\]

\[
L_{(u,v,k)}(c_i) = I(c_i) * \frac{\partial^2 g_{\sigma_k}(c_i)}{\partial u \partial v}
\]

where $*g_{\sigma_k}$ denotes convolution with a Gaussian kernel of size $k$, where $k = 1, 2, 4, 8$ voxels, in the directions $u, v \in \{x, y, z\}$. In the scan acquisition protocol, the reconstructed field of view of each CT scan is set to include the outer rib margins of the subject. These individualized field-of-view settings are recommended for optimal CT coverage. Accordingly, when using number of voxels as the unit for kernel size, the kernel is automatically scaled with respect to the subject’s size. Therefore, the kernel size was expressed in voxels rather than in millimeters.

The Gaussian filters were implemented following the description provided in [19], [32]. A total of 40 texture features were computed.

3) Spatial Features: Spatial features were calculated using the coronary calcium map. First, each scan was registered to the calcium map using the same registration strategy as described in Section II. After the alignment the following features were computed.

Let $T$ be a test image, and $c_i$ the center of gravity for each candidate $C_i$, where $i = 1, \ldots, N$. The location of a candidate was transferred to the atlas coordinate system by the alignment, $T(u_i(c_i)), \forall i$. This provided three spatial features.

Next, minimum, average and maximum probability that a candidate is a coronary calcification were computed

\[
P_{\text{min}}(C_i) = \min_j \{S_\sigma(x_j)\}
\]

\[
P_{\text{mean}}(C_i) = \frac{1}{K_i} \sum_{j=1}^{K_i} S_\sigma(x_j)
\]

\[
P_{\text{max}}(C_i) = \max_j \{S_\sigma(x_j)\}
\]

TABLE I

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Candidate’s volume</td>
</tr>
<tr>
<td>2-4</td>
<td>Minimum, average and maximum distance of a candidate to coronary calcification in the calcium map</td>
</tr>
<tr>
<td>5-7</td>
<td>Minimum, average and maximum probability that a candidate is a coronary calcification</td>
</tr>
<tr>
<td>8-10</td>
<td>Candidate’s location in the atlas coordinate system</td>
</tr>
<tr>
<td>11-50</td>
<td>Gaussian filters up to the second order; scale=1, 2, 4, 8 voxels; $x, y, z$-direction</td>
</tr>
</tbody>
</table>

where $x_j \in C_i$ and $K_i$ is the number of voxels in $C_i$, $\forall i, j = 1, \ldots, K_i$.

Finally, minimum, average and maximum distance of a candidate to coronary calcification in the map was calculated

\[
D_{\text{min}}(C_i) = \min_j \{DT(S_B(x_j))\}
\]

\[
D_{\text{mean}}(C_i) = \frac{1}{K_i} \sum_{j=1}^{K_i} \{DT(S_B(x_j))\}
\]

\[
D_{\text{max}}(C_i) = \max_j \{DT(S_B(x_j))\}
\]

where $DT$ is the 3-D Euclidean distance transform of the binary segmentation $S_B$ ($S_B$ thresholded at 0), $x_j \in C_i$, and $K_i$ is the number of voxels in $C_i$, $\forall i, j = 1, \ldots, K_i$.

In total, a set of nine spatial features was calculated. The complete set of features is listed in Table I.

C. Classification

Before classification, scans were divided into training and test sets. Prior to feature selection and classification, all features were scaled to zero mean and unit variance. Four supervised classifiers and three different classification strategies were evaluated.

Classification was performed using linear discriminant (LDC), quadratic discriminant (QDC), k-nearest neighbor (k-NN) classifiers, and support vector machine (SVM) [33]. Performance of these classifiers was evaluated in a direct classification with a single classifier, in two-stage classification with two different classifiers, and in combinations thereof.

Influence of feature selection on the performance of LDC, QDC, and k-NN classifiers was evaluated. For feature selection, the training set was divided in a training set and a validation set. The training set contained 75%, and the validation set contained 25% of the randomly selected candidates. A floating forward feature selection (SFFS) strategy was used [34]. Note that samples in the classification were candidate lesions. However, as the final goal in this work was to assign a subject to the correct cardiovascular risk group, volume weighted accuracy, i.e., calcium volume was used as a performance criterion in feature selection. With the SVM classifier, a radial basis kernel (a Gaussian function) was used. Optimal settings were determined using a fivefold cross-validation grid-search on the training data to determine the optimal values for $C$ and $\gamma$ [35]. The drawback of this procedure is a long training time, which implied the number
of samples in the training set had to be reduced when doing experiments. Therefore, the performance of SVM was evaluated in two-stage classification approach. In the first stage the size of the training set was reduced with another classifier, and in the second stage SVM was employed.

The following classification strategies were evaluated:

1) Single Classifier Classification: Performance of LDC, QDC and k-NN classifiers was evaluated with and without feature selection. In each feature selection experiment, the maximum number of selected features was set to 50 (all features). The threshold on the posterior probability was set on 0.50. This means that samples which had posterior probability higher than 0.50 for being a negative class were classified as negative, and the remaining samples were classified as coronary calcification. The best setting for the number of neighbors k in the k-NN classifier was determined in pilot experiments where values of k equal to 1, 10, 30, 50, and 100 were evaluated.

2) Two-Stage Classification: In our previous work related to calcification detection, the best results were obtained in a two-stage classification approach [19], [25], [27]. The goal of the first stage was to reduce the number of candidates by discarding samples with a high probability for being in a negative class. In the second stage, a dedicated classifier was used to determine more difficult identifiable coronary calcifications. Considering that the k-NN classifier achieved the best performance in the single classifier classification, this classifier was used in the first classification stage of the current two-stage procedure. The threshold on the posterior probability was set to 0.80. This means that samples which had probability more than 0.80 for being in a negative class were discarded from further analysis, and the remaining samples were reclassified in the second classification stage.

In the second classification stage, performance of the SVM and k-NN classifiers was tested. As described, optimal settings for the SVM classifier were determined using a fivefold cross-validation grid-search on the training data to determine the optimal values for C and γ [35]. With the nearest neighbor classifiers, the number of neighbors k was determined in pilot experiments where odd values between 3 and 15 were evaluated. In feature selection, the maximum number of features to be selected was set to all features.

3) Combination of the Best Performing Classifiers: Combination of several different classifiers can boost classification performance [36]. We evaluated whether a combination of the several above designed classification strategies can improve coronary calcification detection. Performance of the evaluated classifiers/classification strategies was ranked by their ability to assign subjects to the correct cardiovascular risk group, thus by kappa value. Combinations of the best two, three, four, and five classification settings were evaluated. First, each of the employed classifiers independently computed a posterior probability for each candidate. Subsequently, those probabilities were averaged to obtain the final result. Candidates that had an average posterior probability for being a coronary calcification higher than or equal to 0.50 were classified as such.

Classification performance was expressed in terms of the classified candidates, as well as in terms of their volume. Sensitivity, average false positive rate per scan and average number of errors per scan were determined.

D. Importance of the Features Calculated Using the Coronary Calcium Map

After the best classification settings were found, the significance of the features calculated from the coronary calcium map was evaluated. The performance of the chosen system was tested by employing 1) all features, 2) only features calculated using the coronary calcium map, and 3) only features computed without the use of the map. Subsequently, to identify the most important features among those extracted from the coronary calcium map, the experiment employing only those features was performed with feature selection (SFFS with volume weighted accuracy as performance criterion). The number of features to be selected was set to the highest possible value, i.e., nine.

E. Determination of the Cardiovascular Risk

After the classification, the detected positives (coronary calcifications) were quantified by computing volume and Agatston scores [37], [38]. Each subject was assigned a cardiovascular risk category as outlined in Rumberger et al. [6]. Please note that CVD risk directly relating the obtained scores and CVD related morbidity and mortality has not been evaluated in this work. The categories are listed in Table II.

IV. MATERIALS

A. Data

Chest CT scans from a population-based randomized multi-center screening trial of heavy smokers were used in this study [12]. The study was approved by the Ethics Committee of each participating hospital. Baseline and short-term follow-up (mean: 3.1 ± 0.6 months) CT scan pairs of 584 consecutive subjects were selected for coronary calcium scoring (average age 59.9 ± 5.8 years).

Scans were acquired on a 16 detector-row scanner (Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH) in spiral mode with 16 × 0.75 mm collimation. Axial images of 1.0 mm thickness at 0.7 mm increment were reconstructed with a moderately soft kernel (Philips “B”), using the smallest field of view to include the outer rib margins at the widest dimension of the thorax. This resulted in a voxel sizes ranging from 0.55 to 0.87 mm. The peak voltage was 120–140 kVp depending on patient weight, with tube current 30 mAs. No intravenous contrast injection was applied. A detailed description of the inclusion criteria and the scanning protocol is provided by Xu et al. [39]. To reduce image noise and to perform calcium scoring with the usual section thickness, scans were subsampled to 3.1-mm-thick sections by averaging four consecutive 1-mm sections with a total increment of 1.4 mm [4], [40], [41].

<table>
<thead>
<tr>
<th>Category</th>
<th>Risk</th>
<th>Agatston score</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>very low</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>low</td>
<td>0-10</td>
</tr>
<tr>
<td>III</td>
<td>intermediate</td>
<td>10-100</td>
</tr>
<tr>
<td>IV</td>
<td>high</td>
<td>100-400</td>
</tr>
<tr>
<td>V</td>
<td>very high</td>
<td>&gt;400</td>
</tr>
</tbody>
</table>
Baseline scans were used for the development and evaluation of the automatic coronary calcium scoring method. From this set, scans containing metal implants (10 scans), extreme noise levels (four scans), an incomplete volume (one scan), or section thickness different than defined by the protocol (one scan) were removed from the data set. Thus, the set of baseline scans contained 568 CT scans.

B. Reference Standard

Two observers (medical doctors) with two and three years experience in cardiac CT manually identified coronary calcifications in all scans independently. Baseline and follow-up scan pairs were equally divided between the two observers. Thus, each scan was segmented once by one of them. Calcium scoring was performed using an in-house developed software package. Initially, all voxels above the threshold value of 130 HU were marked by a color overlay. Each coronary calcification was manually identified by clicking on one of its voxels. This was followed by 3-D region growing using 26-connectivity [42], and the identified calcification was indicated by a color change in the overlay. Total Agatston and volume scores were calculated for each subject following the algorithms described in [37], [38]. For each subject a cardiovascular risk category was determined [6]. Manual calcium scoring in these scans is a labor-intensive task, and therefore calcium scoring of the same scan by two observers would have been difficult to obtain in a large set of scans. Since only a small change in calcium score is expected between baseline and short-term follow-up imaging [40], the follow-up scans were used for estimation of the inter-scan variability. Coronary calcifications in the baseline and the follow-up scan pairs were identified by the same observer to eliminate the effect of the inter-observer variability. The time interval between reading the baseline and the follow-up scans was at least two weeks.

Finally, the performance of a second observer was estimated on a small set of images. For this purpose nine scans were selected from each cardiovascular risk category in a consecutive manner. The reference standard was set by the first observer. A medical student trained by the first author identified coronary calcifications in these images following the same protocol as the two observers. The first author (with calcium scoring experience in more than 1000 CT scans) reviewed the results and corrected when deemed necessary.

V. EXPERIMENTS AND RESULTS

From the set of 568 baseline scans, 237 were selected to create the coronary calcium map. The scans were selected in alphabetical (subject) order. In this set 37 scans did not contain any coronary calcifications. One of these scans was randomly selected as an atlas scan (U). The remaining 36 zero calcium score scans were removed from further analysis, because they could not contribute to the calcium map. The remaining 200 scans (A1) were elastically registered to the atlas.

After the coronary calcium map had been designed, 331 out of 568 baseline scans remained for training and testing the classification system. These scans were registered to the calcium map and afterwards divided in two sets. The first 100 scans were selected for training the classification system. The remaining 231 scans (T) were used for testing.

After coronary calcifications had been detected in the test scans, coronary calcium scores were computed. Based on the score, each subject was assigned to a cardiovascular risk category.

In total, the test set contained 664 (45,674 mm³) coronary calcifications, and 1,123,893 (22,171,662 mm³) negative candidates were extracted.

A. Single Classifier Classification

Single classifier classification with LDC and QDC resulted in moderate sensitivity and high false positive rate. These classifiers were not able to determine the cardiovascular risk of subjects well. The best results with nearest neighbor classifier were achieved when the number of neighbors was set to 100: 73% of subjects were assigned the correct risk category, 18% were one category off, and 9% were two or three categories off.

After feature selection, performance of all classifiers substantially improved. When employing LDC, a single feature, i.e., the maximum probability that a candidate is a coronary calcification, was selected. With QDC, 23 features were selected: 1 size, 7 features extracted from the coronary calcium map, and 15 texture features. Nearest neighbor classifier (100-NN) selected five features: 1 size, 3 features calculated using the coronary calcium map, and 1 texture feature. The selected features are listed in Table IV. Classification with QDC with selected features resulted in a high false positive volume (about 2300 mm³ on average of false positive volume per scan vs. 198 mm³ of average calculation volume per scan). Performance of LDC with feature selection was comparable to classification with 100-NN without feature selection. 100-NN with feature selection achieved the best performance in terms of the linearly weighted kappa value. It assigned 74% of the subjects to the correct risk category, 17% were one category off and 8% were two categories off. No scan was assigned to a risk group that was three or four categories off. Detailed results of the classification are listed in Table III.

B. Two-Stage Classification

In the first classification stage, the best classifier from single classifier classification, namely a 100-nearest neighbor classifier with five selected features, was employed. All potential calcifications which had more than 80% posterior probability of being a noncoronary-calcification were discarded from further analysis. In this way 1,123,614/1,123,893 (99.9%) negative candidates and 131/664 (19.7%) coronary calcifications were discarded from further analysis. In terms of volume, 22,165,931/22,171,662 (99.9%) negatives and 4,839/45,674 (10.6%) calcifications were discarded.

In the second classification stage, performance of k-NN and SVM were evaluated.

The best performance with the nearest neighbor classifier was achieved when k was set to 11. In the second stage, the 11-NN classifier selected 11 features: four features calculated using the coronary calcium map, and seven texture features. Features that were selected in either classification stage are listed in Table IV.
TABLE III
RESULTS OF DIRECT CLASSIFICATION WITH LDC, QDC, AND 100-NN, WITH FLOATING FORWARD FEATURE SELECTION. AVERAGE SENSITIVITY, NUMBER OF FALSE POSITIVES, AND ERRORS PER SCAN ARE LISTED. ACCURACY OF CARDIOVASCULAR RISK GROUP ASSIGNMENT WAS CALCULATED BASED ON THE AGATSTON SCORE AND IS EXPRESSED IN TERMS OF LINEARLY WEIGHTED KAPPA

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Candidates</th>
<th>Calcium volume (mm$^3$)</th>
<th>Risk group kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>FP/scan</td>
<td>Errors</td>
</tr>
<tr>
<td>LDC</td>
<td>0.67</td>
<td>22.7</td>
<td>23.7</td>
</tr>
<tr>
<td>LDC(SFFS)</td>
<td>0.51</td>
<td>8.8</td>
<td>2.2</td>
</tr>
<tr>
<td>QDC</td>
<td>0.70</td>
<td>11.6</td>
<td>12.5</td>
</tr>
<tr>
<td>QDC(SFFS)</td>
<td>0.80</td>
<td>2.3</td>
<td>3.1</td>
</tr>
<tr>
<td>100-NN</td>
<td>0.49</td>
<td>0.8</td>
<td>2.2</td>
</tr>
<tr>
<td>100-NN(SFFS)</td>
<td>0.58</td>
<td>0.9</td>
<td>2.1</td>
</tr>
</tbody>
</table>

TABLE IV
FEATURES SELECTED WITH A SINGLE CLASSIFIER (LDC, QDC, 100-NN) AND WITH 11-NN CLASSIFIER FROM THE SECOND CLASSIFICATION STAGE IN TWO-STAGE CLASSIFICATION. FEATURES ARE NUMBERED ACCORDING TO THE ANNOTATION IN TABLE I, AND LISTED IN THE ORDER OF SELECTION

<table>
<thead>
<tr>
<th>Stage</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDC</td>
<td>7</td>
</tr>
<tr>
<td>QDC</td>
<td>7, 3, 2, 8, 10, 46, 33, 44, 49, 47, 28, 6, 9, 1, 50, 41, 45, 38, 43, 22, 40, 48, 31</td>
</tr>
<tr>
<td>100-NN</td>
<td>1, 7, 4, 9, 31</td>
</tr>
<tr>
<td>11-NN</td>
<td>33, 27, 44, 8, 37, 36, 9, 10, 31, 32, 2</td>
</tr>
</tbody>
</table>

All candidates with a posterior probability larger than or equal to 0.5 for being coronary calcifications were classified as such.

In terms of the kappa value, SVM performed similarly to 100-NN with selected features, and k-NN in two stages achieved the best results. Based on the Agatston score, 81% of subjects were assigned the correct risk group, and 13%, 5%, and 1% were one, two, and three categories off, respectively. Detailed classification results in terms of candidates and their volumes are shown in Table V.

C. Classifier Combination

After ranking the tested classifiers by the achieved kappa value, combinations of the following classifiers were evaluated.

(i) Two-stage classification with nearest neighbor classifiers (kappa 0.83).

(ii) Two-stage classification with nearest neighbor and SVM classifiers (kappa 0.79).

(iii) Classification with 100-NN using selected features (kappa 0.78).

(iv) Classification with LDC using selected features (kappa 0.77).

(v) Classification with QDC using selected features (kappa 0.66).

This resulted in four classification systems (see Table VI).

Note that single classifier classification with a nearest neighbor classifier was not evaluated although it performed better than the QDC classifier with feature selection in terms of kappa value. This choice was made based on the error analysis of each classifier. QDC had a very low false negative rate and made different errors than the nearest neighbor classifier. The goal was to evaluate whether a combination with QDC might improve overall detection of calcifications.

In terms of kappa value, as well as in terms of classified volume, all combinations performed similarly. In terms of classified candidates, combinations of the best two and the best three classifiers achieved better results. Detailed results of the tested combinations are listed in Table VI.

Subsequently, for each tested combination, the threshold on the posterior probability giving the smallest error volume (sum of false positive and false negative volumes) was determined. If this threshold was different than 0.50, the performance was recomputed and the results are shown in Table VII.

The best results in terms of the kappa value were achieved by a combination of three classifiers, namely two-stage classification with nearest neighbor classifiers, two-stage classification with nearest neighbor classifier and SVM classifier, and single classifier classification with nearest neighbor classifier with selected features. In terms of calcified volume, the method correctly detected 79.2% (36,189/45,674) of coronary calcifications with on average 4 mm$^3$ of false positive volume per scan. The method assigned 190/231 (82.2%) subjects a correct cardiovascular risk category. Thirty-three (14.3%) and eight (3.5%) subjects were one and two categories off, respectively. No subject was assigned to a group that was more than two risk categories off. Table VIII shows this agreement. The agreement obtained from the volume scores resulted in a Spearman’s rho of 0.93. In 110/231 (48%) scans no false positive or false negative error was made. In the remaining scans, false negatives were calcifications in LAD in 35% of the cases, in LCX in 44% of the cases. Most often these were small calcifications of intensity just above the threshold value for calcifications; frequently, they resembled noise. Most of the false negatives in LAD were located in the vessel running next to the pericardium, as shown in Fig. 4(a). In RCA the errors were in the area strongly affected by cardiac motion, as illustrated in Fig. 4(c). Although sensitivity of the system in terms of classified candidates was relatively low (58.6%), sensitivity in terms of classified volume was much better (79.2%). This indicates that false negative lesions were small in volume. Accuracy of cardiovascular risk assignment based on the Agatston score is high (kappa of 86.7%). This suggests that false negatives were also of low intensity, in agreement with visual analysis of the errors.

False positives were image noise in the area of the coronary arteries in 66% of the cases and calcifications in the ascending aorta in 9%. Retrospective analysis showed that in 22% of the cases false positives were likely coronary calcifications not marked in the reference standard. The remaining one (3%) false positive was a part of spine. Analysis showed that this error was not caused by incorrect alignment between the atlas and this test scan, but probably by texture and volume characteristics of
TABLE V
RESULTS OF THE TWO STAGE CLASSIFICATION STRATEGY. IN THE FIRST STAGE 100-NN WITH FEATURE SELECTION WAS USED. IN THE SECOND STAGE SVM AND 11-NN CLASSIFIERS WERE APPLIED. AVERAGE SENSITIVITY, NUMBER OF FALSE POSITIVES, AND ERRORS PER SCAN ARE LISTED. ACCURACY OF CARDIOVASCULAR RISK GROUP ASSIGNMENT WAS CALCULATED BASED ON THE AGATSTON SCORE AND IS EXPRESSED IN TERMS OF LINEARLY WEIGHTED KAPPA

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Candidates</th>
<th>Volume (mm³)</th>
<th>Risk group kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>100-NN; SVM</td>
<td>0.73, 0.8, 1.6</td>
<td>0.88, 38, 62</td>
<td>0.79</td>
</tr>
<tr>
<td>100-NN; 11-NN</td>
<td>0.76, 0.9, 1.6</td>
<td>0.87, 24, 49</td>
<td>0.83</td>
</tr>
</tbody>
</table>

TABLE VI
RESULTS OF THE CLASSIFICATION COMBINATIONS WITH (i) TWO-STAGE CLASSIFICATION WITH NEAREST NEIGHBOR CLASSIFIERS, (ii) TWO-STAGE CLASSIFICATION WITH NEAREST NEIGHBOR AND SVM CLASSIFIERS, (iii) 100-NN USING SELECTED FEATURES, (iv) LDC USING SELECTED FEATURES, AND (v) QDC USING SELECTED FEATURES. AVERAGE SENSITIVITY, NUMBER OF FALSE POSITIVES, AND ERRORS PER SCAN ARE LISTED. ACCURACY OF CARDIOVASCULAR RISK GROUP ASSIGNMENT WAS CALCULATED BASED ON THE AGATSTON SCORE AND IS EXPRESSED IN TERMS OF LINEARLY WEIGHTED KAPPA

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Candidates</th>
<th>Volume (mm³)</th>
<th>Risk group kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i),(ii)</td>
<td>0.76, 0.8, 1.5</td>
<td>0.88, 37, 60</td>
<td>0.81</td>
</tr>
<tr>
<td>(i),(ii),(iii), (iv)</td>
<td>0.76, 0.8, 1.5</td>
<td>0.88, 34, 57</td>
<td>0.82</td>
</tr>
<tr>
<td>(i),(ii),(iii),(iv),(v)</td>
<td>0.69, 0.7, 1.6</td>
<td>0.86, 24, 52</td>
<td>0.83</td>
</tr>
<tr>
<td>(i),(ii),(iii),(iv),(v)</td>
<td>0.73, 1.0, 1.8</td>
<td>0.87, 26, 51</td>
<td>0.82</td>
</tr>
</tbody>
</table>

TABLE VII
PERFORMANCE OF THE CLASSIFIER COMBINATION SYSTEMS WHEN MINIMUM VOLUME ERROR IS ACHIEVED. COMBINATIONS USING (i) TWO-STAGE CLASSIFICATION WITH NEAREST NEIGHBOR CLASSIFIERS, (ii) TWO-STAGE CLASSIFICATION WITH NEAREST NEIGHBOR AND SVM CLASSIFIERS, AND (iii) 100-NN USING SELECTED FEATURES ARE SHOWN. THRESHOLD ON POSTERIOR PROBABILITY, AVERAGE SENSITIVITY, NUMBER OF FALSE POSITIVES, AND ERRORS PER SCAN ARE LISTED. ACCURACY OF CARDIOVASCULAR RISK GROUP ASSIGNMENT WAS CALCULATED BASED ON THE AGATSTON SCORE AND IS EXPRESSED IN TERMS OF LINEARLY WEIGHTED KAPPA

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Threshold</th>
<th>Candidates</th>
<th>Volume (mm³)</th>
<th>Risk group kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i),(ii)</td>
<td>0.30</td>
<td>0.59, 0.1</td>
<td>1.3</td>
<td>0.79, 4, 46</td>
</tr>
<tr>
<td>(i),(ii)</td>
<td>0.35</td>
<td>0.68, 0.5</td>
<td>1.4</td>
<td>0.85, 13, 42</td>
</tr>
</tbody>
</table>

TABLE VIII
RISK CATEGORY AGREEMENT IN 231 SCANS BETWEEN THE REFERENCE (REF) IN ROWS AND THE AUTOMATIC SYSTEM (AUTO) IN COLUMNS. A CORRECT RISK CATEGORY WAS ASSIGNED TO 190/231 (82.2%) SCANS, THIRTY-THREE SCANS (14.3%) WERE ONE CATEGORY OFF, EIGHT SCANS (3.5%) WERE TWO CATEGORIES OFF. NO SCAN WAS THREE OR FOUR CATEGORIES OFF. THE LINEARLY WEIGHTED KAPPA SCORE WAS 0.87

<table>
<thead>
<tr>
<th>Ref</th>
<th>Auto</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>84</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>13</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>III</td>
<td>5</td>
<td>4</td>
<td>36</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IV</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>26</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>27</td>
<td>0</td>
</tr>
</tbody>
</table>

the candidate. This false positive was detected by all of the three combined classifiers.

D. Importance of the Features Calculated Using the Coronary Calcium Map

The importance of the features extracted from the coronary calcium map was assessed. The automatic system employing only features extracted from the coronary calcium map assigned 163/231 (70.6%) of scans a correct risk category. When feature selection was used, slightly worse results were obtained: 161/231 (69.7%) subjects were assigned the correct risk category. The system using only features that were calculated without using the coronary calcium map (size and texture features) correctly detected the risk category in only 111/231 (48.0%) subjects. Spearman’s rho between the automatic system employing only features obtained from the coronary calcium map and the reference standard was 0.84, and the linearly weighted kappa score was 0.76. When feature selection was employed, these values were 0.62 and 0.74, respectively. When only size and texture features were used, Spearman’s rho was 0.51, and the linearly weighted kappa score was 0.31. These results are listed in Table IX. For comparison, the results of the best classifier combination with feature selection, already presented in Table VI, are given in this table as well. Features computed using the coronary calcium map that were selected with the SFFS scheme are listed in Table X.

E. Additional Evaluation

To evaluate automatic coronary calcium scoring, the performance of the automatic system was compared with the inter-scan agreement between the baseline and the short-term follow-up scans. This was possible because only limited progression of coronary calcium is expected in the given time interval [40], [43]. Therefore, the agreement between the volume calcium scores in the baseline (reference) and follow-up scans was evaluated.

The same risk category was assigned to 155/231 (67.1%) by both scans. The linearly weighted kappa was 0.76. The agreement in volume scores resulted in a Spearman’s rho between the reference and follow-up scans of 0.91 (Table XI).

In addition, to compare the results of the automatic system with the second human operator, the performance of the second observer was compared to the reference standard in a subset of 45 scans. To make a fair comparison, the performance of the automatic system was evaluated on the same set of scans. The correct risk category was assigned to 41/45 (accuracy 91.1%) subjects by the second observer, and to 35/45 (accuracy 77.8%) sub-
Fig. 4. Examples of errors generated by the automatic system. Typical examples of false negative lesions: Calcifications in (a) LAD, and (b) LCX are both small in volume and of low intensity. Calcification in (c) RCA is strongly affected by cardiac motion. Typical examples of false positive lesions: (d) Area of increased intensity, possibly calcification of very low intensity or blurred calcification due to cardiac motion, (e) motion artifact, and (f) calcification in the ascending aorta around the location where RCA branches off.

TABLE IX

<table>
<thead>
<tr>
<th>Features</th>
<th>accuracy</th>
<th>1-off</th>
<th>2-off</th>
<th>3-off</th>
<th>4-off</th>
<th>κ</th>
<th>ρ</th>
</tr>
</thead>
<tbody>
<tr>
<td>All features</td>
<td>0.75</td>
<td>0.17</td>
<td>0.06</td>
<td>0.01</td>
<td>0.01</td>
<td>0.78</td>
<td>0.85</td>
</tr>
<tr>
<td>All features with SFFS</td>
<td>0.81</td>
<td>0.12</td>
<td>0.06</td>
<td>0.01</td>
<td>0</td>
<td>0.82</td>
<td>0.94</td>
</tr>
<tr>
<td>Only features using coronary calcium map</td>
<td>0.71</td>
<td>0.22</td>
<td>0.06</td>
<td>0.01</td>
<td>0</td>
<td>0.76</td>
<td>0.84</td>
</tr>
<tr>
<td>Only features using coronary calcium map with SFFS</td>
<td>0.70</td>
<td>0.21</td>
<td>0.07</td>
<td>0.02</td>
<td>0</td>
<td>0.74</td>
<td>0.62</td>
</tr>
<tr>
<td>Only features without coronary calcium map</td>
<td>0.48</td>
<td>0.20</td>
<td>0.17</td>
<td>0.07</td>
<td>0.07</td>
<td>0.31</td>
<td>0.51</td>
</tr>
</tbody>
</table>

TABLE X

| Features selected in the experiment with the best classifier combination when only features calculated from the coronary calcium map were employed. Features were selected with 100-NN and 11-NN classifiers. The features are numbered according to the annotation in Table I, and listed in the order of selection |
| Stage | Feature |
| 100-NN | 3, 2, 4, 8 |
| 11-NN  | 7, 6 |

TABLE XI

Risk category agreement in 231 scans between the reference (Ref) in rows and the follow-up scan (FU) in columns. A correct risk category was assigned to 155/231 (67.1%) scans. Sixty-three scans (27.3%) were one category off, thirteen scans (5.6%) two categories off. Linearly weighted kappa score was 0.76

<table>
<thead>
<tr>
<th>Ref</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>73</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>III</td>
<td>3</td>
<td>8</td>
<td>28</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>IV</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>26</td>
</tr>
</tbody>
</table>

Objects by the automatic method. The achieved linearly weighted kappa between the observers was 0.93, and between the automatic system and the reference 0.86. Spearman’s rho between the volume scores of the observers was 0.95, between the automatic system and the reference 0.97. These results are listed in Tables XII and XIII. Fig. 5 illustrates examples of disagreement between the two observers.
VI. DISCUSSION

A method for coronary calcium scoring in chest CT has been presented. Coronary calcifications appear in CT scans as small bright lesions at very particular locations in a scan. Therefore, a coronary calcium map has been designed which provided an a priori probability for appearance of coronary calcifications at every location in a chest CT scan. Subsequently, a supervised classification was employed to identify coronary calcifications. Prior to feature computation and classification, a registration step was introduced to align a scan with the coronary calcium map. After this alignment (spatial) features of the potential coronary calcifications were computed using the designed coronary calcium map. In addition, texture and size features were calculated.

The created coronary calcium map presented a relatively large region with a high probability for the appearance of coronary calcifications. This is likely due to difficult alignment of coronary arteries in low-dose, noncontrast-enhanced scans owing to very low intensity variations within the heart. In addition, cardiac motion makes arteries and calcifications occasionally appear twice or at least blurry. Note that the features derived from the map proved important for the good performance of the system (see Tables IX and IV). The feature selection experiment with only features calculated from the coronary calcium map revealed that all features describing the distance from the calcifications in the map were of high importance for the 100-NN classifier, while also a candidate’s location in the atlas coordinate system was selected. With the 11-NN classifier, the average and maximum probabilities that a candidate is coronary calcium were the selected features. Future work could investigate whether a coronary calcium map could be created using ECG-synchronized CT where cardiac motion is minimized. This could lead to a more accurate coronary calcium map, and might subsequently lead to a reduction of errors of the automatic system.

The atlas created in this work was generated by registering a number of images to a single randomly chosen scan. This means that the atlas is biased towards that scan. Future work could evaluate if an atlas in which this bias would be reduced (using e.g., the approach described in [44]) could improve overall classification results.

In line with results from our previous publications [19], [25], classification with a nearest neighbor classifier outperformed LDC and QDC. With feature selection, LDC achieved lower sensitivity than QDC and 100-NN. Unlike LDC, QDC achieved high sensitivity both in terms of classified lesions and classified volume but it had a high false positive rate. Visual analysis of the errors showed that false positives were most often within the heart, and only in few cases they were outside it, mostly in the ascending aorta.

Two-stage classification achieved better results than single classifier classification. Although classification with SVM in the second stage performed only slightly better than single-stage
classification with a nearest neighbor classifier and feature selection in terms of kappa value, it achieved higher sensitivity both in terms of classified candidates, and in terms of classified volume. Better results were achieved using two-stage classification with k-NN classifiers. The first stage enabled discarding a large number of negative candidates that were easy to recognize. The subsequent classifier was trained to differentiate between more difficult distinguishable positives and negatives.

Cardiovascular risk was most accurately determined by averaging the results of the three best performing classifiers/classification strategies, namely two-stage classification with nearest neighbor classifiers, two-stage classification with nearest neighbor classifier and SVM classifier, and single classifier classification with a nearest neighbor classifier with selected features. Since the errors of the individual classification approaches (classifiers) were sometimes different, their combination reduced overall error rate. Visual inspection of the results showed that the most frequent false positives represented image noise. Noise removal prior to analysis, may reduce these false positives, but this carries the risk that removing noise may also remove small calcifications in the coronary arteries. The largest false positive errors in the terms of volume were caused by calcifications in the ascending aorta. Applying aortic segmentation and spatial features derived from it, as in [19], might lead to a reduction of such false positives.

The presented method generates more false negative than false positive errors both in number of lesions and in volume of errors (see Table VII). Consequently, the automatic system more often underestimates than overestimates cardiovascular risk (see Table VIII). In 35/231 (15.2%) subjects risk was underestimated and in 6/231 (2.6%) it was overestimated. In 4/6 scans where the risk was overestimated, the overestimation was caused by a single false positive lesion which was also the only misclassified candidate in the scan. Similarly, in 17/35 scans in which the risk category was underestimated, the underestimation was caused by a single false negative. In 16 of those 17 scans, risk was underestimated by one category. Note that coronary calculations which had a high probability for being non-coronary-calculations when classified using 100-NN classifier, were detected as false negatives by all three best performing classification approaches (due to employment of 100-NN in the first classification stage when two-stage classification was used). These calcifications were difficult to detect and were frequently also missed by LDC and QDC. Future work could investigate whether a different set of features is needed to recognize these calcifications.

It should be noted that an exhaustive search on experiment settings has not been performed. Also, only a simple combination of classifiers has been tried out. Future work could perform more detailed investigations in these directions and evaluate whether further fine tuning of parameters in individual classifiers and more advanced classifier combinations would further improve the performance.

The current study also evaluated whether automatic calcium scoring in a screening setting could determine cardiovascular risk of subjects based on the obtained Agatston score. The system correctly detecting cardiovascular risk group in 190/231 (82.2%) subjects (see Table VIII) best serves this purpose, according to kappa statistics. However, if such a system would be applied in screening programs, the exact settings achieving the desired sensitivity and specificity with respect to cost effectiveness would need to be determined.

Note that this study has not correlated automatically obtained Agatston scores and corresponding CVD risk categories with cardiovascular morbidity and mortality. Future work will relate these scores with cardiovascular events in the screening of heavy smokers.

Coronary calcium scores can not be reliably determined in low-dose and noncontrast enhanced scans, but previous work showed that yet these scans do enable estimation of cardiovascular risk that is able to predict cardiovascular events [4]. Our results show that automatic estimation of cardiovascular risk categories based on coronary calcium scores derived from screening of heavy smokers with CTs is more accurate than inter-scan agreement with short-term follow-up scans. The obtained inter-scan variability in Agatston score, and therefore in CVD risk, obtained in our study (Table XI) is in concordance with a previous study [40]. Agreement in CVD risk assignments between two observers is higher than between the automatic system and the reference. This indicates that the difference between the observers lies in lesions of lower intensity value than the differences between the automatic system and the reference. Retrospective analysis of the manual scorings did not reveal (typical) reasons for this disagreement. In the segmentation of the second observer, both false positive and false negative errors were present, and they did not occur at specific locations.

A recent publication by Rozanski et al. [45] showed that patients’ awareness of their coronary calcium score leads to important lifestyle changes and subsequently to a reduction of their cardiovascular risk. Given this observation and the recently increased interest in screening of heavy smokers, especially after the American Lung Association recommended CT lung cancer screening for smokers, the proposed method provides a fully automatic tool for identification and early treatment of asymptomatic individuals at risk, without additional radiation for the subjects and at minimum extra cost.

VII. CONCLUSION

We have presented a method for automatic coronary calcium scoring in low-dose, noncontrast enhanced, non-ECG-synchronized chest CT. First, a probabilistic coronary calcium map was created using a multi-atlas segmentation approach. Subsequently, supervised classification was employed for the detection of coronary calcifications based on size, texture, and spatial features. The spatial features were computed using the coronary calcium map. The method has been applied to low-dose, noncontrast enhanced and non-ECG-synchronized chest CT. The presented method detected on average 157.198 mm³ (sensitivity 79.2%) of coronary calcium volume with on average 4 mm³ false positive volume in subjects participating in a screening of heavy smokers.


