Computerized Medical Imaging and Graphics 97 (2022) 102049

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Automated segmentation of normal and diseased coronary arteries – The ASOCA challenge

Ramtin Gharleghi [a,*], Dona Adikari [b,c], Katy Ellenberger [b,c], Sze-Yuan Ooi [b,c], Chris Ellis [d], Chung-Ming Chen [e], Ruochen Gao [f], Yuting He [h], Raabid Hussain [g], Chia-Yen Lee [j], Jun Li [f], Jun Ma [k], Ziwei Nie [l], Bruno Oliveira [m,n,o], Yaolei Qi [h], Youssef Skandarani [g,i], João L. Vilaça [m], Xiyue Wang [p], Sen Yang [q], Arcot Sowmya [r], Susann Beier [a]

[a] School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney, Australia
[b] Prince of Wales Clinical School of Medicine, UNSW Sydney, Australia
[c] Department of Cardiology, Prince of Wales Hospital, Sydney, Australia
[d] Auckland City Hospital, Auckland, New Zealand
[e] Institute of Biomedical Engineering, National Taiwan University
[f] Institute of Computing Technology, Chinese Academy of Sciences, China
[g] ImViA Laboratory, University of Burgundy, Dijon, France
[h] Southeast University, China
[i] CASIS inc., Dijon, France
[j] Department of Electrical Engineering, National United University, Taiwan
[k] Nanjing University of Science and Technology, China
[l] Nanjing University
[m] 2Ai - School of Technology, Polytechnic Institute of Cávado and Ave, Barcelos, Portugal
[n] Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal
[o] Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal
[p] College of Computer Science, Sichuan University, Chengdu, China
[q] College of Biomedical Engineering, Sichuan University, Chengdu, China
[r] School of Computer Science and Engineering, University of New South Wales, Sydney, Australia

ARTICLE INFO

Keywords:
Coronary arteries
Image segmentation
Machine learning

ABSTRACT

Cardiovascular disease is a major cause of death worldwide. Computed Tomography Coronary Angiography (CTCA) is a non-invasive method used to evaluate coronary artery disease, as well as evaluating and reconstructing heart and coronary vessel structures. Reconstructed models have a wide array of for educational, training and research applications such as the study of diseased and non-diseased coronary anatomy, machine learning based disease risk prediction and in-silico and in-vitro testing of medical devices. However, coronary arteries are difficult to image due to their small size, location, and movement, causing poor resolution and artefacts. Segmentation of coronary arteries has traditionally focused on semi-automatic methods where a human expert guides the algorithm and corrects errors, which severely limits large-scale applications and integration within clinical systems. International challenges aiming to overcome this barrier have focussed on specific tasks such as centreline extraction, stenosis quantification, and segmentation of specific artery segments only. Here we present the results of the first challenge to develop fully automatic segmentation methods of full coronary artery trees and establish the first large standardized dataset of normal and diseased arteries. This forms a new automated segmentation benchmark allowing the automated processing of CTCAs directly relevant for large-scale and personalized clinical applications.

* Corresponding author.
E-mail address: r.gharleghi@student.unsw.edu.au (R. Gharleghi).

https://doi.org/10.1016/j.compmedimag.2022.102049
Received 24 November 2021; Received in revised form 7 February 2022; Accepted 10 February 2022
Available online 18 February 2022

1. Introduction

1.1. Clinical relevance

Cardiovascular disease is a significant cause of death worldwide World Health Organization (2012). Strategies for reducing the mortality and resource burden of cardiovascular disease has been an increasingly important area, with a considerable number of studies focussed on improved understanding of coronary anatomy and its influence on blood flow, performance of medical devices, surgical operations and disease behaviour. CTCA is a standard imaging technique in the assessment of coronary anatomy as well as presence and severity of disease. It is relatively non-invasive nature and relatively high resolution make it a preferred tool for most virtually reconstructed coronary models used for computational studies Sun and Xu (2014); Beier et al. (2016b); Pinho et al. (2019), virtual stenting Sun and Jansen (2019) and stent testing Beier et al. (2016a); Antoine et al. (2016), as well as physical and virtual reality models for education and training purposes Silva et al. (2018); Yoo et al. (2017). Fig. 1 shows a sample of the CTCA image stack and the corresponding 3D reconstruction. Segmentation of the coronary arteries for reconstruction has previously been time consuming, requiring manual annotations and corrections, thus prohibiting large scale coronary studies to date.

Robust, fully automated methods of reconstructing the coronary arteries have the potential to overcome these problems and allow rapidly processing of even an extensive amount of CTCA data. Few publications provide the source code of their evaluation strategy and sufficient information to verify their results. Use of a standardized dataset allows for objective comparison between methods developed by the challenge participants and future advances, whereas relying on private datasets and evaluation metrics requires accounting for the effect of intrinsic differences between the datasets and choice of and specific implementation of the evaluation metrics. Standardized datasets are commonly employed for specific tasks, such as the long running Brain Tumor Segmentation (BraTS) challenge Bakas et al. (2022) and have been successful in encouraging research on the proposed problem and providing a fair and easily accessible method to evaluate the performance of new algorithms.

1.2. Previous work

Few datasets of coronary arteries exist, as developing standardized datasets is time consuming and requires trained experts. Previous datasets such as the Rotterdam Coronary Artery Algorithm Evaluation Framework Schaap et al. (2009); Kirişli et al. (2013) have focussed on the closely related tasks of extracting vessel centrelines, stenosis quantification and segmentation of segments of the coronary arteries. A dataset of healthy coronary vessels is available, albeit without sufficient information on the methodology and without the necessary framework for evaluating the segmentation quality Iaizzo (2016). However, our dataset is a standardized benchmark dataset of the full coronary tree, including a balanced number of cases between healthy vessels and vessels with varying degrees of disease. This has allowed the development of fully automatic segmentation methods of the full coronary artery tree without expert assistance, which is vital for providing meaningful risk prediction attempts such as those for other image-based pathology assessments, because traditional (semi)-manual methods are not a scalable solution for the magnitude of the data processed in such assessments.

1.3. Scope

The Automated Segmentation of Coronary Arteries (ASOCA) challenge focusses on the fully automated segmentation and assessment of normal and diseased coronary arteries in CT Coronary Angiography (CTCA) images for two primary purposes: 1) to establish the first publicly available dataset of fully annotated normal and diseased left and right coronary artery trees, and 2) to establish a segmentation benchmark of this data. In this work we introduce the dataset and describe its development and present and discuss the best performing algorithms before concluding with an overall evaluation in the context of the clinical relevance of this work and its clinical impact.

2. Materials and Methods

2.1. Dataset

A total of 60 cases of CTCA images and data was used from the Coronary Atlas (https://www.coronaryatlas.org/) Medrano-Gracia et al. (2014), (2016). Briefly, the data was collected from Mercy Angiography, Auckland under ethics approval from the University of New South Wales Human Research Ethics Committee (Ref. HC190145) and University of Auckland Human Participants Ethics Committee (Ref. 022961). The selected patients were stratified based on available medical reports such that the selection comprised 30 patients with reported coronary disease, and 30 patients with no disease reported by the cardiologist. CTCA imaging was used due to its robustness and non-invasiveness, using a GE LightSpeed 64 slice CT Scanner. Data was collected using retrospective ECG-gated acquisition with the late diastole time point selected for reconstruction. Beta blockers were administered to lower the resting heart rate below 60bpm, and 60–80 ml Omnipaque 350 used as the contrast medium. The collected images have anisotropic resolution, with an in-plane resolution of 0.3–0.4 mm and out of plane resolution of 0.625 mm.

2.2. Annotation

Here, we describe a benchmark standardized dataset of coronary

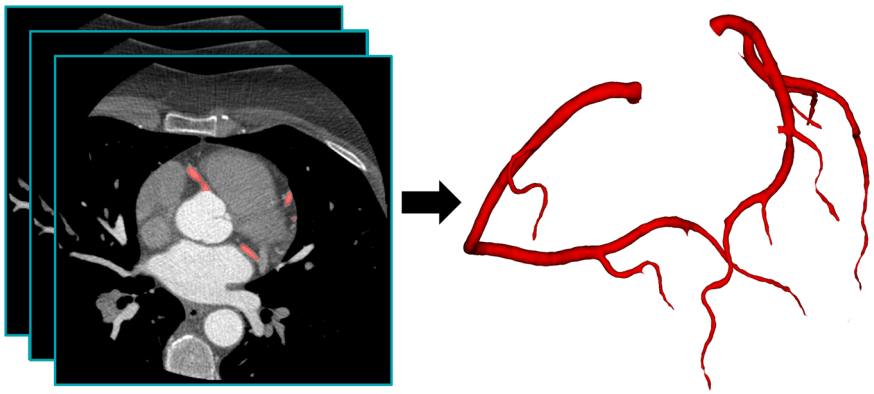

Fig. 1. Coronary vessels are annotated on each CTCA slice, and a 3D model is constructed from the stack of 2D slices.

artery geometry with reference segmentations performed by three experts. A standardised dataset of CTCA image data and expert annotations of the coronary artery was prepared based on the selected 60 patients. Each CTCA image was independently segmented by each annotator using 3D Slicer Fedorov et al. (2012); Kikinis et al. (2014). The segmentation started with a thresholding operation at a threshold chosen by the expert, followed by manually correcting vessel contours and over- and under-segmented vessels.

Segmentations from the three annotators were merged together using majority voting, i.e. pixels that at least two annotators had labelled as coronary vessel were included in the vessel mask to generate the final annotation. Post processing of the vessel mask involved removing thin vessels that would not be clinically significant and connecting vessel segments that were disconnected from the coronary tree due to high variation between annotators (particularly at heavily calcified areas). Annotations for the training set (20 healthy and 20 diseased arteries) were provided to the challenge participants. The training data used in this challenge is publicly available on Synapse Gharleghi et al. (2021).

2.3. Challenge design

The challenge design follows best practices for designing challenge rankings to avoid accidental or deliberate bias Maier-Hein et al. (2018). We used two evaluation metrics to ensure that the participant solutions were well generalized rather than tailored to exploiting a specific evaluation metric.

Dice score Dice (1945) is a commonly used segmentation metric measuring the overlap between the proposed segmentation and the ground truth, defined as

$$\text{DiceCoefficient} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{1}$$

where TP is the number of true positives, FP is the number of false positives and FN the number of false negatives.

The Hausdorff distance is used to evaluate how close corresponding points between the submission and the ground truth are. This provides a measure of accuracy of the segmentation at the boundary of the vessels. Hausdorff Distance is originally defined as

$$\text{HD} = \max\left(\max_{x \in X} \min_{y \in Y} d(x,y), \max_{y \in Y} \min_{x \in X} d(x,y)\right) \tag{2}$$

where X and Y are the two point sets compared and d(x,y) signifies the distance from point x to y. Since Hausdorff distance is very sensitive to noise and outliers, we used the 95th percentile Hausdorff distance, which is commonly used and is more robust in the presence of outliers Taha and Hanbury (2015).

Using multiple evaluation metrics is helpful to ensure that developed methods are fairly evaluated, however it poses challenges in determining the best performing method when results from multiple metrics are combined. Different methods of aggregating these metrics and whether the mean or median is used produce significantly different results Maier-Hein et al. (2018). In this challenge, the participants were assigned a separate rank for Dice coefficient and Hausdorff distance, with the final rank determined based on the sum of the ranks for each metric.

Additional metrics are calculated to better explore the performance of the methods presented. Precision and recall of the proposed methods is calculated, defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

,

where TP is the number of true positives, TN the number of true negatives and FN is the number of false negatives. Coronary vessels are expected to be continuously connected, with only two distinct components for the left and right coronary trees. To evaluate the connectivity of the segmentation produced, the Dice score, Hausdorff distance, precision and recall were also evaluated for the two largest connected components in the segmentation, which will discard small segments not connected to the main coronary arteries. A 6-connected structuring element was used to identify connected components, which were then sorted by voxel count and all components except the two largest were discarded. We also calculated the Hausdorff distance for the topological skeleton of the segmentations, looking at how the centrelines of the segmentation produced compared versus the centrelines based on the ground truth. Topological skeletons of the submitted segmentation and ground truth were constructed using a skeletonization algorithm Lee et al. (1994) and 95% Hausdorff distance was calculated for the two centrelines.

An online evaluation tool allows participants to submit results of their algorithms, which was then automatically evaluated and ranked alongside other participants. The source code for the evaluation framework is publicly available on GitHub Gharleghi (2021), which includes implementation of the evaluation metrics and set up to reproduce the docker image used.

The use of this platform was demonstrated at the Medical Image Computing and Computer-Assisted Intervention (MICCAI 2020) conference Gharleghi et al. (2020), (2021). The 35 participants were asked to develop automated algorithms to segment coronary arteries from CTCA images of normal and diseased cases. First we provided 40 cases, 20 normal and 20 diseased, for the participants to develop their methodology, with the other 20 cases withheld to test and validate the algorithms developed. The top ten participants (two had to drop out, leaving 8 final participants) were invited to presented their results during the conference and are discussed here. An automated scoring system was developed to generate scores based on the submissions of the participants and adjust the leader board accordingly. The challenge website is available for future researchers via asoca.grand-challenge.org to benchmark their algorithms against this standard dataset.

3. Results

Results and scoring are available via https://asoca.grand-challenge.org/MICCAI_Ranking/. These results will continue to be updated with new participants and submissions, and will likely be different than the results described here as the submitted methods improve. An overview of the best performing submissions is provided in Figs. 2 and 3, shown in order of best performance based on the ranking scheme described above. Care must be taken when interpreting these rankings, as the rankings do not capture all differences in the developed algorithms. Specific requirements, such as low computation requirements, may favour some algorithms over others.

In Fig. 2 the performance of ensemble models created by combining the participant solutions is shown. Each model uses majority voting, such that 2, 3 and 4 participants must agree on the assigned label corresponding to the top 3, 5 and 7 ensembles. The ensemble models performed significantly better than the initial submissions, particularly on the Hausdorff distance metric.Table 1.

In Table 2 the methodology and differences in algorithms developed by the participants are summarised. In Fig. 4 examples of the segmentations submitted for the best and worst performing cases are shown. Visually most of the resulting segmentations appear to be physiologically accurate and appropriate for further use. Errors generally occur far away from the ostia, with the thin downstream branches containing missing segments. This is significantly more pronounced in the diseased vessels, whereby even an ensemble of the top three performing participants is missing large segments of the left circumflex artery (Fig. 4 bottom right).

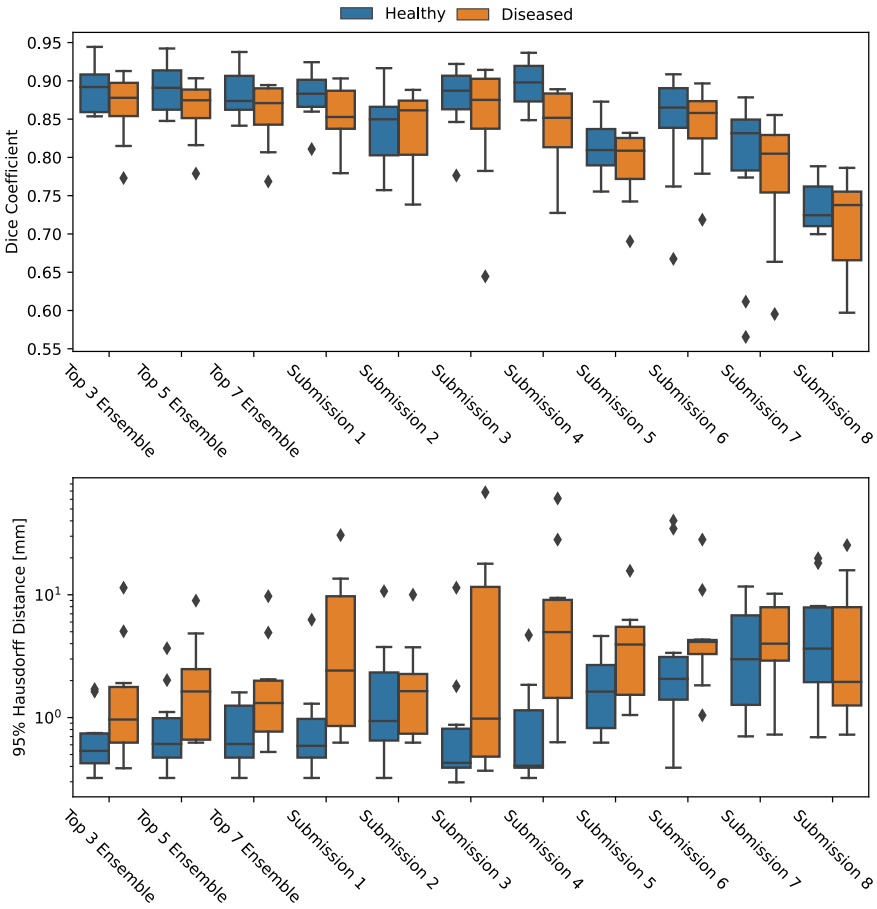

Fig. 2. Results of participants' algorithms on the test set, grouped by presence or absence of disease.

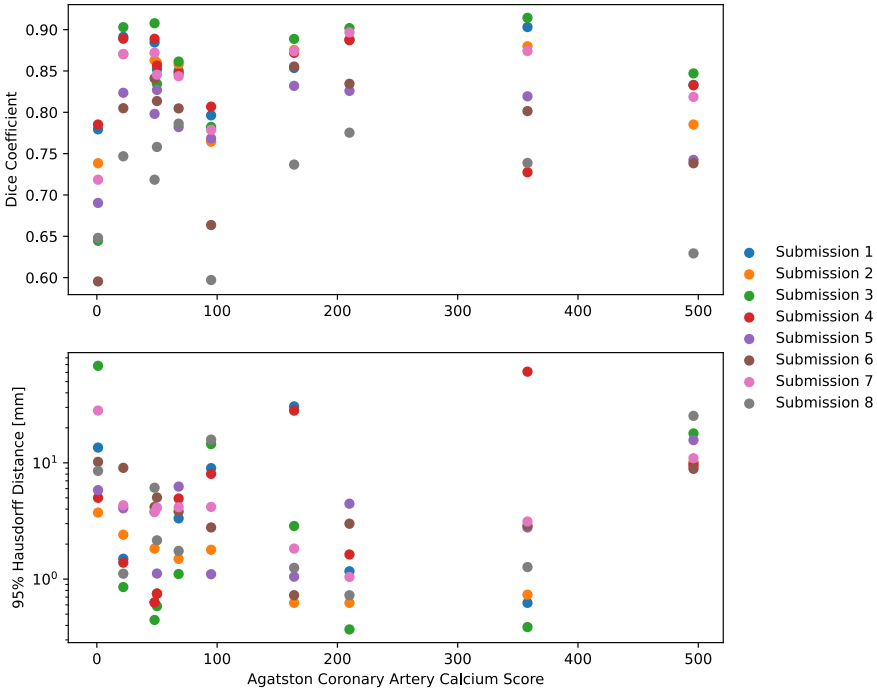

Fig. 3. Influence of calcification on segmentation accuracy of the participants.

Table 1
Summary of the evaluation metrics for each method.

	Submission				Largest Two Connected Bodies				Topological Skeleton
	Dice Score	HD	Precision	Recall	Dice Score	HD	Precision	Recall	HD
Top 3 Ensemble	0.88 ± 0.04	1.56 ± 2.55	0.95 ± 0.04	0.82 ± 0.06	0.87 ± 0.04	5.08 ± 6.79	0.95 ± 0.04	0.81 ± 0.07	8.35 ± 7.82
Top 5 Ensemble	0.88 ± 0.04	1.73 ± 2.08	0.95 ± 0.04	0.81 ± 0.06	0.87 ± 0.04	5.01 ± 6.75	0.96 ± 0.04	0.8 ± 0.07	9.9 ± 7.7
Top 7 Ensemble	0.87 ± 0.04	1.61 ± 2.17	0.96 ± 0.04	0.8 ± 0.06	0.86 ± 0.04	5.0 ± 6.75	0.97 ± 0.03	0.79 ± 0.07	8.8 ± 7.76
Submission 1	0.87 ± 0.04	4.16 ± 7.3	0.94 ± 0.05	0.81 ± 0.06	0.86 ± 0.04	5.46 ± 6.48	0.95 ± 0.04	0.8 ± 0.07	11.38 ± 8.67
Submission 2	0.84 ± 0.05	2.34 ± 2.92	0.92 ± 0.05	0.77 ± 0.06	0.82 ± 0.06	13.34 ± 13.09	0.95 ± 0.04	0.73 ± 0.09	9.19 ± 5.82
Submission 3	0.86 ± 0.07	6.22 ± 15.52	0.89 ± 0.06	0.84 ± 0.09	0.86 ± 0.07	6.34 ± 15.49	0.89 ± 0.06	0.84 ± 0.09	12.73 ± 16.2
Submission 4	0.87 ± 0.05	6.57 ± 14.27	0.92 ± 0.08	0.82 ± 0.07	0.87 ± 0.04	5.23 ± 6.96	0.95 ± 0.05	0.81 ± 0.07	13.35 ± 14.09
Submission 5	0.8 ± 0.04	3.31 ± 3.42	0.93 ± 0.05	0.71 ± 0.07	0.79 ± 0.05	12.81 ± 11.11	0.94 ± 0.04	0.68 ± 0.08	19.38 ± 19.95
Submission 6	0.84 ± 0.06	7.73 ± 11.83	0.94 ± 0.05	0.77 ± 0.09	0.81 ± 0.1	18.9 ± 18.39	0.95 ± 0.05	0.71 ± 0.13	17.06 ± 13.06
Submission 7	0.78 ± 0.1	4.77 ± 3.51	0.91 ± 0.06	0.69 ± 0.12	0.72 ± 0.15	28.37 ± 18.55	0.96 ± 0.03	0.6 ± 0.17	14.07 ± 6.69
Submission 8	0.73 ± 0.05	6.55 ± 7.4	0.83 ± 0.05	0.65 ± 0.07	0.7 ± 0.09	11.29 ± 14.7	0.85 ± 0.06	0.61 ± 0.12	14.43 ± 9.74

Table 2
Summary of the methodology used by each participant. Submission 6 participants were not available.

	Structure	Preprocessing	Post processing	Loss function
Submission 1	nnU-NetIsensee et al. (2020)	Scale map generationWang et al. (2020)	Pretrained neural network used to segment the epicardium and discard vessels outside the region of interest	Soft Dice lossBertels et al. (2019)+ Cross entropy loss Gordon-Rodriguez et al. (2020)
Submission 2	Modified 2D U-Net with residual connections and Squeeze-and-Excitation decoder blockHu et al. (2018); Li et al. (2019),	Image contrast normalized to zero mean and unit variance	–	Soft Dice lossBertels et al. (2019)+ Cross entropy loss Gordon-Rodriguez et al. (2020)
Submission 3	U-Net with vesselness features Ronneberger et al. (2015)	Pretrained U-Net architecture to remove pulmonary vessels, often erroneously classified as coronary vessels.Frangi vesselness filter to highlight vascular structuresFrangi et al. (1998)	Ratio based connected component analysis to remove unrelated components	Focal LossLin et al. (2017)
Submission 4	nnU-NetIsensee et al. (2020)	Hessian vesselness filter	Five fold cross validation, with an ensemble of the five models used to submit predictions	Soft Dice loss + Cross entropy loss
Submission 5	Standard 3D U-NetRonneberger et al. (2015)	Utilizes coarse 3D U-Net to find and segment region of interest, allowing for finer segmentations with low computational cost.Data augmentation using random flips, translation and scaling. Image contrast normalized and images resampled to isotropic spacing.	–	Soft Dice lossBertels et al. (2019) + Cross entropy loss Gordon-Rodriguez et al. (2020)
Submission 7	Modified 2D U-Net with additional layers and different pooling convention, considering one slice at a time	–	–	Soft Dice lossBertels et al. (2019)+ Focal lossLin et al. (2017)
Submission 8	Standard 3D U-NetRonneberger et al. (2015)	Images resampled to same dimensions before training	Segmentation constrained to regions near automatically extracted vessel centrelines	Soft Dice LossBertels et al. (2019)

4. Discussion

While CTCA images try to only capture the cardiac body region, the resulting volumes (512×512×200 on average) are usually too large for deep learning algorithms. Most participants have used various strategies to reduce the computational power required. Several participants have used deep learning to crop the CTCA volumes to only include the heart and surrounding blood vessels, while others have focussed on using 2D implementations that consider single slices, or a subvolume of multiple slices of the dataset, at a time rather than the full volume. In submissions 5 and 8 the dataset was resampled to a lower resolution, however due to the small size of coronary vessels this is expected to have an adverse effect on the overall results and likely contributed to the low ranking of these submissions. Other approaches such as gradient check pointing Kellman et al. (2020) to reduce the memory complexity of the models are possible, however implementation of these methods for arbitrary architectures is difficult and no participant opted to utilize them in this challenge.

The U-Net architecture has proven to be extremely effective in medical image segmentation tasks. The methods used in this challenge are generally based on the U-Net architecture, either directly using it or using various improvements proposed since U-Net's inception. Two participants used the state-of-the-art nnU-Net architecture Isensee et al. (2020), a self-adapting network that can be automatically configured for new tasks without expensive hyper parameter optimization studies. nnU-Net has been shown to perform well for 23 datasets comprised of various organs, 2D vs 3D and different imaging modalities. Given the good performance of this network on a large number of datasets, it is not surprising that it also performs well in this challenge. Three participants used the 2-D version of U-Net, with two of those using a single slice as the input, and one participant using a sliding window of three slices to integrate more information about neighbouring slices.

Several participants used preprocessing filters such as the Frangi vesselness filter to improve the contrast for tubular structures such as blood vessels and suppress other organs in the images. The vesselness features are generally provided as an additional feature to the network, along with the original image intensity. Vessel enhancing filters are expected to work well for these tasks specially as the initial contrast is fairly high, however they generally require relatively high computation costs compared and it may not be feasible to use for large datasets. Calculating multi scale Frangi vesselness for the entire volume can take several minutes per case, whereas the inference time of neural networks

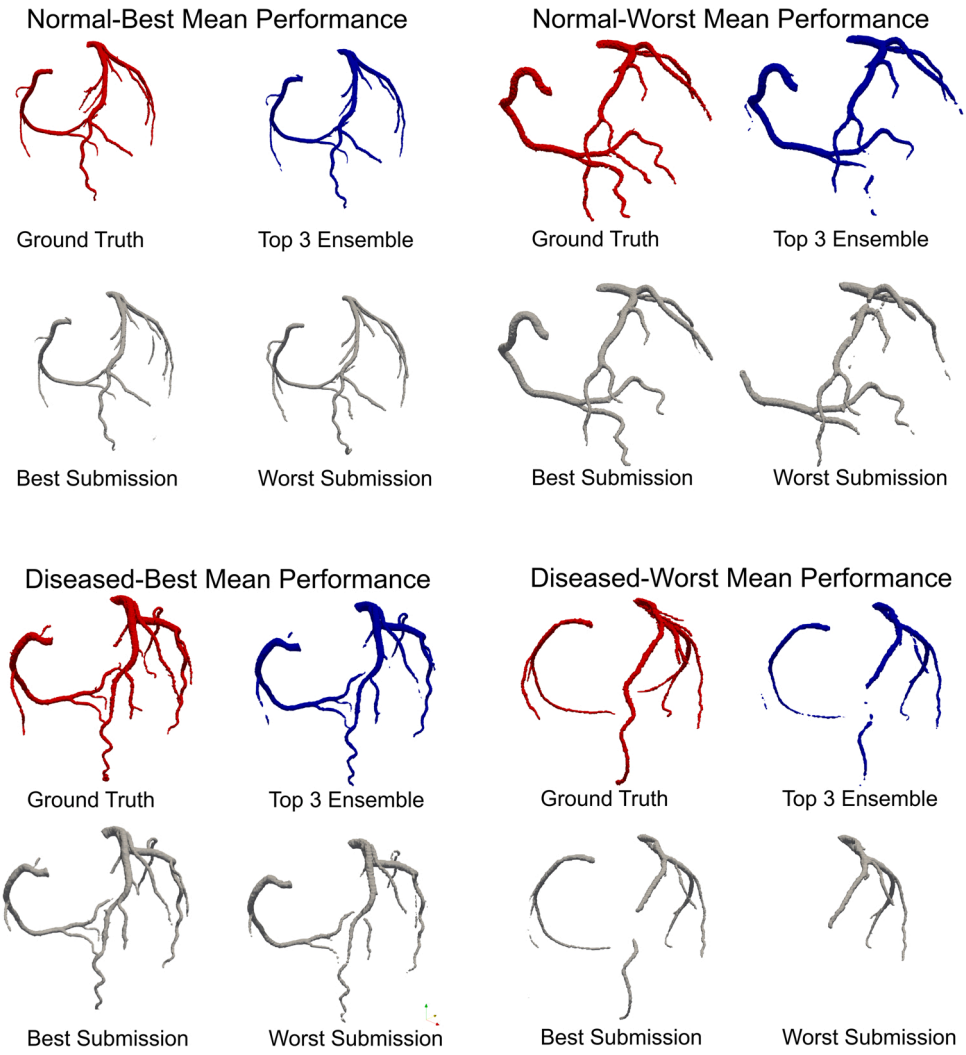

Fig. 4. Highest and lowest segmentation quality for normal and diseased cases.

is usually orders of magnitude faster.

Various structures similar to coronary vessels such as pulmonary vessels or bone segments often get included in the initial segmentation. Some participants have opted for post-processing based on connected components in the segmentation to remove small disconnected components, which would improve the performance of the models. Approaches with no post processing have significantly worse Dice scores and Hausdorff distances.

Soft Dice Loss Bertels et al. (2019) was most commonly chosen as the objective function, often in combination with cross entropy Gordon--Rodriguez et al. (2020) or focal loss Lin et al. (2017). Since coronary vessels are small and comprise a small fraction of the heart and nearby tissue, the methodology used must be robust to heavy class imbalance. Combination of Soft Dice loss and binary cross entropy seems to be the best performing choice. Focal loss is expected to handle such conditions Lin et al. (2017) and also performs well in this challenge. Use of Soft Dice loss alone seems to have significantly worse performance.

The methodology in Submission 1 had the highest rank based on the combined Dice score and Hausdorff distance rank. The second best performing algorithm (with the lowest Hausdorff distance) was Submission 2. Both these methodologies are heavily inspired by U-Net networks, with additional modifications to account for the sparsity and small size of coronary vessels without requiring infeasibly large computational resources. Submission 1 uses an ensemble model of three models to capture features associated with tubular structures and improve segmentation of the coronary vessels, while the second submission uses significantly more advanced encoder and decoder structures, with the addition of squeeze and excitation blocks to capture details of small coronary structures.

It can be seen from Fig. 2 that vessels without coronary disease were easier to segment, resulting in higher Dice scores (0.851 vs 0.829, $p = 0.006$) and lower Hausdorff distance (3.02 mm vs 6.97 mm, $p = 0.0004$) for most participants. Despite this relationship, cases with higher calcium score did not seem to be particularly difficult to segment (Fig. 3, $p = 0.86$ for Dice score and $p = 0.37$ for Hausdorff distance), with the worst performing segmentations associated with the lowest calcium scores. This is likely due to the presence of non-calcified plaque and stenoses in these vessels, which seems to have a larger effect on segmentation performance. Similarly, annotator variability was not influenced by the calcium score of patients.

Dice score and Hausdorff distance showed a significant inverse correlation, with a Kendall Tau rank correlation of -0.52 ($p < 0.0001$). This is to be expected as these metrics were deliberately selected to measure counteracting quality factors, encouraging submissions that perform well in general rather than bespoke solutions for a specific metric. Addition of other segmentation quality metrics in the future may further improve this at the cost of increasing the complexity of aggregating and ranking participants.

The methods investigated showed very high precision, with most methods reaching 0.9 and above. However the recall values are

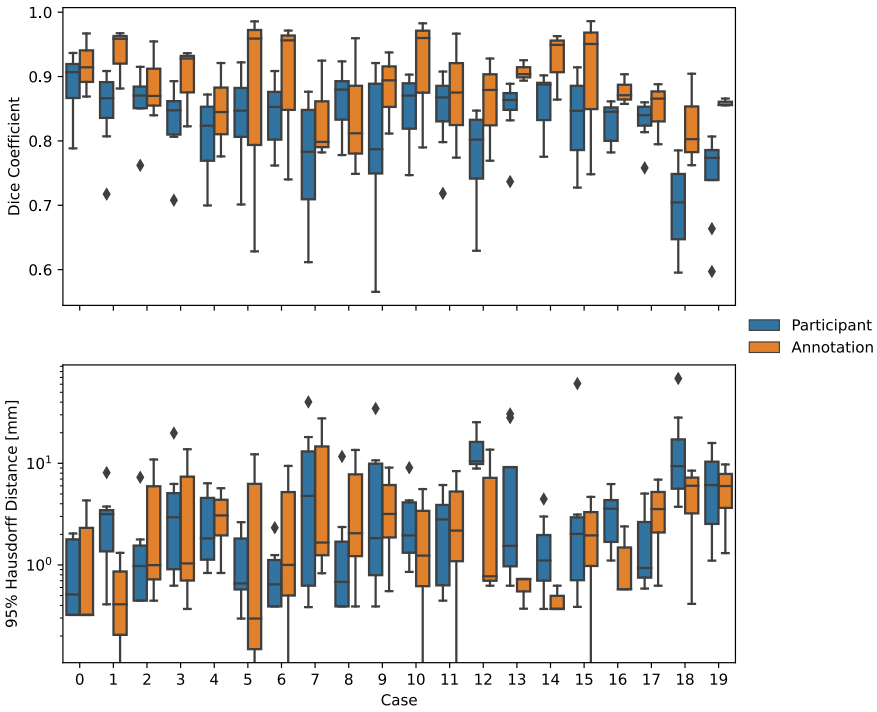

Fig. 5. Comparison of the segmentation score compared to annotation variability.

significantly lower. Considering the image properties involved in the coronary arteries with relatively high noise, potential artefacts and a highly imbalanced dataset low recall values are expected. While the proximal sections of the arteries are accurately segmented, segmentations get progressively less reliable towards distal sections of the arteries, and missing distal segments tend to reduce the recall of this methods. Removing the small disconnected components and keeping only the largest two components tends to have a relatively small effect on the Dice score. In all submissions this slightly reduced the Dice score, as Dice score highly penalizes false negative predictions. The Hausdorff distance was significantly increased by only keeping the largest two components, as the removed segments now represent a very large distance to the ground truth. It is clear that despite these segments being disconnected from the main two coronary trees, they are correctly predicted parts of the vessel and removing these components would significantly harm the performance of these methods. More sophisticated post-processing methods can be used to combat this problem, for example previous literature Han et al. (2016) has used a search strategy to detect disconnected branches and attempt to reconnect it to the coronary tree.

Annotator variability seems to be higher on healthy vessels, with diseased vessels having relatively consistent annotations. As expected, the performance of the algorithms developed heavily depends on annotator variability, so that with few exceptions, cases that were difficult for the experts were also difficult for the developed segmentation algorithms (Fig. 5).

While the CTCA resolution used here is sufficient to resolve clinically significant vessels, most vessels will only be a few voxels across. This significantly exacerbates effects of image artefacts and calcium blooming, such that images with extremely heavy artefacts and calcification would be difficult to segment. We have chosen the main metrics used for ranking participants such that overall segmentation quality is favoured over the connectivity of the resulting coronary tree, due to the assumption that small disconnected segments are likely not clinically significant, however depending on the specific application other metrics may be appropriate. As the current dataset does not include ground truth data regarding stenosis locations, the metrics do not quantify behaviour specifically around stenosis segments. Additionally, all data has been collected from one medical centre and hence the results obtained are likely to be biased towards the imaging protocol and CT machine used. Further studies could extend this dataset with CTCA images from different medical centres.

5. Conclusion

The best performing algorithms were based on U-Net, highlighting the continued dominance of the U-Net architecture for biomedical segmentation tasks as well as the potential improvements that can be achieved to handle unusual cases, including sparse and small coronary arteries. This challenge has established a new benchmark dataset which is publicly available and can be used. The training dataset is also made available for research uses requiring expert curated segmentations. The challenge continues to be available and we invite participants to test their methods and share their findings and improvements.

Ultimately this effort will help to develop better segmentation methods, requiring minimal expert input and thus open new opportunities for large scale coronary artery image assessments to study physiology and pathophysiology, uncovering important trends and relationships for improved patient care in future.

CRediT authorship contribution statement

Ramtin Gharleghi: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Dona Adikari:** Methodology, Data Curation, Writing – original draft, Writing – review & editing. **Katy Ellenberger:** Methodology, Data Curation, Writing – original draft, Writing – review & editing. **Sze-Yuan Ooi:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Chris Ellis:** Data Curation, Writing – original draft, Writing – review & editing. **Chung-Ming Chen:** Software, Writing – original draft. **Ruochen Gao:** Software, Writing – original draft. **Yuting He:** Software, Writing – original draft. **Raabid Hussain:** Software, Writing – original draft. **Chia-Yen Lee:** Software, Writing – original draft. **Jun Li:** Software, Writing – original draft. **Jun Ma:** Software,

Writing – original draft. **Ziwei Nie:** Software, Writing – original draft. **Bruno Oliveira:** Software, Writing – original draft. **Yaolei Qi:** Software, Writing – original draft. **Youssef Skandarani:** Software, Writing – original draft. **João L.Vilaça:** Software, Writing – original draft. **Xiyue Wang:** Software, Writing – original draft. **Sen Yang:** Software, Writing – original draft. **Arcot Sowmya:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Susann Beier:** Conceptualization, Methodology, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Jane Liggins and Miriam Hayward, Intra Imaging for assisting with the collection of the challenge dataset. SB acknowledges the Auckland Academic Health Alliance (AAHA) and the Auckland Medical Research Foundation (AMRF) for their financial support and endorsement. The work of BO was funded in part by the project "NORTE-01–0145-FEDER-000045", supported by Northern Portugal Regional Operational Programme (Norte2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). BO and JV also acknowledge support from FCT and the European Social Found, through Programa Operacional Capital Humano (POCH), in the scope of the PhD grant SFRH/BD/136721/2018.

References

World Health Organization, 2012. The atlas of heart disease and stroke. World Health Organization.

Antoine, E.E., Cornat, F.P., Barakat, A.I., 2016. The stentable in vitro artery: an instrumented platform for endovascular device development and optimization. J. R. Soc. Interface 13, 20160834.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R., Berger, C., Ha, S., Rozycki, M., et al., 2022. Identifying Best. Mach. Learn. Algorithms brain Tumor Segm., Progress. Assess., Overall Surviv. Predict. brats Chall.

Beier, S., Ormiston, J., Webster, M., Cater, J., Norris, S., Medrano-Gracia, P., Young, A., Cowan, B., 2016a. Hemodynamics in idealized stented coronary arteries: important stent design considerations. Ann. Biomed. Eng. 44, 315–329.

Beier, S., Ormiston, J., Webster, M., Cater, J., Norris, S., Medrano-Gracia, P., Young, A., Cowan, B., 2016b. Impact of bifurcation angle and other anatomical characteristics on blood flow-a computational study of non-stented and stented coronary arteries. J. Biomech. 49, 1570–1582.

Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., Blaschko, M.B., 2019. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 92–100.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al., 2012. 3d slicer as an image computing platform for the quantitative imaging network. Magn. Reson. Imaging 30, 1323–1341.

Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering. In: International conference on medical image computing and computer-assisted intervention. Springer, pp. 130–137.

Gharleghi, R., 2021.Ramtingh/ASOCA_MICCAI2020_Evaluation: MICCAI Evaluation.10.5281/zenodo.4460628.

Gharleghi, R., Samarasinghe, G., Sowmya, P.A., Beier, S., 2020. Autom. Segm. Coron. Arter. https://doi.org/10.5281/zenodo.3819799.

Gharleghi, R., Adikari, D., Ellenberger, K., Pua, Q.S., Shen, C., Webster, M., Ellis, C., Sowmya, A., Ooi, S.Y., Beier, S., 2021. Computed tomography coronary angiogram images, lumen annotations and associated data of normal and diseased coronary arteries. Synapse. https://doi.org/10.7303/SYN25684144. ⟨https://repo-prod.prod.sagebase.org/repo/v1/doi/locate?id=syn25684144&type=ENTITY⟩.

Gordon-Rodriguez, E., Loaiza-Ganem, G., Pleiss, G., Cunningham, J.P., 2020.Uses and abuses of the cross-entropy loss: case studies in modern deep learning. arXiv: 2011.05231.

Han, D., Shim, H., Jeon, B., Jang, Y., Hong, Y., Jung, S., Ha, S., Chang, H.J., 2016. Automatic coronary artery segmentation using active search for branches and seemingly disconnected vessel segments from coronary ct angiography. PLoS One 11, e0156837.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 7132–7141.

Iaizzo, P.A., 2016.The visible heart® project and free-access website 'atlas of human cardiac anatomy'. EP Europace 18, iv163-iv172.

Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2020. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods 1–9.

Kellman, M., Zhang, K., Markley, E., Tamir, J., Bostan, E., Lustig, M., Waller, L., 2020. Memory-efficient learning for large-scale computational imaging. IEEE Trans. Comput. Imaging 6, 1403–1414.

Kikinis, R., Pieper, S.D., Vosburgh, K.G., 2014. 3d slicer: a platform for subject-specific image analysis, visualization, and clinical support. In: Intraoperative imaging and image-guided therapy. Springer, pp. 277–289.

Kirişli, H., Schaap, M., Metz, C., Dharampal, A., Meijboom, W.B., Papadopoulou, S.L., Dedic, A., Nieman, K., de Graaf, M.A., Meijs, M., et al., 2013. Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. Med. Image Anal. 17, 859–876.

Lee, T.C., Kashyap, R.L., Chu, C.N., 1994. Building skeleton models via 3-d medial surface axis thinning algorithms. CVGIP: Graph. Models Image Process. 56, 462–478.

Li, X., Wang, W., Hu, X., Yang, J., 2019. Selective kernel networks. Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. 510–519.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. Proc. IEEE Int. Conf. Comput. Vis. 2980–2988.

Maier-Hein, L., Eisenmann, M., Reinke, A., Onogur, S., Stankovic, M., Scholz, P., Arbel, T., Bogunovic, H., Bradley, A.P., Carass, A., et al., 2018. Why rankings of biomedical image analysis competitions should be interpreted with care. Nat. Commun. 9, 1–13.

Medrano-Gracia, P., Ormiston, J., Webster, M., Beier, S., Ellis, C., Wang, C., Young, A.A., Cowan, B.R., 2014. Construction of a coronary artery atlas from ct angiography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 513–520.

Medrano-Gracia, P., Ormiston, J., Webster, M., Beier, S., Young, A., Ellis, C., Wang, C., Smedby, Ö., Cowan, B., 2016. A computational atlas of normal coronary artery anatomy. Eur.: J. Eur. Collab. Work. Group Interv. Cardiol. Eur. Soc. Cardiol. 12, 845–854.

Pinho, N., Castro, C.F., António, C.C., Bettencourt, N., Sousa, L.C., Pinto, S.I.S., 2019. Correlation between geometric parameters of the left coronary artery and hemodynamic descriptors of atherosclerosis: Fsi and statistical study. Med. Biol. Eng. Comput. 57, 715–729.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241.

Schaap, M., Metz, C.T., van Walsum, T., van der Giessen, A.G., Weustink, A.C., Mollet, N. R., Bauer, C., Bogunović, H., Castro, C., Deng, X., et al., 2009. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. Med. Image Anal. 13, 701–714.

Silva, J.N., Southworth, M., Raptis, C., Silva, J., 2018. Emerging applications of virtual reality in cardiovascular medicine. JACC: Basic Transl. Sci. 3, 420–430.

Sun, Z., Jansen, S., 2019. Personalized 3d printed coronary models in coronary stenting. Quant. Imaging Med. Surg. 9, 1356.

Sun, Z., Xu, L., 2014. Computational fluid dynamics in coronary artery disease. Comput. Med. Imaging Graph. 38, 651–663.

Taha, A.A., Hanbury, A., 2015. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. BMC Med. Imaging 15, 1–28.

Wang, Y., Wei, X., Liu, F., Chen, J., Zhou, Y., Shen, W., Fishman, E.K., Yuille, A.L., 2020. Deep distance transform for tubular structure segmentation in ct scans. Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. 3833–3842.

Yoo, S.J., Spray, T., Austin III, E.H., Yun, T.J., van Arsdell, G.S., 2017. Hands-on surgical training of congenital heart surgery using 3-dimensional print models. J. Thorac. Cardiovasc. Surg. 153, 1530–1540.