Show, Attend and Detect: Towards Fine-grained Assessment of Abdominal Aortic Calcification on Vertebral Fracture Assessment Scans

Syed Zulqarnain Gilani[1,2,3*], Naeha Sharif[1,2,3*], David Suter[1,2,3], John T. Schousboe[4], Siobhan Reid[5], William D. Leslie[6], and Joshua R. Lewis[1]

[1] Nutrition & Health Innovation Research Institute, Edith Cowan University.
s.gilani@ecu.edu.au
[2] Centre for AI&ML, School of Science, Edith Cowan University.
[3] Computer Science and Software Engineering, the University of Western Australia.
[4] Park Nicollet Clinic and HealthPartners Institute, HealthPartners, Minneapolis.
[5] Department of Electrical and Computer Engineering, University of Manitoba.
[6] Departments of Medicine and Radiology, University of Manitoba, Canada.
[*] Joint First Authors

Abstract. More than 55,000 people world-wide die from Cardiovascular Disease (CVD) each day. Calcification of the abdominal aorta is an established marker of asymptomatic CVD. It can be observed on scans taken for vertebral fracture assessment from Dual Energy X-ray Absorptiometry machines. Assessment of Abdominal Aortic Calcification (AAC) and timely intervention may help to reinforce public health messages around CVD risk factors and improve disease management, reducing the global health burden related to CVDs. Our research addresses this problem by proposing a novel and reliable framework for automated "fine-grained" assessment of AAC. Inspired by the vision-to-language models, our method performs sequential scoring of calcified lesions along the length of the abdominal aorta on DXA scans; mimicking the human scoring process.

Keywords: Abdominal Aortic Calcification · Sequential Prediction · Dual-Energy Xray.

1 Introduction

Cardiovascular Disease (CVD) is the leading cause of death globally, and a significant contributor to disability worldwide [16]. Vascular calcification, a stable marker of asymptomatic CVD, occurs when calcium builds up within the walls of the arteries undergoing the atherosclerotic process, and often begins decades before clinical events such as heart attacks or strokes [13]. The abdominal aorta is one of the first vascular beds where calcification is seen, and is a marker for generalised atherosclerosis at other vascular beds [21, 11]. The presence and extent of Abdominal Aortic Calcification (AAC) is associated with increased risk of future cardiovascular hospitalizations and death [8]. Given that AAC often occurs well before clinical events, this paper provides a window of opportunity to identify people at risk and intervene in a timely manner before they suffer cardiovascular events such as heart attacks or strokes [14].

The extent and severity of AAC can be assessed using lateral-lumbar radiographs, lateral spine Vertebral Fracture Assessment (VFA) Dual-energy X-ray

Absorptiometry (DXA) and Quantitative Computed Tomography (QCT). Out of these, VFA DXA scans have the least amount of radiation but are of lower resolution and contain more noise. These scans can be used to semi-quantify AAC using the widely adopted Kauppila 24-point scoring method (See Section 2.1 for more details), which measures the calcification along the length of abdominal aorta from L1-L4. However, acquiring manual assessments for DXA images is not only time-consuming and expensive but also subjective [17].

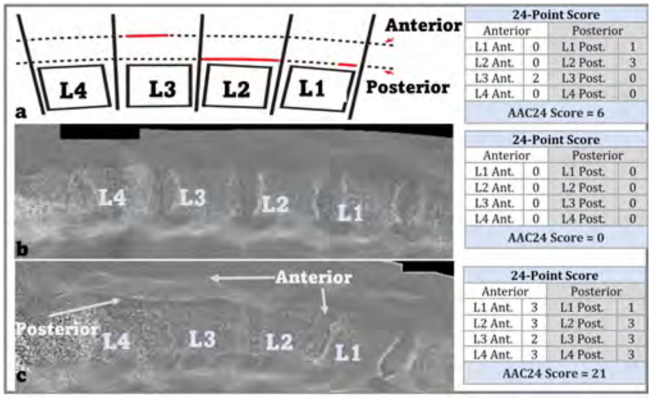
The published methods [4, 3, 15] predict an overall AAC-24 score for each scan. A brief summary of these methods is give in Section 2.2. For the sake of reliability and explainablility, it is pertinent that the overall AAC-24 assessment mimics the human scoring process. While a single AAC-24 score provides clinically useful information on levels of cardiovascular risk, more granular (location of calcification) scoring is required to provide results to general practitioners and patients. Additionally, more granular scoring may provide better understanding of why and how AAC develops and progresses. Finally, AAC in different parts of the abdominal aorta may be more or less important for clinical outcomes such as heart attacks, stroke or death.

We address the shortcomings of the existing methods by proposing an effective framework to generate the fine-grained scores in a sequential manner without the need for ground truth annotations for the lumbar regions. We draw our inspiration from the vision-to-language domain, in particular image captioning [2], where the task is to transform visual information into a sequence of words. We intend to transform the images into a sequence of fine-grained AAC-24 scores. Our attention based encoder-decoder network is a step towards mimicking the human-like AAC-24 scoring method (see Section 2.1 for more details). Though such models are quite popular in the language/vision-to-language domain, this is the first time they have been used to address the particular problem of AAC-24 scoring. However, this domain adaptation comes with its own challenges. Language has a syntax and a structure which is comparatively easier to learn by the sequential models, provided they are trained on a large corpus of data. Moreover, language is flexible as there can be a number of plausible solutions, and successive words in a sentence are predictable. In contrast, AAC-24 scores are rigid (only three possibilities per segment per vertebrae), random, and the margin of error is small. We also lack the advantage of having large annotated datasets.

Our model focuses on the most salient aortic regions while generating a sequence of scores. Moreover, our algorithm can classify patients into the three risk categories of low, medium and high, with an accuracy, sensitivity, and specificity of 82%, 74% and 80% respectively on the test set. The AAC-24 scores generated by our algorithm are highly correlated (>80%) with human assessments.

In this context our contributions are as follows:

- For the first time, we frame the problem of generating fine-grained AAC assessments as *translating DXA scans to sequential scores.*
- We propose an effective framework to generate fine-grained AAC-24 scores. Our scoring process is more understandable because it is easier to compare with the way humans score, and can point out regions adjacent to particular vertebrae that are highly calcified.



24-Point Score			
Anterior		Posterior	
L1 Ant.	0	L1 Post.	1
L2 Ant.	0	L2 Post.	3
L3 Ant.	2	L3 Post.	0
L4 Ant.	0	L4 Post.	0
AAC24 Score = 6			

24-Point Score			
Anterior		Posterior	
L1 Ant.	0	L1 Post.	0
L2 Ant.	0	L2 Post.	0
L3 Ant.	0	L3 Post.	0
L4 Ant.	0	L4 Post.	0
AAC24 Score = 0			

24-Point Score			
Anterior		Posterior	
L1 Ant.	3	L1 Post.	1
L2 Ant.	3	L2 Post.	3
L3 Ant.	2	L3 Post.	3
L4 Ant.	3	L4 Post.	3
AAC24 Score = 21			

Fig. 1. AAC-24 scoring to quantify the severity of AAC. The scores of all eight segments along with the AAC-24 scores are given in the tables alongside each image.

- Our attention-based model has independent decoders for assessing the anterior and posterior aortic walls, which leads to better performance when compared to a single decoder for assessing both aortic walls.
- We show that despite the limited size of our dataset and severe class imbalance, our model achieves a high level of correlation with human assessment.

2 Related Work

2.1 Kauppila Scale and AAC classes

We have used the 24-point semi-quantitative scale [6] (commonly known as the AAC-24 scale), to quantify the extent of calcification in abdominal aorta. This is the most widely used scale [17, 19] to assess the location, severity and progression of calcified lesions on the anterior and posterior abdominal aortic walls in the region parallel to the lower lumbar spine L1 – L4.

A score of ‘1’ is given if $\leq 1/3$ of the aortic wall is calcified, ‘2’ if $> 1/3$ to $\leq 2/3$ is calcified, or ‘3’ if $> 2/3$ is calcified. The anterior and posterior walls and segments are then summed up, for a possible score of up to 24. The scores for the anterior and posterior wall segments are summed (for a possible score of up to 24). The whole process is time consuming ($\sim$10 minutes), needs specialised equipment (radiology monitor) and subjective (depends on training/experience of reader). Furthermore, the aorta is not visible in the scans if there is no calcification (Figure 1(b)).

Figure 1 shows three examples of the AAC-24 scoring, Figure 1(a) depicts the anterior and posterior aortic walls. Images (b) and (c) are from our dataset, where (b) reflects the difficulty in localizing the aorta when there is no calcification. Figure 1(c) shows a severe case of AAC: calcific deposits can be observed in each segment. We use the established severity categories: low (AAC-24 score 0 or 1), moderate (score 2–5) or high-risk (score $\geq$ 6) [18, 10, 9].

2.2 Automatic AAC Classification

We now summarize three relevant pieces of work available in the literature, that perform automatic AAC classification based on the overall AAC-24 score per image. Elmasri et al. [4] trained an Active Appearance Model on 20 DXA VFA images to localize the vertebra L1-L4 and the part of aorta adjacent to these vertebra. Next, they fitted this model to 53 test scans to localize the aorta and extract visual features to perform three-class classification using SVM and KNN techniques. They used non-standard class boundaries to classify the test images with an average accuracy of 92.9%. Chaplin et al. [3] followed a similar process as [4] on 195 DXA VFA scans to extract the Region of Interest (ROI), except that, they used a statistical shape model. The ROI in each scan was then warped to straighten the spine which generally leads to loss of information in a calcified aorta. Two separate U-net architectures were then used to segment the calcification in the ROI as a whole, as well as segment-wise (anterior and posterior). They report R^2 coefficient of only 0.68 between ground truth scores and segment-wise scores and $R^2 = 0.58$ with predicted AAC-24 scores for the whole image. Finally, Reid et al. [15] used a battery of CNNs to classify 1100 DXA VFA scans into three classes. They do not perform cross-validation, rather they reported their results from a single train/validate/test run and selected the network that gave them the best results. Their R^2 coefficient between the ground truth and predicted AAC-24 scores (for the whole scan) was 0.86 with an accuracy of 88.1%. It is important to note that none of the methods discussed above produce fine-grained AAC-24 scores.

3 Proposed Framework

Figure 2 (a) shows our proposed framework. It starts with an image pre-processing module (for details see section 4.2) which crops and resizes the images. Once the image is pre-possessed, it is passed on to the visual encoder to extract visual feature maps. We choose a pre-trained Resnet152v2 as the encoder. However, this is not a rigid choice and in future this could be replaced by other/better models. We extract feature maps from the last convolutional layer without using the classification layer of the pre-trained CNN. The feature maps are fed as input to two individual decoders, each of which independently maximizes the log likelihood over the parameter space:

$$\theta^* = \arg\max_{\theta} \sum_{(V,y)} \log p(y|V;\theta) \tag{1}$$

where θ represents the model parameters, V= [v^1, v^2, ..., v^n] represents the visual feature maps extracted from the pre-processed image, and $y = \{y^1, y^2 ...,$ $y^t\}$ is the sequence of segmented scores. The log-likelihood of the joint probability distribution $\log p(y|V;\theta)$ can be decomposed as:

$$\log p(y|V;\theta) = \sum_{(t=1)}^{T} \log p(y^t|y^1, ..., y^{t-1}, V;\theta) \tag{2}$$

We use a Long-Short Term Memory (LSTM) module to generate y, therefore the conditional probability $\log p(y|V)$ (dropping θ for convenience) can be modeled as $\log p(y^t|y^1, ..., y^{t-1}, V) = g(h^t, c^t)$ where g is a nonlinear function, c^t is the context vector and h^t is the hidden state of the LSTM at time t. We can model h^t as $h^t = LSTM(s^t, h^{t-1}, m^{t-1})$ where s^t is the input vector, and h^{t-1} and m^{t-1} are hidden state and memory cell vectors at time t-1, respectively.

To compute the context vector c^t we use an attention module, such that the context is dependent on specific regions in the image (via image feature maps) as well as the decoder outputs. Therefore, c^t can be defined as $c^t = q(V, h^{t-1})$

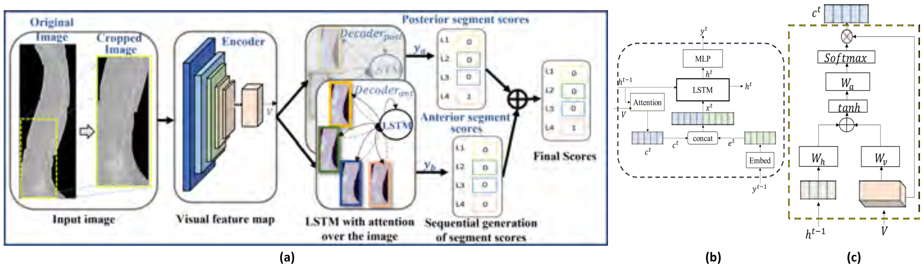


Fig. 2. Our proposed (a) framework for automatic fine-grained AAC scoring,(b) detailed schematic of our attention-based decoders, and (c) attention module. Note that $Decoder_{ant}$ and $Decoder_{post}$ have the same architecture.

where q is the attention function, and h^t is the hidden state of the LSTM at time t. The distribution of attention over the feature maps V (corresponding to various regions of the image) is computed using a feed-forward network and can be formalized as $z^t = W^a \tanh(W^v V + W^h h^{t-1})$ and $\beta^t = softmax(z^t)$ where W^a, W^v and W^h are the learnable parameters, and β is the attention weight over the feature maps V. Finally, c^t can be computed as:

$$c^t = \sum_{i=1}^{n} \beta^{ti} v^{ti} \tag{3}$$

We train our model using weighted cross-entropy loss, where we set the weights for each class based on the data distribution. We do not fine-tune our encoder due the limited size of our dataset. Our two decoders, $Decoder_{ant}$ and $Decoder_{post}$ (see Figure 2(b) and (c)), have similar architecture and are trained independently to maximize the objective function given in Equation 2.

4 Experiments

4.1 Dataset

Our dataset is comprised of randomly selected 1,916 bone-density machine-derived lateral-spine scans, obtained using iDXA GE machines [15] with a resolution of at least 1600 x 300 pixels. The disease severity distribution of the 1,916 scan is: *low risk* 829, *moderate risk* 445 and *high risk* 642. Although, these scans come with expert annotated AAC-24 scores [6], the location of calcified pixels is not annotated on the scans. The distribution of AAC-24 scores in the dataset (see Table 1) is very challenging as it has severe class imbalance. Specifically, this distribution of zero scores is highly skewed for L1 and L2 perhaps because vascular calcification usually starts around L4 and L3 and then progresses upwards [12]. In terms of sequences for anterior segments, our data has 176 unique (out of $4^4 = 256$ possible) combinations but only 29 of them appear more than 10 times. The most frequent sequence is [0,0,0,0], which appears 904 times followed by [0,0,0,1], which appears 77 times. For posterior segments, our data has 190 unique combinations, out of which only 30 appear more than 10 times. Once again, [0,0,0,0] is the most frequent combination and appears 786 (41%) times.

4.2 Pre-processing

To obtain the ROI i.e., the area around the lower lumbar vertebrae, we follow the pre-processing guidelines of Reid et al. [15] and crop 50% from the top, 40% from the left and 10% from the right side of each scan. Then the cropped images are

Score Segment	0	1	2	3
Anterior	5390	1255	518	501
Posterior	5098	1251	685	630

Table 1. Distribution of calcification scores for the anterior and posterior segments. Note the skewed distribution of score '0'.

resized to 900 x 300 pixels using the nearest neighbor interpolation, and re-scaled to values between 0 and 1. We augment the dataset by applying various affine transformations to the images, such as translation [+20, -20], scaling [+20, -20], shear [0.01°, 0.05°] and rotation [+10°,-10°]. We used the TorchVision library for data augmentation and the PyTorch Machine Learning Library for model training and evaluation.

4.3 Model and Training Parameters

We train our model M_{fgs} ("fgs" stands for fine-grained scoring model) with stochastic gradient descent using the Adam optimizer [7]. The initial learning rate and batch size were set to $1e^{-4}$ and 10 respectively. We use ResNet152v2 [5] pretrained on ImageNet as an encoding model (to extract the visual feature maps), but do not fine-tune it on our data. For an input image size of 900 x 300, the size of the extracted feature map is 29 x 10 x 2048. We flatten the feature maps to 290 x 2048 and feed them individually to the two decoding networks, which we term as $Decoder_{ant}$, and $Decoder_{post}$.

The two decoders, are trained independently with sequences of anterior and posterior segment ground truth scores, respectively. Furthermore, after training is complete, the output scores of both decoders (for a given test image) are summed to get a single score corresponding to each lumbar vertebrae. Finally, the scores of L1 -L4 are summed to obtain the AAC-24 scores. Both decoding pipelines are comprised of an LSTM, with a hidden size of 512, and based on an attention module, where the output sequence length is 4. We perform 10-fold stratified cross validation (where the data is split based on the distribution of AAC-24 scores, such that this distribution is maintained across all splits). In each fold 1,724 examples are used to train the network and 192 for validation. We perform early stopping based on the average Pearson correlation between the predicted and ground truth segment scores. We also use dropout (first after the hidden layer of LSTM (alpha=0.5), then another (alpha= 0.4) before the last FC layer) as a regularization strategy [20]. Our trained network and scripts are publicly available [1].

4.4 Evaluation

To the best of our knowledge, this is the first paper that predicts fine-grained AAC-24 scores for each vertebrae; instead of a single score for L1-L4 lumbar regions. Therefore, to compare our results with the state-of-the-art we use the sum of all individual granular scores. Since Reid et al. [15] have analysed the same dataset as ours, we implement their pipeline (albeit with minor modifications) to compare with our results. Following [15], we train a baseline CNN with Resnet152v2 as its encoder. The decoder consists of a global pooling layer, followed by a dense layer with Relu activation, and another dense layer with a

	Low ($n = 829$)		Moderate ($n = 445$)		High ($n = 642$)		Mean	
	M_{base}	M_{fgs}	M_{base}	M_{fgs}	M_{base}	M_{fgs}	M_{base}	M_{fgs}
Accuracy	71.14	**82.52**	62.06	**75.52**	79.12	**87.89**	70.77	**81.98**
Sensitivity	55.49	**86.37**	**59.33**	37.53	**54.83**	80.22	56.55	**68.04**
Specificity	**83.07**	79.58	62.88	**87.02**	91.37	**91.76**	79.11	**86.12**
NPV	70.99	**88.45**	**83.63**	82.16	80.06	**90.20**	78.23	**86.93**
PPV	71.43	**76.33**	32.59	**46.65**	76.19	**83.06**	60.07	**68.68**

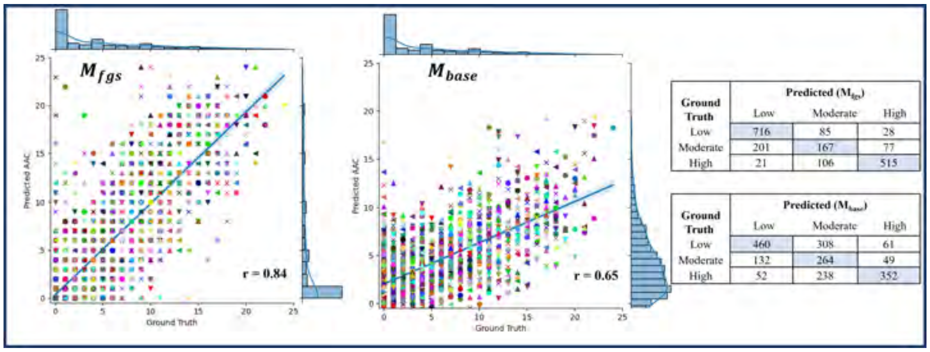
Table 2. Performance comparison of our model with the baseline [15] (NPV is Negative Predictive Value and PPV is Positive Predictive Value) in one-vs-rest setting using the cumulative AAC-24 predicted scores.

linear activation, the same as in [15]. The generated AAC-24 scores are classified into three risk levels, based on the thresholds discussed in Section 2. Note that, for fairness and transparency, we do not report results directly from [15] as we could not obtain their train/validation split and they did not perform ten-fold cross validation. The baseline model M_{base} follows the same stratified cross validation strategy as our proposed M_{fgs} model. As stated (see Section 4.3 and Figure 2), our model has two decoders, for anterior and posterior segments of the lumbar regions L1-L4. It would be natural to ask whether predicting AAC scores in two segments is better than predicting them horizontally across each lumbar region e.g. L1 or L2. To ascertain this, we train a variant of our model (call it M^*_{fgs}) with a single decoder to predict a sequence of scores for each lumbar vertebra, L1-L4, where the score for L1, would be the sum of $L1_{ant}$ and $L1_{post}$. We report our results in the following section.

4.5 Results and Discussion

Table 2 reports one-vs-rest performance of our model M_{fgs} compared to the baseline M_{base} [15] after 10-fold cross validation. Our average classification accuracy $81.98 \pm 2.5\%$ is significantly better than the base line accuracy of $70.77 \pm 3.2\%$. Similarly, our average 3-class classification accuracy is $72.8 \pm 2.9\%$ while that of the baseline is $55.8 \pm 3.2\%$. Note that our AAC-24 scores are obtained by summing up the individual fine-grained scores. Our model predicts AAC-24 scores more accurately compared to the baseline model [15].

To assess the efficacy of our model at a more granular level, we compare the predicted scores with fine-grained ground truth scores. The scatter plots and confusion matrix are shown in Figure 3. Since the baseline model [15] does not



Ground Truth	Predicted (M_{fgs})		
	Low	Moderate	High
Low	716	85	28
Moderate	201	167	77
High	21	106	515

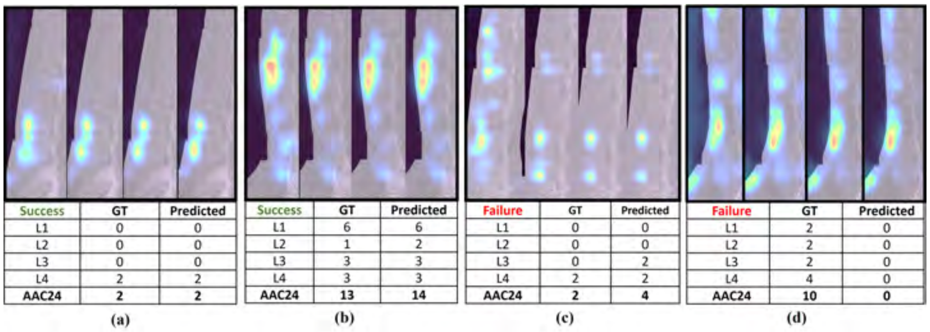
Ground Truth	Predicted (M_{base})		
	Low	Moderate	High
Low	460	308	61
Moderate	132	264	49
High	52	238	352

Fig. 3. Scatter plots and confusion matrix of fine-grained ground truth scores vs predicted scores for our proposed model M_{fgs} and ground truth vs the baseline M_{base} [15] overall AAC-24 score per scan.

	L1		L2		L3		L4	
	M^*_{fgs}	M_{fgs}	M^*_{fgs}	M_{fgs}	M^*_{fgs}	M_{fgs}	M^*_{fgs}	M_{fgs}
Pearson Correlation ↑	0.40	**0.49**	0.56	**0.64**	0.56	**0.70**	0.64	**0.69**
Kendall Correlation ↑	0.34	**0.44**	0.49	**0.56**	0.47	**0.60**	0.54	**0.58**
Mean Absolute Error ↓	0.96	**0.60**	0.73	**0.67**	1.20	**0.88**	1.13	**1.10**

Table 3. Correlation and error metrics between the ground truth and predicted scores for each lumbar segment. Note that $p << 0.001$ for both correlation metrics.

have the capability to perform fine-grained scoring, we compare its output of a single AAC-24 score (for all lumbar regions) with the corresponding ground truth scores. Note that our model is very good at classifying low and high risk patients. The figure provides evidence that fine-grained scoring results in significantly (p«0.01) better prediction and higher correlation with human-scores.



Success	GT	Predicted
L1	0	0
L2	0	0
L3	0	0
L4	2	2
AAC24	**2**	**2**

(a)

Success	GT	Predicted
L1	6	6
L2	1	2
L3	3	3
L4	3	3
AAC24	**13**	**14**

(b)

Failure	GT	Predicted
L1	0	0
L2	0	0
L3	0	2
L4	2	2
AAC24	**2**	**4**

(c)

Failure	GT	Predicted
L1	2	0
L2	2	0
L3	2	0
L4	4	0
AAC24	**10**	**0**

(d)

Fig. 4. Our qualitative results show the attention maps generated by our decoding pipeline by combining the weights of $Decoder_{ant}$ and $Decoder_{post}$ for simplicity.

Table 3 shows the comparison between predicting AAC scores horizontally across each vertebrae vs predicting the scores vertically for each segment (anterior and posterior), i.e. comparison between M^*_{fgs} and M_{fgs}. It makes sense that the two decoders in our model M_{fgs} 'attend' to the two vertical segments and perform better than a model that looks at each vertebrae horizontally. Thus the correlation between human annotated scores of those predicted by M_{fgs} is significantly better (p<0.01) than the correlation produced by our variant M^*_{fgs}.

Figure 4 shows some examples where our model succeeds (a-b) or fails (c-d). The four sub-figures in each section are from four different time stamps of our sequential attention model. The model "sees" a particular vertebrae at a given time stamp, "attends" to it and "detects" the amount of calcification. It then moves on to the next vertebrae in the sequence. Figure 4(a-b) show how the model attends to each vertebrae and correctly scores the calcification. Figure 4(c) shows failure cases where the model over-estimates the score of L3 while (d) portrays a case where it totally fails to identify the heavy calcification. However, Figure 4(d) is very interesting as the aorta in the DXA scan produced by the GE iDXA machine is masked for radiation dose reduction. The human experts have not scored L2 and L3 anterior sections of this scan because they are not visible. Our model is unable to "see" the aorta and hence outputs a zero score. (This is good because a higher predicted score would have meant that the model is not paying attention to the aorta in the score generation process).

5 Conclusion

This is the first work to adapt sequential attention-based models from vision-language domain to address the challenge of fine-grained AAC-24 scoring. This preliminary study on a dataset of 1,916 low resolution DXA scans, not only overcomes the bottlenecks of domain adaptation, but also provides evidence that sequential “fine-grained” scoring yields higher agreement (correlation) with expert human annotated scores. Furthermore, it highlights the necessity of developing larger LFA DXA scan datasets with granular ground truth scores to validate this technique in large population based studies.

Acknowledgement

The work was funded by a National Health and Medical Research Council (NH&MRC) of Australia Ideas grants (APP1183570). The salary of JRL is supported by a National Heart Foundation of Australia Future Leader Fellowship (ID: 102817). The study was approved by the Health Research Ethics Board for the University of Manitoba (HREB H2004:017L, HS20121). The Manitoba Health Information Privacy Committee approved access to the Manitoba data (HIPC 2016/2017–29). The results and conclusions are those of the authors and no official endorsement by Manitoba Health and Seniors Care, or other data providers is intended or should be inferred.

References

1. https://github.com/NaehaSharif/Show-Attend-and-Detect
2. Bernardi, R., Cakici, R., Elliott, D., Erdem, A., Erdem, E., Ikizler-Cinbis, N., Keller, F., Muscat, A., Plank, B.: Automatic description generation from images: A survey of models, datasets, and evaluation measures. Journal of Artificial Intelligence Research **55**, 409–442 (2016)
3. Chaplin, L., Cootes, T.: Automated scoring of aortic calcification in vertebral fracture assessment images. In: Medical Imaging 2019: Computer-Aided Diagnosis. vol. 10950, pp. 811–819. SPIE (2019)
4. Elmasri, K., Hicks, Y., Yang, X., Sun, X., Pettit, R., Evans, W.: Automatic detection and quantification of abdominal aortic calcification in dual energy x-ray absorptiometry. Procedia Computer Science **96**, 1011–1021 (2016)
5. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European Conference on Computer Vision. pp. 630–645. Springer (2016)
6. Kauppila, L.I., Polak, J.F., Cupples, L.A., Hannan, M.T., Kiel, D.P., Wilson, P.W.: New indices to classify location, severity and progression of calcific lesions in the abdominal aorta: a 25-year follow-up study. Atherosclerosis **132**(2), 245–250 (1997)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
8. Leow, K., Szulc, P., Schousboe, J.T., Kiel, D.P., Teixeira-Pinto, A., Shaikh, H., Sawang, M., Sim, M., Bondonno, N., Hodgson, J.M., et al.: Prognostic value of abdominal aortic calcification: a systematic review and meta-analysis of observational studies. Journal of the American Heart Association **10**(2), e017205 (2021)
9. Lewis, J.R., Eggermont, C.J., Schousboe, J.T., Lim, W.H., Wong, G., Khoo, B., Sim, M., Yu, M., Ueland, T., Bollerslev, J., et al.: Association between abdominal aortic calcification, bone mineral density, and fracture in older women. Journal of Bone and Mineral Research **34**(11), 2052–2060 (2019)

10. Lewis, J.R., Schousboe, J.T., Lim, W.H., Wong, G., Wilson, K.E., Zhu, K., Thompson, P.L., Kiel, D.P., Prince, R.L.: Long-term atherosclerotic vascular disease risk and prognosis in elderly women with abdominal aortic calcification on lateral spine images captured during bone density testing: a prospective study. Journal of Bone and Mineral Research **33**(6), 1001–1010 (2018)
11. Lewis, J.R., Schousboe, J.T., Lim, W.H., Wong, G., Zhu, K., Lim, E.M., Wilson, K.E., Thompson, P.L., Kiel, D.P., Prince, R.L.: Abdominal aortic calcification identified on lateral spine images from bone densitometers are a marker of generalized atherosclerosis in elderly women. Arteriosclerosis, thrombosis, and vascular biology **36**(1), 166–173 (2016)
12. Lillemark, L., Ganz, M., Barascuk, N., Dam, E.B., Nielsen, M.: Growth patterns of abdominal atherosclerotic calcified deposits from lumbar lateral x-rays. The international journal of cardiovascular imaging **26**(7), 751–761 (2010)
13. Pickhardt, P.J., Graffy, P.M., Zea, R., Lee, S.J., Liu, J., Sandfort, V., Summers, R.M.: Automated CT biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: a retrospective cohort study. The Lancet Digital Health **2**(4), e192–e200 (2020)
14. Radavelli-Bagatini, S., Bondonno, C.P., Sim, M., Blekkenhorst, L.C., Anokye, R., Connolly, E., Bondonno, N.P., Schousboe, J.T., Woodman, R.J., Zhu, K., et al.: Modification of diet, exercise and lifestyle (model) study: a randomised controlled trial protocol. BMJ open **10**(11), e036366 (2020)
15. Reid, S., Schousboe, J.T., Kimelman, D., Monchka, B.A., Jozani, M.J., Leslie, W.D.: Machine learning for automated abdominal aortic calcification scoring of dxa vertebral fracture assessment images: A pilot study. Bone **148**, 115943 (2021)
16. Roth, G.A., Mensah, G.A., Johnson, C.O., Addolorato, G., Ammirati, E., Baddour, L.M., Barengo, N.C., Beaton, A.Z., Benjamin, E.J., Benziger, C.P., et al.: Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. Journal of the American College of Cardiology **76**(25), 2982–3021 (2020)
17. Schousboe, J.T., Lewis, J.R., Kiel, D.P.: Abdominal aortic calcification on dual-energy x-ray absorptiometry: methods of assessment and clinical significance. Bone **104**, 91–100 (2017)
18. Schousboe, J.T., Taylor, B.C., Kiel, D.P., Ensrud, K.E., Wilson, K.E., McCloskey, E.V.: Abdominal aortic calcification detected on lateral spine images from a bone densitometer predicts incident myocardial infarction or stroke in older women. Journal of Bone and Mineral Research **23**(3), 409–416 (2008)
19. Schousboe, J.T., Wilson, K.E., Kiel, D.P.: Detection of abdominal aortic calcification with lateral spine imaging using dxa. Journal of Clinical Densitometry **9**(3), 302–308 (2006)
20. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research **15**(1), 1929–1958 (2014)
21. Strong, J.P., Malcom, G.T., McMahan, C.A., Tracy, R.E., Newman III, W.P., Herderick, E.E., Cornhill, J.F., of Atherosclerosis in Youth Research Group, P.D., of Atherosclerosis in Youth Research Group, P.D., et al.: Prevalence and extent of atherosclerosis in adolescents and young adults: implications for prevention from the pathobiological determinants of atherosclerosis in youth study. Jama **281**(8), 727–735 (1999)