

Supplementary Material: A Hybrid CNN-Transformer Feature Pyramid Network for Granular Abdominal Aortic Calcification Detection from DXA Images

Anonymous MICCAI 2024 Submission - 1214

No Institute Given

Table 1. Ablation Study: Effect of CNN backbone type on model’s performance. The evaluation metric here is the Pearson Correlation Coefficient.

CNN Backbone	L1A	L1P	L2A	L2P	L3A	L3P	L4A	L4P
ResNet152v2 FPN + EFFM + DRSAT	0.56	0.62	0.63	0.72	0.70	0.72	0.72	0.74
DenseNet121 FPN + EFFM + DRSAT	0.60	0.62	0.62	0.69	0.72	0.74	0.72	0.75
EfficientNetV2S FPN + EFFM + DRSAT	0.64	0.68	0.65	0.73	0.74	0.76	0.75	0.78

Table 2. Granular Level Analysis on DE DXA VFA Dataset from GE iDXA Machine: Correlation and Error metrics comparison between the ground truth and predicted scores for abdominal aortic sections adjacent to vertebrae L1 to L4.

	Method	Accuracy	Sensitivity	Specificity	NPV	PPV
Low (n=829)	Gilani et al. [1]	82.52	86.37	79.58	88.45	76.33
	Saleem et al. [2]	86.01	86.61	85.55	89.33	82.05
	Proposed	87.78	85.64	89.42	89.09	86.06
Moederate (n=445)	Gilani et al. [1]	75.52	37.53	87.02	82.16	46.65
	Saleem et al. [2]	77.92	55.50	84.70	86.28	52.33
	Proposed	80.01	65.17	84.50	88.91	55.98
High (n=642)	Gilani et al. [2] [1]	87.89	80.22	91.76	90.20	83.06
	Saleem et al. [2]	89.50	78.66	94.97	89.82	88.75
	Proposed	90.03	79.75	95.21	90.32	89.35
Overall	Gilani et al. [1]	81.98	68.04	86.12	86.93	68.68
	Saleem et al. [2]	84.48	73.59	88.41	88.48	74.38
	Proposed	85.94	76.85	89.71	89.44	77.13

In the works of Gilani et al., [1] and [2], one of the metrics to measure the model’s performance is the three-class classification (one-vs-all) accuracy measured using an overall AAC score (range: 0 to 24). The defined classes are Low-Risk (AAC score: 0 – 1), Medium-Risk (AAC score: 2 – 5), and High-Risk (AAC 6 – 24). For overall AAC scores different approaches have been used i.e. in [2], the overall AAC score is predicted using a single regression output and there are no granular scores, whereas in the case of [1], as the model predicts granular scores only, the overall AAC score is calculated using the sum of granular scores. As we are predicting granular scores, so for a fair comparison, we retrained the model of Saleem et al. [2] under granular output settings and reported the cumulative AAC score based on the summation of granular scores for it. The

three-class classification results calculated based on the cumulative AAC score for all three models are shown in Table 2. Our proposed framework performed well compared to previous works. It is worth noting that the three-class accuracy calculated using the sum of granular scores is less than the accuracy calculated using the overall AAC score predicted using a single regression output.

Table 3. Comparative Granular Level Analysis of the proposed framework with the baseline model on 3 different DXA VFA datasets. The baseline is the FPN with CNN-Backbone (EfficientNetV2S).

DE DXA Dataset GE iDXA (1,916 scans)									
Evaluation Metric	Method	L1		L2		L3		L4	
		Ant	Post	Ant	Post	Ant	Post	Ant	Post
Pearson Correlation \uparrow	Baseline	0.55	0.61	0.59	0.65	0.71	0.69	0.71	0.73
	Hybrid-FPN-AACNet	0.64	0.68	0.65	0.73	0.75	0.76	0.76	0.78
Kendall Tau \uparrow	Baseline	0.36	0.40	0.45	0.48	0.51	0.55	0.51	0.57
	Hybrid-FPN-AACNet	0.42	0.46	0.48	0.52	0.57	0.59	0.56	0.61
Mean Absolute Error \downarrow	Baseline	0.35	0.37	0.38	0.39	0.51	0.53	0.52	0.57
	Hybrid-FPN-AACNet	0.29	0.32	0.35	0.36	0.44	0.45	0.43	0.49
SE DXA Dataset GE iDXA (1,916 scans)									
Evaluation Metric	Method	L1		L2		L3		L4	
		Ant	Post	Ant	Post	Ant	Post	Ant	Post
Pearson Correlation \uparrow	Baseline	0.57	0.62	0.61	0.68	0.72	0.72	0.73	0.75
	Hybrid-FPN-AACNet	0.66	0.69	0.66	0.76	0.74	0.78	0.74	0.79
Kendall Tau \uparrow	Baseline	0.38	0.42	0.47	0.50	0.54	0.58	0.52	0.58
	Hybrid-FPN-AACNet	0.44	0.45	0.49	0.54	0.56	0.60	0.54	0.62
Mean Absolute Error \downarrow	Baseline	0.32	0.34	0.36	0.36	0.49	0.50	0.49	0.53
	Hybrid-FPN-AACNet	0.28	0.31	0.33	0.33	0.43	0.44	0.43	0.47
SE DXA Dataset Hologic (508 scans)									
Evaluation Metric	Method	L1		L2		L3		L4	
		Ant	Post	Ant	Post	Ant	Post	Ant	Post
Pearson Correlation \uparrow	Baseline	0.16	0.45	0.38	0.52	0.53	0.57	0.54	0.56
	Hybrid-FPN-AACNet	0.25	0.53	0.44	0.61	0.62	0.65	0.57	0.59
Kendall Tau \uparrow	Baseline	0.09	0.17	0.24	0.29	0.28	0.33	0.33	0.35
	Hybrid-FPN-AACNet	0.16	0.22	0.27	0.34	0.33	0.38	0.34	0.37
Mean Absolute Error \downarrow	Baseline	0.24	0.23	0.25	0.24	0.36	0.38	0.45	0.52
	Hybrid-FPN-AACNet	0.19	0.14	0.20	0.20	0.32	0.31	0.39	0.45

References

1. Gilani, S.Z., Sharif, N., Suter, D., Schousboe, J.T., Reid, S., Leslie, W.D., Lewis, J.R.: Show, Attend and Detect: Towards Fine-Grained Assessment of Abdominal Aortic Calcification on Vertebral Fracture Assessment Scans. In: 2022 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 439-450 (2022).
2. Saleem, A., Ilyas, Z., Suter, D., Hassan, G.M., Reid, S., Schousboe, J.T., Prince, R., Leslie, W.D., Lewis, J.R., Gilani, S.Z.: SCOL: Supervised Contrastive Ordinal Loss for Abdominal Aortic Calcification Scoring on Vertebral Fracture Assessment Scans. In: 2023 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) pp. 273-283 (2023).