SynSeq: End-to-End SYNTAX Score Prediction from Coronary Angiography Videos

Christoph Baumann[1,2], Ronny Schweitzer[3], Noemi Pavo[3], Ulrike Attenberger[4], Christian Loewe[4], and Philipp Seeböck[1,2]

[1] MANO Group, Computational Imaging Research Lab, Medical University of Vienna, Austria
[2] Comprehensive Center for Artificial Intelligence in Medicine, Medical University of Vienna, Austria
[3] Department of Cardiology, Medical University of Vienna, Austria
[4] Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Austria
{christoph.a.baumann,philipp.seeboeck}@meduniwien.ac.at

Abstract. The SYNTAX score is an established tool for assessing coronary artery disease and guiding revascularization treatment decisions. However, its manual estimation from coronary angiography videos by clinical experts is time-consuming and subject to inter-reader variability. While machine learning has shown promise in automating this process, prior work has primarily focused on lesion detection, characterization, or binary disease classification, leaving direct SYNTAX score prediction relatively unexplored. We propose *SynSeq*, a video-based method for direct SYNTAX score prediction. It combines targeted preprocessing with a tailored training strategy using a zero-inflation-aware loss and linear target scaling. Evaluated on the public CardioSyntax dataset, SynSeq significantly outperforms previous state-of-the-art methods, improving R^2 by 0.55, reducing prediction bias by 93.1% and achieving more consistent performance across annotations from three independent expert graders. In addition, SynSeq achieves a weighted F_1-score of 0.80 for revascularization treatment recommendations, slightly below inter-expert agreement. These results demonstrate the potential of SynSeq to provide consistent, automated SYNTAX score assessment and reliable decision support for coronary revascularization planning.

Keywords: Coronary angiography · SYNTAX score prediction · Multi-view video regression

1 Introduction

Coronary artery disease (CAD) remains a leading cause of mortality worldwide, requiring precise diagnosis and management to optimize patient outcomes. When managing complex CAD, clinicians rely on the SYNTAX score as the established tool to grade lesion severity, complexity, and location from invasive coronary angiography (ICA) examinations [14]. Scores above a critical threshold ($\geq$33)

strongly indicate superior long-term survival with coronary artery bypass grafting (CABG) over percutaneous coronary intervention (PCI) [11]. Recognizing its predictive power, the recent European Society of Cardiology (ESC) guidelines highlight that integrating machine learning into this workflow could significantly streamline the revascularization selection process and improve prognostic assessments [17]. Despite its clinical utility, widespread adoption of the SYNTAX score remains bottlenecked. Manual calculation is limited by inter-observer variability due to anatomical ambiguities and is time-consuming, reducing its feasibility in everyday clinical practice [3,4]. In this work, we address this gap by introducing an automated, end-to-end framework capable of predicting patient-level SYNTAX scores directly from multi-view coronary angiography videos.

Related Work Prior automated coronary angiography efforts have focused on sub-tasks like vessel segmentation, lesion detection, and stenosis quantification rather than full SYNTAX score estimation [2,8,9,15,18]. While these methods have demonstrated promising performance on their respective tasks, they address only individual components of the reasoning required for SYNTAX score assessment. Moreover, the majority of prior work relies on single-frame datasets [10,13], limiting their ability to exploit the temporal and multi-view information inherent to angiography studies. Recent methods have leveraged full video sequences to exploit temporal and multi-view projections. CathAI [1] introduces an end-to-end pipeline for automated ICA interpretation and stenosis estimation. DeepCoro [7] and DeepCORO-CLIP [5] further advanced video-based representation learning through large-scale spatiotemporal and multi-view pretraining. However, these still focus primarily on stenosis localization and anatomical characterization rather than direct SYNTAX score estimation, which requires integrating lesion severity, vessel dominance, and anatomical complexity. The CardioSyntax benchmark [12] is the first large-scale dataset for end-to-end SYNTAX score prediction from multi-view videos. While the accompanying baseline models include continuous score prediction, they primarily emphasize zero-versus-nonzero discrimination (healthy versus diseased). While useful for screening, this provides limited insight into high-complexity cases where clinical decision-making between PCI and CABG is most critical. Furthermore, the zero-inflation and long-tailed distribution properties of the SYNTAX score task remain unaddressed, limiting performance on highly diseased, underrepresented cases. Our work directly targets these challenges, focusing on robust continuous score prediction across the full range of SYNTAX values.

Contribution We introduce *SynSeq*, a video-based framework for automated SYNTAX score prediction that focuses on directly optimizing the continuous scoring objective. We build SynSeq on the previously introduced Multi-View-Look (MVL) method [12]. Our main contributions are: (i) We propose targeted preprocessing for multi-view coronary angiography videos to enhance feature extraction from complex temporal and cross-view inputs. (ii) We design a tailored optimization strategy with two distinct adjustments to handle the highly skewed data distribution: (1) a sample-weighting penalty function to counteract skewed

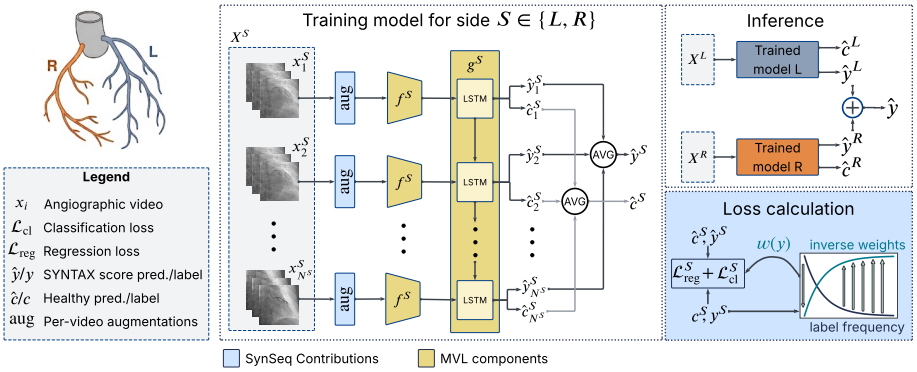

Fig. 1: Overview of the proposed *SynSeq* method. Coronary angiographic video sequences of the left (X^L) and right (X^R) vessel tree are processed by two independent branches with identical architecture but separate weights. Each sequence is encoded by a 3D-ResNet (f^S), with resulting features being fed into a LSTM-based dual-head predictor (g^S), estimating a SYNTAX score ($\hat{y}^S$) and vessel status ($\hat{c}^S$) per branch. Inverse frequency weighted losses ($\mathcal{L}^S_{\text{reg}}$, $\mathcal{L}^S_{\text{cl}}$) are computed independently for each branch. During inference, summing left and right predictions results in a patient-level SYNTAX score $\hat{y} = \hat{y}^L + \hat{y}^R$. The temporary head h used in pre-training is omitted.

label distribution, and (2) a linear rather than logarithmic target scaling to preserve score granularity in high-severity cases. (iii) We present a rigorous evaluation on the full CardioSyntax dataset, demonstrating significantly improved performance over prior state-of-the-art methods and increased robustness across multiple expert annotations.

2 Method

Our method *SynSeq* predicts the patient-level ***SYN***TAX score $\hat{y} \in \mathbb{R}$ from an arbitrary number of multi-view angiographic video ***seq***uences X, using a backbone network f and prediction head g (Figure 1). For each patient, SynSeq processes left and right coronary artery trees independently, with the corresponding sets of angiography video sequences denoted as X^S for $S \in L, R$. The left and right sets are defined as $X^L = \{x_i^L\}_{i=1}^{N^L}$ and $X^R = \{x_i^R\}_{i=1}^{N^R}$, where N^L and N^R denote the number of video sequences per side. Each set contains at least one video and is treated as an unordered collection. Each coronary artery tree is processed by an independent 3D-ResNet backbone f^S with parameters θ_f^S to extract per-video feature representations, followed by a dual-head LSTM g^S with parameters θ_g^S to obtain predictions for each video. The LSTM outputs (i) a scalar estimate of the SYNTAX sub-score and (ii) a binary logit indicating the probability of disease presence (SYNTAX score > 0). To account for varying numbers of acquisitions,

we obtain $(\hat{y}_i^S, \hat{c}_i^S) = g^S(f^S(x_i^S))$ and compute the side-specific predictions by averaging:

$$(\hat{y}^S, \hat{c}^S) = \frac{1}{N^S} \sum_{i=1}^{N^S} (\hat{y}_i^S, \hat{c}_i^S). \tag{1}$$

During inference, the final patient-level SYNTAX score is obtained by summing contributions from both coronary trees, with negative values clamped to zero:

$$\hat{y} = \max(0, \hat{y}^L) + \max(0, \hat{y}^R) \tag{2}$$

Inputs & Preprocessing To simulate realistic acquisition artifacts, domain tailored data augmentation is applied during training. Random spatial cropping is performed with a relative crop area $\kappa \in [\kappa_{\min}, \kappa_{\max}]$ and an aspect ratio $\eta \in [\eta_{\min}, \eta_{\max}]$. Minor video rotations are sampled from $\rho \in [-\rho_{\max}, \rho_{\max}]$, and brightness and contrast are adjusted by a multiplicative factor δ_{int}. The backbone input is reduced to a single channel for grayscale angiography, using training-set mean μ and standard deviation σ for input normalization. Unlike standard three-channel color image augmentations as used in [12], these adjustments are tailored to X-ray angiography to mimic real-life acquisition variance. Our method assumes each patient has at least one video for both the left and right coronary vessel trees, labeled by artery side and grouped by patient ID. For all videos, a temporal center crop of length τ is used; sequences shorter than τ are padded via temporal repetition.

Continuous long-tailed distribution handling We explicitly address the zero-inflated and skewed distribution of SYNTAX scores, characterized by a high prevalence of healthy cases ($y_i^S = 0$) and a sparse distribution of diseased samples ($y_i^S \geq 1$). Specifically, each sample is assigned a weight inversely proportional to the frequency of its ground-truth sub-score y_i^S:

$$w(y_i^S) = \left(\sum_{j=1}^{T} \mathbb{I}\left(\lfloor y_j^S \rfloor = \lfloor y_i^S \rfloor \right) \right)^{-1} \tag{3}$$

where the function $\mathbb{I}$ is the identity function that maps the boolean input to 0 or 1 and T denotes the total number of training samples of all patients for a given side S.

Loss functions SynSeq is trained using a multi-task objective combining regression and binary classification. The regression head uses a sample-weighted Mean Squared Error (MSE) loss, with targets scaled to $[0, 1]$ via a normalization parameter λ that is utilized instead of logarithmic scaling used in [12]

$$\mathcal{L}_{\text{reg}}^S(\hat{y}^S, y^S) = \frac{1}{M} \sum_{k=1}^{M} w(y_k^S) \cdot \left(\hat{y}_k^S - \frac{y_k^S}{\lambda} \right)^2 \tag{4}$$

where M denotes the total number of samples in the respective mini-batch. In addition, the classification heads predict whether a system is diseased ($c_k^S = 1$ if $y_k^S > 0$, else 0) via a sample-weighted Binary Cross-Entropy loss:

$$\mathcal{L}_{\text{cl}}^S(\hat{c}^S, c^S) = -\frac{1}{M}\sum_{k=1}^{M} w(y_k^S) \cdot \left[c_k^S \log \sigma(\hat{c}_k^S) + (1 - c_k^S)\log(1 - \sigma(\hat{c}_k^S))\right] \tag{5}$$

The final unified multi-task loss for each coronary artery tree ($S \in \{L, R\}$) combines both losses, balanced by hyperparameter α:

$$\mathcal{L}^L = \mathcal{L}_{\text{reg}}^L + \alpha\mathcal{L}_{\text{cl}}^L, \quad \mathcal{L}^R = \mathcal{L}_{\text{reg}}^R + \alpha\mathcal{L}_{\text{cl}}^R \tag{6}$$

Training procedure For each coronary arterial tree L and R, networks f^S and g^S are trained in separate optimization loops. Parameters θ_f^L, θ_g^L and θ_f^R, θ_g^R are learned using the corresponding left- and right-branch losses. In both cases, training follows a two-stage procedure, conducted independently for each tree. First, the backbone network f^S is pretrained on a per-video classification task. A temporary linear head h^S maps the backbone features to a single logit predicting the health probability for each side $\hat{c}_{\text{pre}}^S = h^S\left(f^S(x_i^S)\right)$ and is trained with the binary cross entropy-loss $\mathcal{L}_{\text{cl}}^S(\hat{c}_{\text{pre}}^S, c^S)$. Second, h^S is discarded and the full architecture is trained by optimizing the dual-head LSTM g^S while the backbone f^S is frozen, followed by a joint fine-tuning of both. The sub-scores are aggregated only during inference to yield the final patient-level SYNTAX score. This overall training procedure recreates the training steps from the MVL method [12]. The implementation will be released at https://github.com/cirmuw/SynSeq.

3 Experimental Setup

Data We use the publicly available CardioSyntax dataset [12] for our experiments. The dataset comprises 1,844 patients, providing a total of 9,590 left coronary artery (LCA) and 3,970 right coronary artery (RCA) angiographic videos. Each video has a resolution of 512×512 pixels at 15 frames per second, with lengths ranging from 2 to 355 frames (90% between 26 and 77). The ground-truth SYNTAX sub-scores are provided for the LCA and RCA separately by a single primary interventional cardiologist. For a subset of 60 patients, consensus labels from two additional independent interventional cardiologist experts are provided within the dataset repository.

Experiments We evaluate SynSeq against the Multi-View-Look (MVL) baseline [12], a baseline based on the domain-pretrained DeepRV model [2] that utilizes a single view per patient and is not fine-tuned beyond linear probing in our experiments, and the naive mean baseline for better context. To ensure reproducibility, we establish a strict patient-wise split: 60 multi-expert annotated patients are held out as an independent test set, and the remaining data is partitioned via 5-fold cross-validation. Only for final performance evaluation, models

from each fold are evaluated on the test set, with performance reported as mean ± standard deviation across the folds.

We assess regression performance using the coefficient of determination (R^2), mean absolute error (MAE), and mean error (bias). R^2 is also provided for inter-expert agreement. To determine if our method significantly reduces MAE and bias magnitude compared to baselines, a one-tailed Wilcoxon signed-rank test is performed on patient-level ensemble predictions, obtained by averaging outputs from the five-fold models. Particularly this is done for the sample-level metrics (MAE, bias), while R^2 is excluded from significance testing as it is an aggregate dataset-level metric. Additionally, we evaluate the revascularization treatment recommendation performance using the weighted F_1-score computed over the standard risk categories: low ($\leq$22), intermediate (23–32), and high ($\geq$33) risk. We report the (macro) precision, recall, F1-score and accuracy as mean ± standard deviation across the folds. A boundary analysis of misclassified patients describes the deviance in terms of distance between the predicted SYNTAX score and the correct class. Furthermore, we provide a Bland-Altman analysis to investigate systematic errors. We also report the binary classification performance of the classification head after pretraining and in the final model.

To assess the sensitivity of our performance estimates to data partitioning, we evaluate two partition schemes: The 60-patient consensus test split (Split I) and an 80/20 SYNTAX-stratified split of the full cohort with these patients in the evaluation set (Split II). Results use Split I unless stated otherwise.

Ablations To isolate design effects, we evaluate individual components by cumulatively removing them from the proposed method and measuring changes in R^2, MAE, and bias. Specifically, we analyze: (i) the impact of long-tailed distribution handling via sample-weighting penalty removal in the loss (SW); (ii) the effect of optimizing in log-space instead of linear target space ($LinS$); and (iii) the contribution of domain-specific angiography video preprocessing ($APre$).

Implementation Details SynSeq uses a Kinetics-400 pretrained 3D-ResNet18 (R3D-18) backbone f [6,16] mapping videos to 512-D embeddings, and a single-layer projected LSTM g (512 input, 128 hidden) directly outputting $\hat{y}^S$ and $\hat{c}^S$. All stages use Adam (batch size 4, 30 epochs) and a OneCycle learning rate (LR) scheduler. Training follows two phases: (1) pretraining f (max LR 10^{-4}), and (2) training g with f frozen (max LR 10^{-4}), followed by joint end-to-end fine-tuning with f unfrozen (max LR 10^{-5}). Here, we followed the hyper-parameter configuration of the MVL method [12]. Other parameters include augmentation settings ($\kappa_{\min} = 0.9$, $\kappa_{\max} = 1$, $\eta_{\min} = 0.9$, $\eta_{\max} = 1.1$, $\rho_{\max} = 5°$, $\delta_{\text{int}} = 0.1$), normalization constants ($\mu = 0.55871$, $\sigma = 0.151$), loss weight $\alpha = 1.0$, crop length $\tau = 32$, and scale factor $\lambda = 70$.

4 Results

SynSeq clearly outperforms the baselines across all metrics and evaluation scenarios. Table 1 shows regression performance, with SynSeq achieving an R^2 of

Table 1: Regression performance on the test set using Expert 1 labels. $\bar{y}$ denotes the naive mean baseline. Bold: best per column.

	All samples			**Only SYNTAX > 0**		
Method	$R^2 \uparrow$	bias	MAE $\downarrow$	$R^2 \uparrow$	bias	MAE $\downarrow$
$\bar{y}$	-0.18	-6.92	12.23	-0.40	-10.40	13.53
DeepRV [2]	$-0.11_{\pm 0.11}$	$0.78_{\pm 3.23}$	$13.88_{\pm 1.49}$	$-0.17_{\pm 0.1}$	$-2.28_{\pm 3.16}$	$14.09_{\pm 1.07}$
MVL [12]	$0.07_{\pm 0.07}$	$-8.29_{\pm 0.63}$	$9.62_{\pm 0.59}$	$-0.14_{\pm 0.08}$	$-10.47_{\pm 0.80}$	$11.90_{\pm 0.72}$
SynSeq	$\mathbf{0.62_{\pm 0.06}}$	$\mathbf{-0.57_{\pm 1.72}}$	$\mathbf{6.74_{\pm 0.29}}$	$\mathbf{0.54_{\pm 0.07}}$	$\mathbf{-0.90_{\pm 2.14}}$	$\mathbf{8.24_{\pm 0.45}}$

Table 2: Model performance in R^2 across three independent experts scores.

	Exp. 1	**Exp. 2**	**Exp. 3**
$\bar{y}$	-0.18	-0.21	-0.33
DeepRV [2]	$-0.11_{\pm 0.11}$	$-0.03_{\pm 0.12}$	$-0.07_{\pm 0.13}$
MVL [12]	$0.07_{\pm 0.07}$	$0.03_{\pm 0.07}$	$-0.10_{\pm 0.04}$
SynSeq	$\mathbf{0.62_{\pm 0.06}}$	$\mathbf{0.60_{\pm 0.06}}$	$\mathbf{0.60_{\pm 0.10}}$

0.62 compared to 0.07 for MVL [12] and -0.11 for DeepRV [2]. The gap widens on diseased samples (SYNTAX score > 0), with R^2 0.54 vs. $-0.14/-0.17$, bias -0.90 vs. $-10.47/-2.28$, and MAE 8.24 vs. 11.90/14.09. The improvement is consistent across folds and statistically significant (bias: $p<0.001$, MAE: $p<0.05$). This demonstrates reduced systematic bias of SynSeq, with substantially less underestimation of SYNTAX score compared to both MVL and DeepRV, which is also reflected in the scatter plot in Fig. 2(a). Split II with the extended evaluation set gives the same picture (SynSeq/MVL: R^2 0.61/0.15, bias -1.21/-8.03 , MAE 6.82/9.13), indicating stability of performance estimates with respect to data partitioning. Fig. 2(a) further illustrates revascularization treatment recommendations performance for SynSeq and MVL, with decision boundaries separating the low-, intermediate- and high-risk categories. SynSeq improves classification performance over the MVL baseline (weighted F_1: 0.80 ± 0.04 vs. 0.64 ± 0.05), with more predictions falling into the correct risk category (Fig. 2(a)). Its performance is slightly below inter-expert agreement, with weighted F_1-scores of 0.85 between experts 1 and 2, and 0.82 between experts 1 and 3.

SynSeq misclassified 11 of 60 patients across the clinical SYNTAX cutoffs ($\leq$ 22, 23-32, $\geq$ 33), compared to 17 of 60 for MVL. Its errors were both smaller (4.2 vs. 9.8 distance in SYNTAX points from the true class boundary on average, over misclassified patients) and less consequential: only 3 crossed multiple decision thresholds in a way that would alter treatment strategy, versus 12 for MVL. Bland–Altman analysis supports this with a lower mean bias (0.57 vs. 8.29) and narrower 95% limits of agreement ([-17.76, 18.89] vs. [-17.34, 33.92]).

Table 2 summarizes performance across all of the three independent experts. SynSeq achieves consistent R^2 of 0.60 to 0.62 across all experts, indicating ro-

Table 3: Ablation analysis of SynSeq components evaluated against Expert 1.

Method	SW	LinS	APre	R^2 ↑	bias	MAE ↓
SynSeq	✓	✓	✓	$0.62_{\pm 0.06}$	$-0.57_{\pm 1.72}$	$6.74_{\pm 0.29}$
$\text{SynSeq}_{LinS-APre}$	X	✓	✓	$0.50_{\pm 0.04}$	$-4.58_{\pm 0.27}$	$7.27_{\pm 0.23}$
SynSeq_{LinS}	X	✓	X	$0.49_{\pm 0.05}$	$-1.81_{\pm 1.31}$	$7.84_{\pm 0.49}$
SynSeq_{APre}	X	X	✓	$0.38_{\pm 0.05}$	$-6.59_{\pm 0.48}$	$7.84_{\pm 0.29}$
SynSeq_{SW}	✓	X	X	$0.07_{\pm 0.34}$	$-7.09_{\pm 3.68}$	$9.58_{\pm 1.84}$

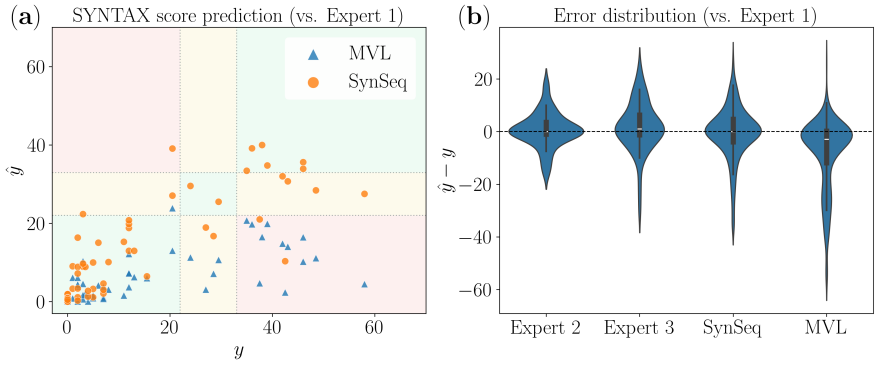

Fig. 2: Performance comparison with expert 1 as reference. (a) Scatter plot of predicted $\hat{y}$ vs expert 1 SYNTAX scores y, with revascularization zones (green: agreement; yellow/red: minor/major disagreement). (b) Violin plot of SYNTAX errors $\hat{y} - y$ for MVL, SynSeq, and experts 2&3 relative to expert 1.

bustness to inter-rater variability. Although these values are below the highest inter-expert agreement ($R^2 = 0.83$ between experts 1 and 2), they are close to the agreement observed between experts 1 and 3 ($R^2 = 0.65$) and experts 2 and 3 ($R^2 = 0.67$). Figure 2(b) supports these findings by showing reduced error variability for SynSeq relative to the MVL baseline, approaching the variability observed between human experts.

Table 3 shows the SynSeq ablation results, demonstrating the relevance of all proposed components. Removing sample-weighting in the loss (SW) substantially increases systematic bias while degrading MAE and R^2. This drop is also reflected in risk category classification performance, with weighted F1-score reduction from 0.8 to 0.68. Replacing the linear optimization objective with a log-transformed objective ($LinS$) also increases bias and reduces R^2. Removing the preprocessing adaptations ($APre$) leads to a substantial deterioration across all evaluation metrics as well. Zero-clamping introduced in Eq. 2 changes the result slightly, with an average of 16.6 samples out of the 60 test samples being clamped for each side across folds. If zero-clamping is not applied, R^2 decreases to 0.58 ± 0.07, bias increases to -2.15 ± 1.23 and MAE increases to 7.87 ± 1.17.

The separate binary classification head (healthy vs. diseased) achieved a macro F1-score of 0.727 ± 0.008 and an accuracy of 0.735 ± 0.008 after pretraining, increasing to 0.747 ± 0.047 macro F1 and 0.783 ± 0.051 accuracy after full training. The MVL baseline shows the same trend, but remains below these values (0.703 ± 0.020 macro F1, 0.743 ± 0.023 accuracy).

5 Discussion and Conclusion

SynSeq demonstrates strong agreement with expert-derived SYNTAX scores and achieves performance approaching inter-expert variability for clinically relevant revascularization decisions. The model maintains stable performance across three independent expert annotators despite being trained using labels from a single expert. This suggests that SynSeq learns a robust representation that is aligned with underlying angiographic patterns rather than overfitting to individual annotation styles. Compared to the MVL baseline [12], SynSeq consistently reduces error and improves agreement with expert annotations. A key observation is that model improvements emerge from the interaction of architecture design, target formulation, and domain-specific preprocessing. First, optimizing in log-space was inferior to linear-space training, which is consistent with the additive structure of the SYNTAX score. Since the SYNTAX score is constructed as a cumulative severity index, linear-space regression better preserves proportional differences between cases, whereas log-space optimization may compress clinically relevant separations. Second, explicit sample-weighting reduces systematic underestimation, which is particularly important in clinical decision thresholds where small score shifts can alter CABG versus PCI recommendations. Third, coronary angiography videos exhibit substantial variability due to projection angles, motion, contrast dispersion, and field-of-view differences. We hypothesize that domain-specific preprocessing and augmentation leads to a more consistent representation of vessel structures across frames and promotes invariance to task-irrelevant factors.

Limitations CardioSyntax lacks an official split, limiting direct comparability with its reported results [12]. The original study reports 5-fold cross-validation without a separate held-out test set, thus the evaluation protocol may differ from ours. Moreover, while we used a patient-level split, we cannot determine whether the original evaluation did as well. The reported metrics therefore seem to estimate different quantities, so we do not read them as directly comparable. To address this, we perform cross-validation across multiple seeds, define a held-out test cohort scored by all three experts, and confirm stability of performance estimates across data partitions.

Another limitation is the small number of baselines, as model weights and training data are not publicly available for CathAI [1], DeepCoro [7] and DeepCoroCLIP [5]. Natural clinical prevalence causes class imbalance affecting model calibration and performance, and our results show that applying explicit mitigation techniques is essential for automated SYNTAX score prediction. Finally,

although SynSeq shows robust performance in the public CardioSyntax dataset, external validation across institutions and imaging protocols is required to assess generalizability beyond the current dataset.

Conclusion This work presents SynSeq, a method for predicting SYNTAX scores directly from multi-view coronary angiography. The results indicate that accurate and stable prediction requires (i) domain-specific angiography preprocessing to reduce acquisition variability and enforce invariance to task-irrelevant factors, (ii) optimization in a linear target space consistent with the additive structure of SYNTAX scoring, and (iii) explicit handling of long-tailed score distributions to reduce systematic bias. Together, these design choices yield significant performance improvements compared to state-of-the-art and remains stable across independent annotators. SynSeq demonstrates the potential of learning-based SYNTAX estimation to support more consistent and scalable cardiovascular treatment planning.

Acknowledgments. This research was conducted within an Inter-University Cluster Project jointly funded by the University of Vienna and the Medical University of Vienna. AI-POD has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement 101080302, and from the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 23.00174. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HaDEA. Neither the European Union nor the granting authority can be held responsible for them.

References

1. Avram, R., Olgin, J.E., Ahmed, Z., Verreault-Julien, L., Wan, A., Barrios, J., Abreau, S., Wan, D., Gonzalez, J.E., Tardif, J.C., et al.: CathAI: fully automated coronary angiography interpretation and stenosis estimation. npj Digital Medicine **6**(1), 142 (2023)
2. Fawzi, F.Z., Menkovic, I., Dostie, N., Tremblay-Gravel, M., Parent, M.C., Asslo, G., Tanguay, J.F., Marquis-Gravel, G., Ansari, M., Barrios, J.P., Tison, G.H., Delfrate, J., Avram, R.: Automated assessment of right ventricular systolic function from coronary angiograms with video-based artificial intelligence algorithms: development, validation, comparison against humans, and prospective deployment. European Heart Journal - Digital Health **7**(4) (Apr 2026). https://doi.org/10.1093/ehjdh/ztag059
3. Garg, S., Girasis, C., Sarno, G., Goedhart, D., Morel, M.A., Garcia-Garcia, H.M., Bressers, M., Es, G.A.v., Serruys, P.W.: The syntax score revisited: a reassessment of the syntax score reproducibility. Catheterization and Cardiovascular Interventions **75**(6), 946–952 (2010)
4. Généreux, P., Palmerini, T., Caixeta, A., Cristea, E., Mehran, R., Sanchez, R., Lazar, D., Jankovic, I., Corral, M.D., Dressler, O., et al.: Syntax score reproducibility and variability between interventional cardiologists, core laboratory technicians, and quantitative coronary measurements. Circulation: Cardiovascular Interventions **4**(6), 553–561 (2011)

5. Harrabi, S., Wu, Y., Tison, G.H., Ansari, M., Vukadinovic, M., Ouyang, D., Barrios, J.P., Delfrate, J., Avram, R.: DeepCORO-CLIP: A multi-view foundation model for comprehensive coronary angiography video-text analysis and external validation (2026). https://doi.org/10.48550/ARXIV.2603.17675
6. Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., Suleyman, M., Zisserman, A.: The kinetics human action video dataset (2017). https://doi.org/10.48550/ARXIV.1705.06950
7. Labrecque Langlais, É., Corbin, D., Tastet, O., Hayek, A., Doolub, G., Mrad, S., Tardif, J.C., Tanguay, J.F., Marquis-Gravel, G., Tison, G.H., Kadoury, S., Le, W., Gallo, R., Lesage, F., Avram, R.: Evaluation of stenoses using ai video models applied to coronary angiography. npj Digital Medicine **7**(1), 138 (2024). https://doi.org/10.1038/s41746-024-01134-4
8. Lin, H., Liu, T., Katsaggelos, A., Kline, A.: StenUNet: Automatic stenosis detection from x-ray coronary angiography (2023). https://doi.org/10.48550/ARXIV.2310.14961
9. Liu, T., Lin, H., Katsaggelos, A.K., Kline, A.: YOLO-Angio: An Algorithm for Coronary Anatomy Segmentation (2023). https://doi.org/10.48550/ARXIV.2310.15898
10. Mahmoudi, S.S., Alishani, M.M., Emdadi, M., Hosseiniyan Khatibi, S.M., Khodaei, B., Ghaffari, A., Oskui, S.D., Ghaffari, S., Pirmoradi, S.: X-ray coronary angiogram images and syntax score to develop machine-learning algorithms for chd diagnosis. Scientific Data **12**(1) (2025). https://doi.org/10.1038/s41597-025-04727-0
11. Mohr, F.W., Morice, M.C., Kappetein, A.P., Feldman, T.E., Ståhle, E., Colombo, A., Mack, M.J., Holmes, D.R., Morel, M.A., Van Dyck, N., et al.: Coronary artery bypass graft surgery versus percutaneous coronary intervention in patients with three-vessel disease and left main coronary disease: 5-year follow-up of the randomised, clinical syntax trial. The Lancet **381**(9867), 629–638 (2013)
12. Ponomarchuk, A., Kruzhilov, I., Mazanov, G., Utegenov, R., Shadrin, A., Zubkova, G., Bessonov, I., Blinov, P.: CardioSyntax: End-to-end syntax score prediction-dataset, benchmark and method. In: 2025 IEEE/CVF WACV. pp. 5873–5883. IEEE (2025)
13. Popov, M., Amanturdieva, A., Zhaksylyk, N., Alkanov, A., Saniyazbekov, A., Aimyshev, T., Ismailov, E., Bulegenov, A., Kuzhukeyev, A., Kulanbayeva, A., et al.: Dataset for automatic region-based coronary artery disease diagnostics using x-ray angiography images. Scientific Data **11**(1), 20 (2024)
14. Serruys, P.W., Onuma, Y., Garg, S., Sarno, G., van den Brand, M., Kappetein, A.P., Van Dyck, N., Mack, M., Holmes, D., Feldman, T., et al.: Assessment of the syntax score in the syntax study. EuroIntervention **5**(1), 50–56 (2009)
15. Sun, D., Liu, Z., Yu, W., Liu, H., Zhai, B., Zhao, B., Xin, H., Hu, H., Yu, Y.: A multisource cue fusion-based method for coronary artery stenosis detection in digital subtraction angiography. Journal of Thoracic Disease **18**(3), 241 (2026)
16. Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M.: A closer look at spatiotemporal convolutions for action recognition (2017). https://doi.org/10.48550/ARXIV.1711.11248
17. Vrints, C., Andreotti, F., Koskinas, K.C., Rossello, X., Adamo, M., Ainslie, J., Banning, A.P., Budaj, A., Buechel, R.R., Chiariello, G.A., Chieffo, A., Christodorescu, R.M., Deaton, C., Doenst, T., Jones, H.W., Kunadian, V., Mehilli, J., Milojevic, M., Piek, J.J., Pugliese, F., Rubboli, A., Semb, A.G., Senior, R., ten Berg, J.M., Van Belle, E., Van Craenenbroeck, E.M., Vidal-Perez, R., Winther, S., Group,

E.S.D.: 2024 ESC guidelines for the management of chronic coronary syndromes: Developed by the task force for the management of chronic coronary syndromes of the European Society of Cardiology (ESC) endorsed by the European Association for Cardio-Thoracic Surgery (EACTS). European Heart Journal **45**(36), 3415–3537 (2024). https://doi.org/10.1093/eurheartj/ehae177

18. Yang, Q., Yi, H., Yi, L., Liu, M., Chen, X.: Accurate segmentation and labeling of coronary artery segments in x-ray angiography with an improved unet-based cgan architecture. Biomedical Signal Processing and Control **112**, 108812 (2026)