



Systematic benchmarking of five GAN architectures for 12-lead ECG synthesis · Blinded clinician evaluation of synthetic ECG interpretability

Health and Medical Devices

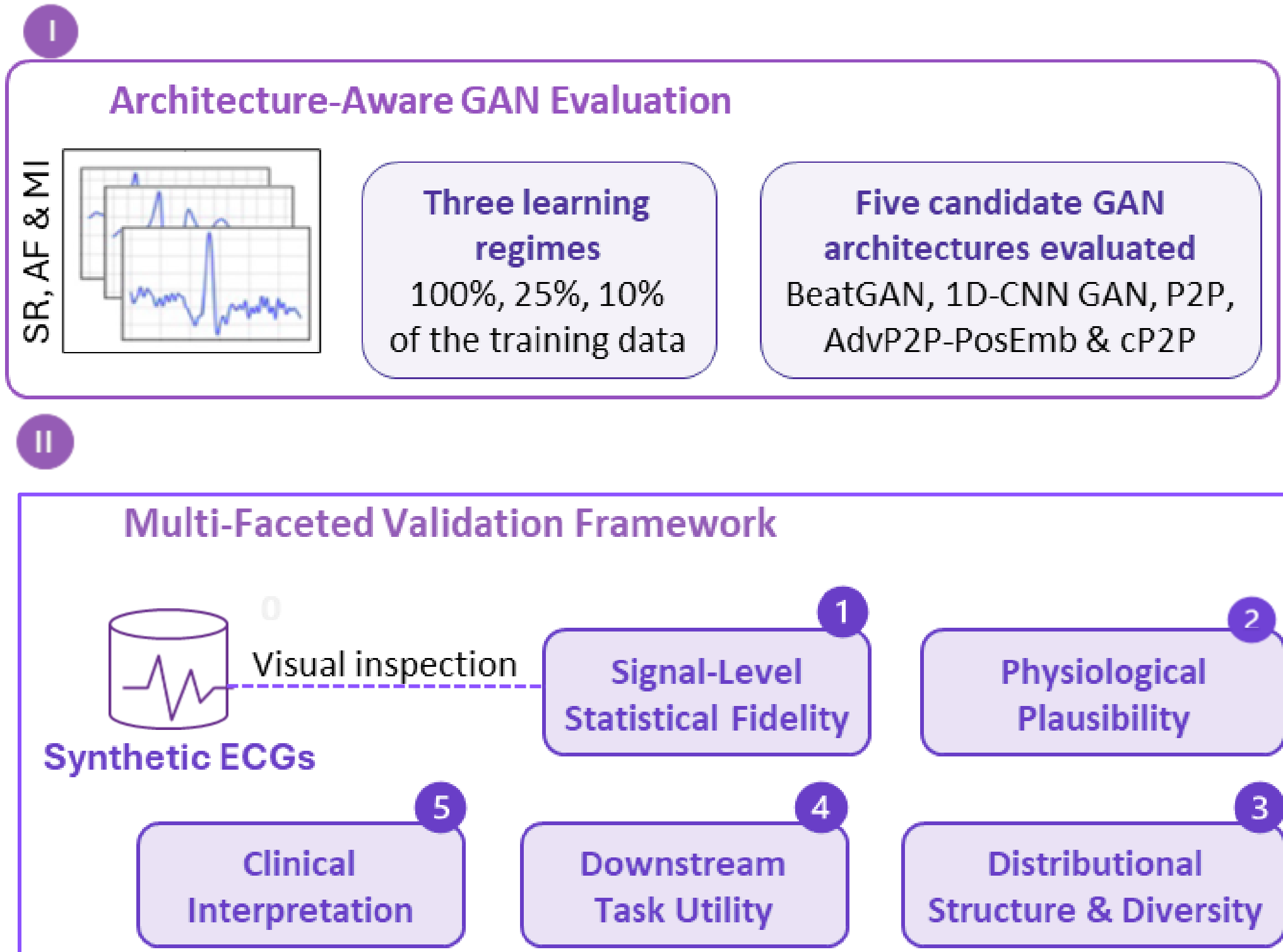
INTRODUCTION

- ECG datasets are scarce and class-imbalanced, limiting supervised learning for rare cardiac conditions
- Clinical validity of GAN-generated ECGs remains poorly characterised
- Synthetic ECG data can augment training sets for personalised cardiovascular modelling (Thambawita et al., 2021)

OBJECTIVES

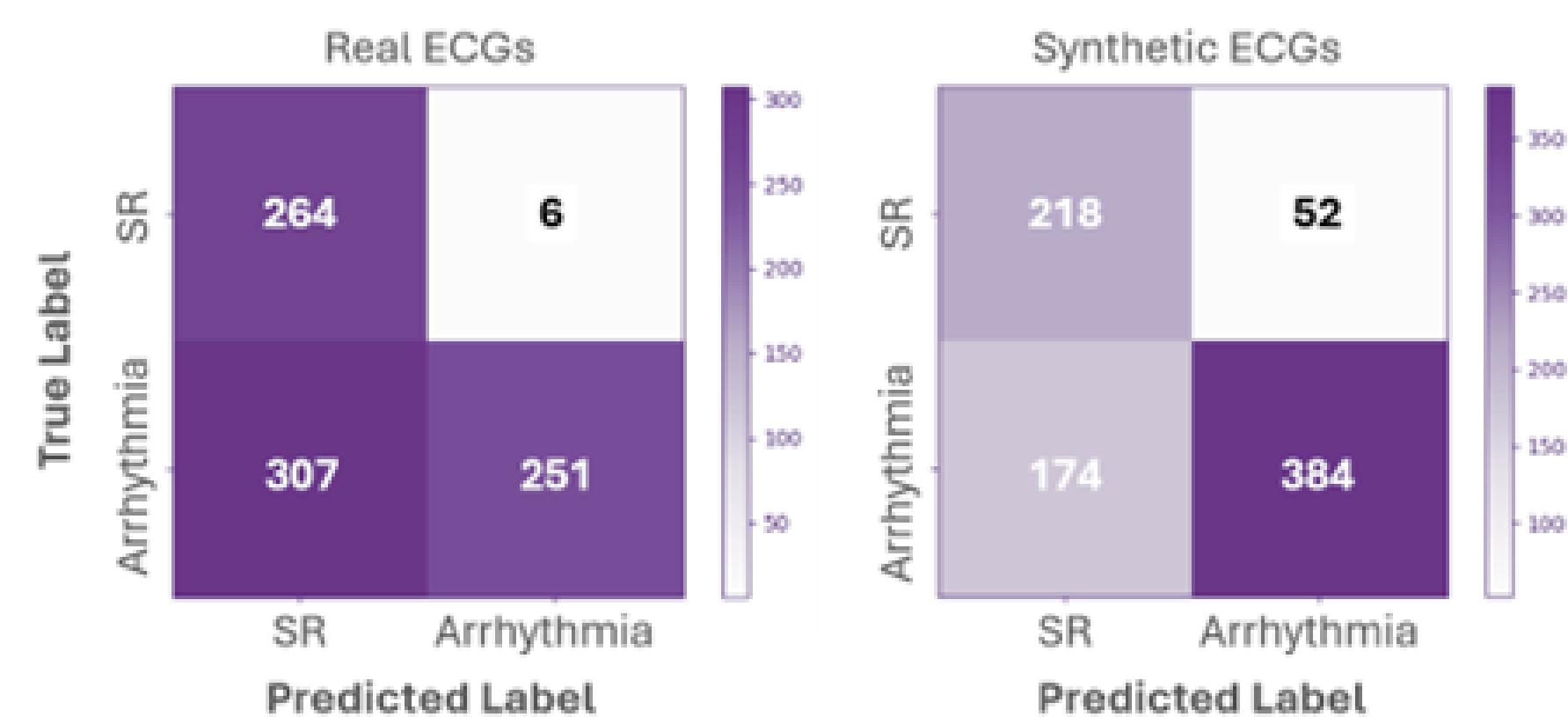
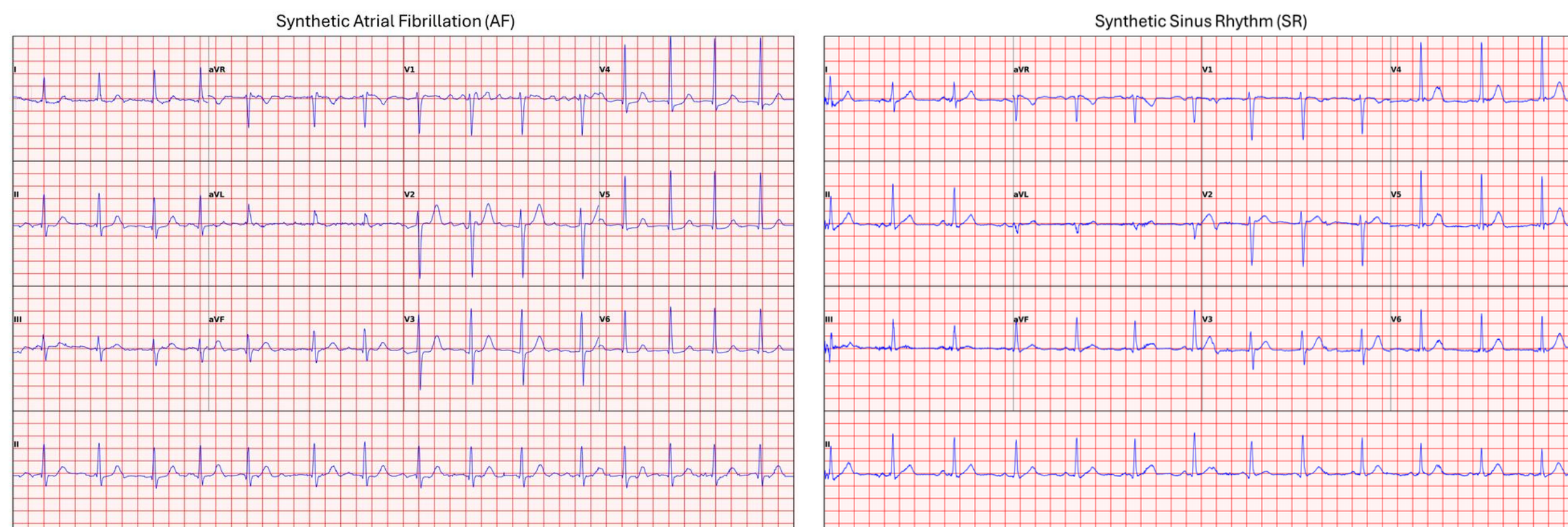
1. Benchmark five GAN architectures for multi-lead ECG synthesis under clinically realistic data constraints
2. Develop a multi-layer validation framework assessing signal fidelity, physiological plausibility, and downstream classification utility (Xia et al., 2023)
3. Conduct blinded clinician review (n=9 clinicians; 186 ECGs) to evaluate clinical interpretability of synthetic signals

METHODOLOGY



RESULTS

Across all 12 leads, GAN-based synthetic signals reproduce consistent waveform morphology with high intra-lead fidelity. P2P trained on SR and AF, achieved the lowest FID scores (7.192, 7.602) and indicating strong alignment with real data



SR vs. Arrhythmia (True vs. Predicted) results from clinical interpretation

- Synthetic SR correctly classified in 94.7% of cases vs. 91.3% for real SR
- Mean binary classification accuracy: real 0.737 ± 0.043 vs. synthetic 0.741 ± 0.058 (no significant difference)
- No clinician showed a statistically significant difference in performance between real and synthetic ECGs

STRENGTHS

- First systematic architecture-aware comparison of five GANs for 12-lead ECG synthesis
- Multi-layer validation (FID, physiological features, classification utility, blinded clinician review)

LIMITATIONS

- 10-second fixed segments; longer temporal dynamics not captured
- Single-cohort training (PTB-XL); inter-lead spatial coherence not fully preserved

CONCLUSION

- GAN-based multi-lead ECG synthesis achieves clinically interpretable outputs, validated by blinded expert review
- Moderate synthetic data augmentation improves downstream classification performance
- Multi-layer validation framework enables rigorous, clinically-grounded assessment of synthetic ECG quality

REFERENCES

[1] Thambawita, V. et al. (2021). DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-01295-2>

[2] Xia, Y., Wang, W., & Wang, K. (2023). ECG signal generation based on conditional generative models. *Biomedical Signal Processing and Control*, 82. <https://doi.org/10.1016/j.bspc.2023.104587>

[3] Li, W., Tang, Y. M., Yu, K. M., & To, S. (2022). SLC-GAN: An automated myocardial infarction detection model based on generative adversarial networks and convolutional neural networks with single-lead electrocardiogram synthesis. *Information Sciences*, 589, 738-750. <https://doi.org/10.1016/j.ins.2021.12.083>



Corresponding :
 Prof Sandra Ortega-Martorell
 S.OrtegaMartorell@ljmu.ac.uk
 Artificial Intelligence and Digital Technologies Research Institute
 Liverpool John Moores University,
 Byrom Street, Liverpool, L3 3AF

Co-authors :
 Zainab Mahmood, Ivan Olier, Dharmesh Mistry, Robyn Lotto, Brian W. Johnston, Hani Essa, Riccardo Proietti, Anil M. Tuladhar, Tim ten Cate, Theodora Benedek, Eva Giralt-Steinhauer, Joan Jiménez-Balado, Jordi Jiménez-Conde, Timothy Fairbairn, Gregory Y. H. Lip, Sandra Ortega-Martorell