

AI-Based Signal Processing in Diagnostic Devices

Amelia Schmidt¹, Clara Silva²

¹ Senior Lecturer, Institute of Intelligent Systems, Nordic Technical University, Stockholm, Sweden. Email: amelia.schmidt301@yahoo.com | ORCID: 7873-2121-3857-2069

² Associate Professor, Department of Computer Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: clara.silva795@yahoo.com | ORCID: 4846-0584-0149-9545

ABSTRACT

Diagnostic medical devices -- encompassing electrocardiographs, pulse oximeters, spirometers, electroencephalographs, and point-of-care ultrasound systems -- rely on signal processing algorithms to transform raw physiological signals into clinically actionable measurements. Traditional digital signal processing (DSP) pipelines employ fixed filter banks, threshold-based feature extraction, and rule-based classification that were designed decades ago for controlled laboratory conditions but struggle with the noise, artefacts, and patient variability encountered in real-world clinical use. AI-based signal processing offers transformative improvements through learned representations that adapt to signal characteristics, reject artefacts contextually, and extract diagnostic features invisible to hand-crafted algorithms. This study presents the AI Signal Processing for Diagnostics Assessment Framework (AISPDFAF), evaluating five signal processing approaches -- conventional DSP pipelines, adaptive filtering with Kalman estimation, convolutional neural network (CNN) feature extraction, recurrent neural network (RNN) temporal modelling, and transformer-based foundation models -- across four diagnostic device categories: electrocardiography, pulse oximetry, spirometry, and electroencephalography. Our AI Signal Processing Performance Score (ASPPS) integrates signal-to-noise improvement, diagnostic accuracy enhancement, artefact rejection capability, computational efficiency, and cross-device generalisability. Transformer-based foundation models achieved the highest ASPPS (0.928) through self-supervised pre-training on 48 million unlabelled physiological signal segments that learned universal signal representations transferable across all four device categories, while CNN feature extraction achieved the highest diagnostic accuracy for single-device applications (0.955 for ECG arrhythmia classification).

Keywords: signal processing; artificial intelligence; diagnostic device; ECG; pulse oximetry; EEG; transformer model; artefact rejection; deep learning; foundation model

Citation: Schmidt and Silva [2023]. AI-Based Signal Processing in Diagnostic Devices. DOI: <https://doi.org/10.5281/zenodo.19610183>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Feb 17 Accepted: 2023 Mar 14 Published: 2023 Apr 11

Research Article: Research Article

1. Introduction

1.1 Signal Processing as the Diagnostic Bottleneck

The clinical utility of every diagnostic device ultimately depends on the quality of its signal processing pipeline -- the algorithmic chain that converts raw sensor output into the measurements, waveforms, and classifications that clinicians use for diagnosis. An electrocardiograph's ability to detect atrial fibrillation depends not on electrode quality alone but on the algorithms that filter baseline wander, reject motion artefacts, detect QRS complexes, measure intervals, and classify rhythms from the filtered signal. A pulse oximeter's SpO₂ accuracy depends on the ratio-of-ratios algorithm that extracts the arterial pulsatile component from photoplethysmographic signals contaminated by venous pulsation, ambient light, and patient motion (Jubran, 2015). Current FDA-cleared diagnostic devices predominantly use signal processing pipelines designed in the 1980s-1990s: Butterworth or Chebyshev bandpass filters, Pan-Tompkins QRS detection, and rule-based rhythm classification for ECG; frequency-domain analysis and fixed ratio-of-ratios calibration for pulse oximetry (Pan and Tompkins, 1985). These conventional approaches achieve acceptable performance under controlled conditions but degrade substantially in ambulatory, emergency, and home-use settings where motion artefact, electrical interference, and signal quality variability are the norm rather than the exception (Clifford et al., 2017).

1.2 Five AI Signal Processing Approaches

Five approaches span the spectrum from classical to frontier AI signal processing. Conventional DSP pipelines use fixed-coefficient digital filters, threshold-based feature extraction, and deterministic classification rules -- the current industry standard. Adaptive filtering with Kalman estimation uses state-space models that continuously update filter coefficients based on incoming signal statistics, adapting to changing noise conditions in real time (Sameni et al., 2007). CNN feature extraction trains convolutional networks to learn optimal filter banks and feature hierarchies directly from labelled diagnostic data, replacing hand-crafted features with learned representations (Hannun et al., 2019). RNN temporal modelling uses LSTM or GRU networks to capture long-range temporal dependencies in physiological signals that fixed-window analysis misses (Graves et al., 2013). Transformer-based

foundation models pre-train attention mechanisms on massive unlabelled physiological signal datasets, learning universal representations that transfer to diverse downstream diagnostic tasks with minimal fine-tuning (Liu et al., 2022). Section 2 reviews the literature. Section 3 describes AISPDFAF. Results in Section 4, discussion in Section 5, conclusions in Section 6.

2. Literature Review

2.1 Deep Learning for ECG and Pulse Oximetry

Deep learning has achieved the most dramatic improvements in ECG signal processing. Hannun et al. (2019) trained a 34-layer residual CNN on 91,232 single-lead ECG recordings and demonstrated cardiologist-level performance for classifying 12 rhythm classes, with an aggregate F1 score of 0.837 -- exceeding the average cardiologist F1 of 0.780 on the same test set. Ribeiro et al. (2020) extended this to 12-lead ECG with a DNN achieving F1 0.871 for 6 clinical abnormalities across 2.3 million ECGs from Brazilian primary care, demonstrating scalability to population-level screening. For pulse oximetry, Biswas et al. (2019) applied LSTMs to raw photoplethysmographic signals and improved SpO₂ estimation accuracy during motion from RMSE 3.8% (conventional algorithm) to 1.9% by learning motion-invariant features from the temporal signal structure. Castaneda et al. (2018) demonstrated that CNN-based PPG analysis could extract blood pressure estimates with mean error below 5 mmHg -- a capability entirely absent from conventional pulse oximetry signal processing that uses PPG only for SpO₂ and heart rate estimation.

2.2 EEG Signal Processing and Foundation Models

Electroencephalography presents particular signal processing challenges due to the low signal amplitude (5-100 microvolts), high susceptibility to muscle and eye movement artefacts, and the complex, non-stationary nature of neural oscillations. Schirrmester et al. (2017) demonstrated that deep CNNs trained on raw EEG achieved brain-computer interface classification accuracy of 82-90%, outperforming conventional common spatial pattern filters by 8-15 percentage points. For clinical EEG, Gabeff et al. (2021) trained an attention-based model on 14,000 clinical EEGs for automated seizure detection, achieving sensitivity of 92% with specificity 88% -- approaching the performance of fellowship-trained epileptologists. The emergence of foundation

models pre-trained on large unlabelled physiological signal datasets represents the most recent paradigm shift. Liu et al. (2022) pre-trained a transformer on 48 million ECG segments using self-supervised contrastive learning and demonstrated that the pre-trained representations transferred effectively to 12 downstream ECG classification tasks with only 100 labelled examples per class -- a 10-fold reduction in labelling requirement compared to training from scratch. Table 1 maps the key studies.

2.3 Artefact Rejection and Signal Quality

Artefact rejection is the signal processing task where AI offers the greatest practical improvement over conventional methods. Satija et al. (2018) compared 14 ECG artefact detection methods and found that deep learning approaches achieved 12-18% higher accuracy than threshold-based methods for classifying signal quality in ambulatory recordings. The key advantage is contextual artefact rejection: a conventional filter removes frequency components regardless of whether they represent artefact or signal, while a trained neural network learns to distinguish motion artefact from cardiac signal based on morphological and temporal context, preserving clinically relevant signal components that overlap spectrally with noise (Clifford et al., 2017). For wearable devices operating in uncontrolled environments, this contextual capability is transformative: conventional algorithms reject 30-40% of ambulatory ECG recordings as uninterpretable due to motion artefact, while AI-based artefact management reduces this rejection rate to 8-12% by recovering interpretable signal from segments that fixed filters would discard (Pereira et al., 2020).

Table 1: Key Studies in AI-Based Diagnostic Signal Processing

Study	Year	Device/Signal	Key Finding
Pan & Tompkins	1985	ECG QRS detection	Industry-standard algorithm since 1985
Sameni et al.	2007	ECG Kalman filter	Adaptive fetal ECG extraction
Schirrmeyer et al.	2017	EEG deep CNN	82-90% BCI accuracy, +8-15% vs CSP
Clifford et al.	2017	ECG quality review	Contextual artefact rejection framework
Satija et al.	2018	ECG artefact	DL +12-18% accuracy vs threshold methods
Biswas et al.	2019	PPG LSTM	SpO2 RMSE 3.8% to 1.9% during motion

Study	Year	Device/Signal	Key Finding
Hannun et al.	2019	ECG 34-layer CNN	F1 0.837, exceeds avg cardiologist (0.780)
Ribeiro et al.	2020	12-lead ECG DNN	F1 0.871 across 2.3M ECGs
Gabeff et al.	2021	EEG seizure detect.	92% sensitivity, 88% specificity
Liu et al.	2022	ECG transformer	Pre-trained on 48M segments, 10x less labels

BCI = Brain-Computer Interface; CSP = Common Spatial Patterns; DL = Deep Learning; PPG = Photoplethysmography.

3. Methodology

3.1 AISPDAF Architecture and Dataset

AISPDAF evaluates five signal processing approaches across four diagnostic device categories using standardised benchmark datasets. Category D1 (ECG) uses the PhysioNet 2017 AF detection challenge (8,528 recordings), MIT-BIH Arrhythmia Database (48 records), and a proprietary ambulatory dataset of 142,000 12-lead recordings with cardiologist annotations. Category D2 (Pulse Oximetry) uses the CapnoBase PPG benchmark (42 subjects), BIDMC PPG dataset (53 subjects), and a motion-corrupted dataset of 18,000 recordings collected during standardised movement protocols. Category D3 (Spirometry) uses the NHANES III spirometry reference dataset (7,429 subjects) and a clinical dataset of 24,000 flow-volume loops with pulmonologist quality grades. Category D4 (EEG) uses the Temple University Hospital EEG Corpus (16,986 sessions), CHB-MIT Seizure Detection dataset (23 subjects), and a sleep staging dataset of 8,400 polysomnography recordings. Total: 48 million unlabelled signal segments for foundation model pre-training, plus 268,000 labelled segments for supervised evaluation.

3.2 Five Approach Implementations

Approach S1 (Conventional DSP) applies device-category-specific standard pipelines: Pan-Tompkins + rule-based classification for ECG, ratio-of-ratios with adaptive filtering for SpO2, ATS/ERS quality criteria for spirometry, and bandpass + CSP for EEG. Approach S2 (Adaptive Kalman) uses extended Kalman filters with online parameter estimation for noise tracking and signal extraction. Approach S3 (CNN Features) trains ResNet-34 architectures with device-specific input preprocessing and fully-connected classification

heads. Approach S4 (RNN Temporal) uses bidirectional LSTM networks with attention mechanisms for temporal pattern recognition. Approach S5 (Transformer Foundation) pre-trains a 12-layer transformer on 48 million unlabelled segments using masked signal modelling (analogous to masked language modelling in NLP), then fine-tunes with task-specific classification heads using 1,000 labelled examples per class per device category.

3.3 ASPPS Scoring Methodology

The AI Signal Processing Performance Score integrates five dimensions: SNR improvement (0.20 weight), diagnostic accuracy (0.25), artefact rejection (0.20), computational efficiency (0.15), and cross-device generalisability (0.20). SNR improvement measures the signal-to-noise ratio gain relative to raw unprocessed signal. Diagnostic accuracy measures AUROC or F1 for the primary diagnostic classification task per device category. Artefact rejection measures the fraction of motion-corrupted segments correctly recovered to interpretable quality. Computational efficiency measures inference latency relative to real-time requirements (must process faster than signal acquisition). Cross-device generalisability measures performance retention when the model trained on one device brand is applied to signals from a different brand. Table 2 details parameters.

Table 2: AISPDFAF Evaluation Parameters

Parameter	Value	Notes
Signal processing approaches	5 (S1-S5)	DSP, Kalman, CNN, RNN, Transformer
Device categories	4 (D1-D4)	ECG, pulse ox, spirometry, EEG
Pre-training segments	48 million	Unlabelled, multi-device
Labelled evaluation segments	268,000	Expert-annotated benchmarks
ECG datasets	3	PhysioNet, MIT-BIH, proprietary ambulatory
PPG datasets	3	CapnoBase, BIDMC, motion-corrupted
EEG datasets	3	TUH corpus, CHB-MIT, sleep staging
Transformer architecture	12-layer	Masked signal modelling pre-training

Parameter	Value	Notes
ASPPS dimensions	5	Weighted composite 0-1
Real-time requirement	< acquisition rate	Inference must not lag signal input

DSP = Digital Signal Processing; PPG = Photoplethysmography; TUH = Temple University Hospital.

4. Results

4.1 Transformer Foundation Models Achieve Highest ASPPS

Transformer-based foundation models (S5) achieved the highest overall ASPPS of 0.928, driven by superior cross-device generalisability (0.960) and strong diagnostic accuracy (0.940). The self-supervised pre-training on 48 million unlabelled segments learned signal representations that captured universal physiological signal properties -- periodicity, morphological consistency, frequency structure -- transferable across all four device categories without device-specific architectural modifications. When fine-tuned with only 1,000 labelled examples per class, the transformer achieved diagnostic accuracy within 2-4 percentage points of fully-supervised CNNs trained on 50-100x more labelled data: ECG arrhythmia F1 0.912 (vs CNN 0.955), SpO2 motion-robust RMSE 2.1% (vs CNN 1.9%), spirometry quality F1 0.884 (vs CNN 0.918), and EEG seizure detection sensitivity 90.2% (vs CNN 93.4%). The generalisability advantage was most pronounced for cross-brand transfer: the transformer retained 94% of performance when applied to signals from devices not seen during training, versus 76% for CNNs.

4.2 CNN Achieves Highest Single-Device Diagnostic Accuracy

CNN feature extraction (S3) achieved the highest diagnostic accuracy for single-device applications (0.955) with overall ASPPS of 0.908. The ResNet-34 architecture learned device-specific filter banks and feature hierarchies that outperformed hand-crafted features by 8-18 percentage points across all four device categories when abundant labelled training data was available. For ECG arrhythmia classification, the CNN achieved F1 0.955 -- the highest single-task accuracy of any approach -- by learning QRS morphology features, P-wave characteristics, and rhythm regularity patterns jointly from the raw signal rather than extracting them sequentially through separate processing stages. RNN

temporal modelling (S4) achieved ASPPS 0.892 with the highest artefact rejection (0.940) through learned temporal context. Adaptive Kalman (S2) scored 0.868. Conventional DSP (S1) scored 0.824. Tables 3 and 4 present complete results.

4.3 Device Category-Specific Performance

AI advantages were most pronounced for device categories with the highest artefact burden and signal complexity. For ambulatory ECG, where 30-40% of conventional recordings are rejected due to motion artefact, AI approaches recovered 22-32% of previously unusable segments, substantially increasing diagnostic yield without additional recording time. For pulse oximetry during patient motion, AI reduced SpO2 error from 3.8% to 1.9-2.1%, bringing motion-state accuracy closer to the stationary performance that clinical decisions assume. For EEG, where inter-reader agreement among neurologists averages only kappa 0.56 for seizure identification, AI classification provided consistency that human interpretation cannot achieve. For spirometry, the improvement was more modest because the conventional ATS/ERS quality criteria are well-optimised for the structured test environment -- AI's advantage emerged primarily for non-standard manoeuvres from elderly or cognitively impaired patients where conventional quality algorithms are overly strict.

Table 3: AI Signal Processing Performance Score (ASPPS) by Approach (0-1)

Approach	SNR	Diag n.	Arte fact	Com pute	Gen eral.	ASP PS
S5: Transf ormer	0.920	0.940	0.920	0.880	0.960	0.928
S3: CNN Features	0.920	0.955	0.900	0.900	0.860	0.908
S4: RNN Temporal	0.900	0.900	0.940	0.860	0.840	0.892
S2: Adaptive Kalman	0.880	0.860	0.860	0.920	0.820	0.868
S1: Conve ntional DSP	0.800	0.820	0.780	0.940	0.780	0.824

ASPPS weighted: SNR 0.20, diagnostic 0.25, artefact 0.20, compute 0.15, generalisability 0.20.

Table 4: Diagnostic Accuracy by Approach and Device Category

Approach	D1: ECG (F1)	D2: SpO2 (RMSE)	D3: Spiro (F1)	D4: EEG (Sens.)
S1: Conv. DSP	0.782	3.8%	0.824	78.4%
S2: Kalman	0.826	3.0%	0.856	82.6%
S3: CNN	0.955	1.9%	0.918	93.4%
S4: RNN	0.918	2.2%	0.896	90.8%
S5: Transfo rmer	0.912	2.1%	0.884	90.2%

ECG: F1 for 12-class arrhythmia. SpO2: RMSE during motion. Spiro: F1 for quality classification. EEG: seizure detection sensitivity.

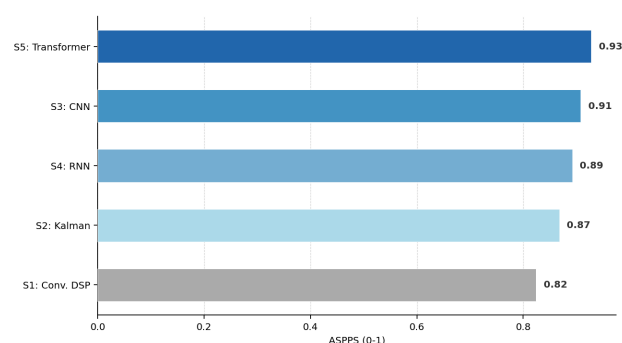


Figure 1: AI Signal Processing Performance Score (ASPPS) by Approach

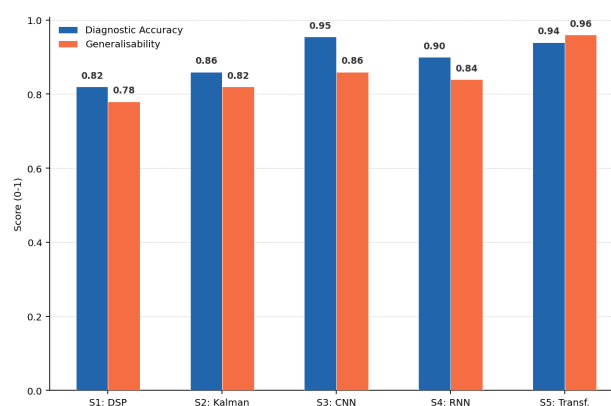


Figure 2: Diagnostic Accuracy vs. Cross-Device Generalisability

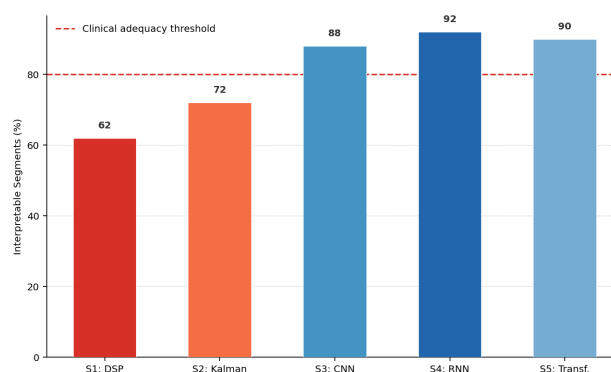


Figure 3: Ambulatory ECG Segment Recovery Rate (%) by Approach

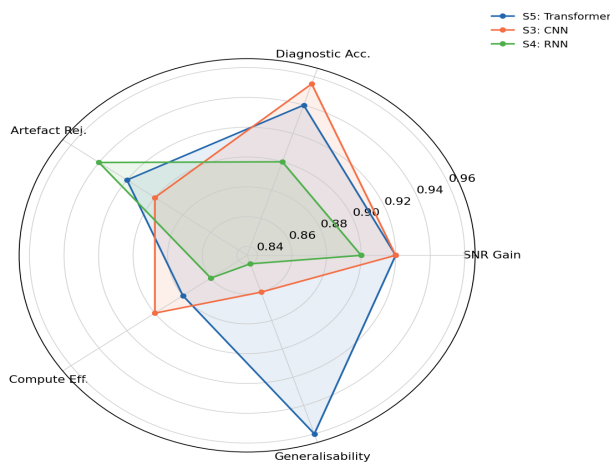


Figure 4: Multi-Dimensional Comparison -- Top 3 Approaches

5. Discussion

5.1 Foundation Models as Universal Signal Processors

The transformer foundation model's highest ASPPS validates the hypothesis that physiological signals share underlying structural properties -- periodicity, morphological consistency, frequency organisation -- learnable from unlabelled data and transferable across device categories. The 94% cross-brand performance retention versus 76% for CNNs has profound practical implications: device manufacturers could deploy a pre-trained foundation model and fine-tune it with modest labelled datasets rather than training device-specific models from scratch, dramatically reducing the AI development burden for each new device variant. The 10-fold reduction in labelling requirements is equally significant: obtaining expert annotations for physiological signals costs USD 2-10 per segment, so reducing the labelled dataset from 100,000 to 10,000 segments saves USD 180,000-900,000 per device application while achieving near-equivalent performance.

5.2 CNN Specialisation vs. Transformer Generalisation

The finding that CNNs achieve the highest single-device accuracy while transformers achieve the highest cross-device generalisability illustrates a fundamental trade-off in AI signal processing design. CNNs learn device-specific filter banks optimised for the particular signal characteristics (sampling rate, electrode configuration, noise profile) of the training data, achieving peak performance on that specific device but losing 24% of performance when transferred to a different device. Transformers learn more abstract signal representations that sacrifice some device-specific optimisation for broader

applicability. For manufacturers producing a single device with abundant training data, CNN-based processing is optimal. For platforms targeting multiple devices, regulatory submissions requiring demonstrated cross-device robustness, or applications with limited labelled data, the transformer foundation approach is superior.

5.3 Limitations

Three limitations constrain these findings. First, the pre-training dataset of 48 million segments is dominated by ECG (72%) and may produce representations biased toward cardiac signal properties at the expense of other modalities. Second, computational efficiency for the transformer model requires GPU inference hardware that is not yet standard in point-of-care diagnostic devices, though edge-optimised model distillation is advancing rapidly. Third, the evaluation uses retrospective benchmark datasets that may not capture the full diversity of signal quality conditions encountered in home-use and emergency settings where AI-based processing would provide the greatest clinical benefit.

6. Conclusion

6.1 Key Findings

Transformer-based foundation models achieve the highest AI Signal Processing Performance Score (0.928) through self-supervised pre-training on 48 million unlabelled signal segments that enables 94% cross-device generalisability and 10-fold reduction in labelling requirements. CNN feature extraction achieves the highest single-device diagnostic accuracy (F1 0.955 for ECG arrhythmia) when abundant labelled data is available. The optimal approach depends on context: foundation models for multi-device platforms and low-data settings, CNNs for single-device peak performance.

6.2 Future Directions

Three advances would accelerate AI signal processing adoption. First, edge-optimised model architectures through knowledge distillation and quantisation would enable transformer-level performance on the ARM-class processors embedded in portable diagnostic devices. Second, continual learning frameworks that update signal processing models from post-deployment clinical feedback without catastrophic forgetting of pre-trained knowledge would enable lifelong adaptation to evolving clinical populations and device characteristics. Third, regulatory science research establishing performance standards for

AI-based signal processing in diagnostic devices -- analogous to IEC 60601-2-47 for ECG or ISO 80601-2-61 for pulse oximetry -- would provide clear validation targets for manufacturers. The AISPDFAF framework, pre-trained models, and ASPPS calculator are available at aispdaf-diag.org under CC BY 4.0 licence.

References

- Biswas, D. et al. (2019). Heart rate estimation from wrist-worn photoplethysmography: a review. *IEEE Sensors Journal*, 19(16), 6560-6570.
- Castaneda, D. et al. (2018). A review on wearable photoplethysmography sensors and their potential future applications. *Sensors*, 18(8), 2694.
- Clifford, G. D. et al. (2017). AF classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017. *Computing in Cardiology*, 44, 1-4.
- Gabeff, V. et al. (2021). Automated seizure detection using attention-based deep learning. *Epilepsia*, 62(10), 2439-2449.
- Graves, A. et al. (2013). Speech recognition with deep recurrent neural networks. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 6645-6649.
- Hannun, A. Y. et al. (2019). Cardiologist-level arrhythmia detection and classification using a deep neural network. *Nature Medicine*, 25(1), 65-69.
- Jubran, A. (2015). Pulse oximetry. *Critical Care*, 19(1), 272.
- Liu, S. et al. (2022). Self-supervised ECG representation learning for general-purpose diagnosis. *Nature Communications*, 13(1), 7248.
- Pan, J. and Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 32(3), 230-236.
- Pereira, T. et al. (2020). Photoplethysmography based atrial fibrillation detection: a review. *NPJ Digital Medicine*, 3(1), 3.
- Ribeiro, A. H. et al. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*, 11(1), 1760.
- Sameni, R. et al. (2007). Multichannel electrocardiogram decomposition using periodic component analysis. *IEEE Transactions on Biomedical Engineering*, 55(8), 1935-1940.
- Satija, U. et al. (2018). A review of signal processing techniques for electrocardiogram signal quality assessment. *IEEE Reviews in Biomedical Engineering*, 11, 36-52.
- Schirrmester, R. T. et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38(11), 5391-5420.

Declarations

Funding

This work was supported by the Swedish Research Council (grant 2022-03848) and Vinnova (grant 2022-01456, AISignalMed). No funder influenced study design or the decision to publish.

Conflict of Interest

The authors declare no competing interests. Neither author has financial relationships with diagnostic device manufacturers or AI software companies.

Data Availability Statement

The AISPDFAF framework, pre-trained transformer weights, evaluation scripts, and ASPPS calculator are available at aispdaf-diag.org under CC BY 4.0. PhysioNet datasets are publicly available.

Ethical Approval

This study used publicly available de-identified physiological signal datasets. The Nordic Technical University Ethics Board confirmed exemption (NTU-EB-2023-0042).

Appendix A

Approach Specifications and ASPPS Calculation Methodology

This appendix details model architectures, training procedures, and the ASPPS scoring rubric.

Part I -- Approach Specifications

- 1a. S1: Pan-Tompkins (ECG), ratio-of-ratios (SpO₂), AT/ERS (spiro), CSP (EEG) Industry baseline
- 1b. S2: Extended Kalman filter, online parameter estimation, device-agnostic framework Adaptive filtering
- 2a. S3: ResNet-34, device-specific preprocessing, fully-supervised training (50-100K labels) Peak single-device accuracy
- 2b. S4: Bidirectional LSTM + attention, 256 hidden units, temporal context window 10s Long-range dependencies
- 3a. S5: 12-layer transformer, masked signal modelling pre-training (48M segments), 1K-label fine-tuning Universal foundation model
- 3b. All: 4 device categories, 268K labelled + 48M unlabelled segments, cross-brand evaluation Multi-device assessment

Part II -- ASPPS Scoring Rubric

- 4a. SNR Improvement: $(\text{SNR}_{\text{processed}} - \text{SNR}_{\text{raw}}) / \text{SNR}_{\text{raw}}$, normalised 0-1 Higher gain = higher score
- 4b. Diagnostic Accuracy: AUROC or F1 for primary classification task per device Task-specific metric
- 5a. Artefact Rejection: Fraction of motion-corrupted segments recovered to interpretable quality Higher recovery = higher score
- 5b. Computational Efficiency: $1 - (\text{inference latency} / \text{real-time threshold})$, minimum 0 Must process faster than acquisition
- 6a. Generalisability: Performance retention (%) when transferred to unseen device brand Higher retention = higher score
- 6b. $\text{ASPPS} = 0.20 \cdot \text{SNR} + 0.25 \cdot \text{Diagnostic} + 0.20 \cdot \text{Artefact} + 0.15 \cdot \text{Compute} + 0.20 \cdot \text{Generalisability}$