RESEARCH PAPER

Foundation Models and Self-Supervised Learning in ECG Signal Processing: Towards Scalable, Multimodal, and Trustworthy Cardiovascular Artificial Intelligence

Dr. Baby Paul[1]

[1]Associate Professor, Department of Electronics, Baselios Poulose II Catholicos College, Piravom, Ernakulam Dist, Kerala, India

Abstract

Artificial intelligence (AI) has entered a transformative era defined by large-scale foundation models trained through self-supervised learning (SSL). Unlike conventional supervised approaches that depend heavily on annotated datasets, SSL leverages intrinsic data structure to learn robust and transferable representations from vast quantities of unlabelled data. In electrocardiography (ECG), this paradigm represents a significant breakthrough, addressing longstanding challenges related to annotation cost, dataset bias, limited cross-population generalization, and scalability across devices and clinical environments. Cardiovascular diseases remain the leading cause of global mortality, underscoring the urgent need for reliable, scalable, and interpretable automated ECG analysis systems.

Foundational advances in transformer architectures (Vaswani et al., 2017), contextual pretraining (Devlin et al., 2018), contrastive representation learning (Chen et al., 2020; He et al., 2020), masked modelling (He et al., 2021), and probabilistic generative modelling (Kingma & Welling, 2014) have collectively established the theoretical and architectural basis for foundation-level ECG intelligence. These innovations enable models to capture both local waveform morphology and long-range rhythm dependencies, while supporting transfer learning across diverse downstream tasks such as arrhythmia detection, risk stratification, and anomaly identification.

This review synthesizes the evolution of SSL paradigms applied to ECG analysis up to 2022, critically examining architectural design choices, benchmarking methodologies, domain adaptation techniques, interpretability mechanisms, privacy-preserving training strategies, and requirements for clinical validation. Furthermore, I propose a forward-looking roadmap toward multimodal, federated, and continually learning biomedical foundation models capable of robust cross-institutional deployment and trustworthy clinical integration. The convergence of SSL and ECG analytics signals a paradigm shift from narrow task-specific classifiers toward universal, adaptive cardiovascular representation learning systems that can underpin next-generation precision cardiology.

How to cite this article: Paul B. Foundation Models and Self-Supervised Learning in ECG Signal Processing: Towards Scalable, Multimodal, and Trustworthy Cardiovascular Artificial Intelligence. Int J Drug Deliv Technol. 2026;16(13s): 173-180. DOI: 10.25258/ijddt.16.13s.18

1. Introduction

Electrocardiography (ECG) remains one of the most ubiquitous, accessible, and cost-effective diagnostic modalities in cardiovascular medicine. Since its clinical adoption in the early twentieth century, ECG interpretation has played a central role in diagnosing arrhythmias, myocardial infarction, conduction abnormalities, electrolyte imbalances, and structural heart diseases. The simplicity, non-invasiveness, and rapid acquisition of ECG recordings have made them indispensable in emergency care, outpatient cardiology, intensive care units, and increasingly, wearable and remote monitoring systems. Despite major technological advancements in acquisition hardware and digital storage, interpretation of ECG signals continues to rely heavily on expert cardiologists. This dependence limits scalability, introduces inter-observer variability, and constrains timely diagnosis in resource-limited settings.

The emergence of deep learning marked a turning point in automated ECG analysis. Convolutional neural networks (CNNs) demonstrated remarkable performance in arrhythmia detection and rhythm classification tasks, in some cases achieving cardiologist-level accuracy (Rajpurkar et al., 2017; Hannun et al., 2019). These advances were fuelled by the availability of curated labelled datasets such as the MIT-BIH Arrhythmia Database (Moody & Mark, 2001) and large repositories hosted by PhysioNet (Goldberger et al., 2000). Supervised deep learning models trained on such datasets showed strong capacity for extracting morphological features and identifying complex rhythm patterns.

However, supervised learning approaches face critical structural limitations. First, the annotation bottleneck remains a major barrier, as high-quality ECG labelling requires clinical expertise and significant time investment. Second, demographic and institutional biases within datasets can limit model generalizability across populations. Third, models trained on specific acquisition devices or pre-processing pipelines often struggle with cross-device generalization. Finally, supervised models frequently exhibit poor robustness under distribution shift, such as noise variation, electrode misplacement, or differing sampling frequencies.

Self-supervised learning (SSL) offers a compelling alternative by enabling models to learn meaningful representations from large volumes of unlabelled ECG data (LeCun et al., 2021). In parallel, the concept of foundation models—large-scale pretrained architectures transferable across multiple downstream tasks—has reshaped natural language processing and computer vision (Devlin et al., 2018; Vaswani et al., 2017). Within ECG research, the convergence of SSL and foundation modelling holds promise for reducing reliance on manual annotations, enhancing cross-dataset transferability, enabling scalable representation learning, and facilitating multimodal integration with other physiological signals. This review provides a comprehensive and critical examination of this rapidly evolving paradigm and its implications for next-generation cardiovascular AI.

2. From Supervised ECG Learning to Foundation Paradigms

2.1 Limitations of Supervised ECG Models

Supervised ECG classifiers require extensive labelled datasets to achieve clinically reliable performance. Although several public repositories exist, most are derived from limited geographic regions and specific patient populations, restricting demographic diversity and potentially embedding systemic biases into trained models. As a result, models may perform well on benchmark datasets yet fail when deployed across different hospitals, wearable devices, or acquisition settings. Overfitting to particular electrode configurations, sampling frequencies, or pre-processing pipelines remains a persistent issue. Furthermore, supervised training typically optimizes representations for a single predefined task, such as arrhythmia classification, thereby limiting transferability to other clinically relevant objectives without substantial retraining.

2.2 Emergence of Self-Supervised Learning

Self-supervised learning (SSL) reframes representation learning by introducing pretext tasks that leverage the intrinsic structure of data rather than relying on manually annotated labels (LeCun et al., 2021). This paradigm is particularly well suited to ECG signals, which possess rich and structured temporal dynamics. First, ECG recordings exhibit strong periodicity, reflecting the cyclical nature of cardiac activity. Second, they demonstrate morphological consistency in waveform components such as the P wave, QRS complex, and T wave. Third, ECG signals have a hierarchical temporal structure, encompassing beat-level morphology, rhythm-level patterns, and longer-term physiological variability, all of which SSL can effectively model.

3. Theoretical Foundations of SSL for ECG

3.1 Contrastive Learning

Contrastive learning aims to learn discriminative representations by maximizing agreement between different augmented views of the same signal (positive pairs) while minimizing similarity between representations of distinct signals (negative pairs) (Chen et al., 2020; He et al., 2020). Instead of relying on explicit diagnostic labels, the model is trained to recognize intrinsic identity-preserving characteristics of the data. This objective encourages the encoder to capture stable and meaningful features that remain invariant under realistic transformations, making contrastive learning particularly attractive for bio signals such as ECG.

In the context of ECG, however, augmentation design requires careful physiological consideration. Unlike images, where aggressive transformations may still preserve semantic meaning, ECG signals are tightly constrained by cardiac electrophysiology. Physiologically valid augmentations must preserve core signal characteristics, including QRS morphology, rhythm periodicity, and inter-lead relationships in multi-lead recordings. Distorting these elements risks introducing artificial patterns that the model may mistakenly learn as discriminative features.

Consequently, contrastive ECG pretraining typically employs controlled augmentations such as temporal jittering (small time shifts that preserve waveform shape), lead masking (simulating missing leads without altering cardiac structure), mild noise injection (reflecting realistic acquisition noise), and frequency filtering (mimicking baseline wander or power-line

interference). These transformations encourage robustness while maintaining clinical plausibility.

Importantly, Franceschi et al. (2019) demonstrated that unsupervised representation learning for time series significantly improves downstream task performance under label scarcity. Their findings suggest that contrastive objectives can extract general-purpose temporal embeddings transferable across classification tasks. Applied to ECG, this paradigm enables models to learn morphology- and rhythm-aware representations that require substantially fewer labelled examples for fine-tuning, thereby reducing annotation burden while enhancing generalization.

3.2 Masked Modelling

Masked modelling is a self-supervised learning paradigm in which a model is trained to reconstruct deliberately removed or masked portions of the input signal (Devlin et al., 2018; He et al., 2021). Rather than distinguishing between different samples as in contrastive learning, masked modelling focuses on predicting missing segments based on surrounding context. This objective forces the model to develop a deep understanding of the internal structure and temporal dependencies within the data. Originally introduced in natural language processing through masked language modelling and later extended to vision via masked autoencoders, this strategy has demonstrated strong potential for structured time-series signals such as ECG.

For ECG analysis, masked modelling offers several important advantages. First, it enforces contextual rhythm understanding. Because the model must infer missing waveform segments from neighbouring beats, it learns to capture periodicity, inter-beat relationships, and rhythm regularity. Second, it captures long-range temporal dependencies. Cardiac abnormalities often manifest over multiple cycles rather than within a single beat; reconstructing masked regions encourages the encoder to model dependencies across extended temporal windows. Third, masked modelling encourages morphology reconstruction. To accurately predict masked QRS complexes or T waves, the model must internalize the characteristic shapes and amplitude patterns of normal and abnormal cardiac waveforms.

3.3 Generative Representation Learning

Generative representation learning provides an alternative self-supervised framework by explicitly modelling the underlying data distribution of ECG signals. Variational autoencoders (VAEs) (Kingma & Welling, 2014) encode input signals into structured latent probabilistic spaces and reconstruct them through a decoder network. Unlike deterministic encoders, VAEs impose a probabilistic prior on the latent space, typically Gaussian, enabling smooth interpolation between cardiac states and principled uncertainty estimation. For ECG signals, this probabilistic structure is particularly valuable because cardiac rhythms exhibit both physiological variability and pathological deviations.

Latent representations learned through VAEs support several clinically meaningful applications. First, uncertainty estimation allows models to quantify confidence in predictions, which is critical in safety-sensitive medical settings. Second, anomaly detection can be performed by measuring reconstruction error or latent likelihood; abnormal rhythms often produce higher reconstruction discrepancies compared to normal sinus rhythms. Third, synthetic data generation becomes feasible by sampling from the latent distribution, enabling augmentation of rare cardiac conditions that are underrepresented in clinical datasets.

Generative adversarial networks (GANs) have also demonstrated success in producing realistic ECG waveforms (Golany & Radinsky, 2019). By training a generator to produce synthetic signals that are indistinguishable from real ECGs to a discriminator, GAN-based models can learn high-fidelity waveform distributions. These synthetic signals can enhance training diversity, mitigate class imbalance, and facilitate privacy-preserving data sharing by generating de-identified yet physiologically plausible ECG samples.

Despite their promise, generative models must ensure clinical fidelity and avoid producing morphologically implausible rhythms. Careful evaluation by domain experts remains essential before deployment in real-world cardiology workflows.

4. Comparative Analysis of SSL Paradigms for ECG

Paradigm	Foundational Work	Learning Objective	ECG Strengths	Limitations
Contrastive	Chen et al. (2020); He et al. (2020)	Instance discrimination	Label efficiency; domain	Batch-size sensitivity

Paradigm	Foundational Work	Learning Objective	ECG Strengths	Limitations
			invariance	
Masked Reconstruction	Devlin et al. (2018); He et al. (2021)	Signal reconstruction	Global context modelling	Computationally heavy
Generative (VAE/GAN)	Kingma & Welling (2014); Golany & Radinsky (2019)	Distribution modelling	Uncertainty & synthesis	Clinical fidelity concerns

5. Architectural Evolution Toward ECG Foundation Models

5.1 CNN-Based Models

Convolutional neural networks (CNNs) have been the dominant architecture in early deep learning approaches for ECG analysis due to their strong inductive bias for local feature extraction. By applying one-dimensional convolutional filters across time, CNNs effectively capture localized waveform characteristics such as QRS complexes, P-wave morphology, ST-segment deviations, and T-wave variations. These morphological patterns are critical for diagnosing arrhythmias, ischemic changes, and conduction abnormalities. The hierarchical structure of CNNs—stacking multiple convolutional layers with nonlinear activations—enables progressive abstraction from raw voltage signals to high-level diagnostic features.

Furthermore, CNNs are computationally efficient and well-suited for real-time deployment in embedded or wearable devices. Their parameter sharing and localized receptive fields make them robust to small temporal shifts and noise variations. However, CNNs inherently rely on fixed receptive fields, which can limit their ability to model long-range temporal dependencies spanning multiple cardiac cycles. Although deeper architectures or dilated convolutions can partially expand receptive fields, capturing global rhythm context remains challenging. As a result, CNN-based models may struggle with tasks requiring extended temporal reasoning, such as atrial fibrillation detection or heart rate variability analysis, motivating exploration of transformer-based and hybrid architectures.

5.2 Transformer Architectures

Transformer architectures, originally introduced for sequence modelling in natural language processing (Vaswani et al., 2017), have increasingly been adopted for time-series analysis, including ECG signal processing. Unlike convolutional neural networks, which rely on fixed local receptive fields, transformers employ self-attention mechanisms that allow each time step to attend to every other time step in the sequence. This global attention capability enables the model to capture long-range dependencies and complex temporal relationships across multiple cardiac cycles.

In ECG analysis, rhythm abnormalities such as atrial fibrillation, atrioventricular block, or ventricular tachycardia often manifest over extended temporal intervals rather than within a single beat. Transformers are particularly well suited for modelling such rhythm-level patterns because self-attention dynamically weights relevant segments of the signal, regardless of their temporal distance. This flexibility improves the model's ability to detect irregular periodicity, subtle conduction delays, and inter-beat variability.

Additionally, positional encoding mechanisms allow transformers to retain temporal order information while maintaining parallel computation efficiency. However, transformer models typically require larger datasets and higher computational resources than CNNs. Despite these demands, their superior capacity for global rhythm modelling makes them a compelling backbone for emerging ECG foundation models.

5.3 Hybrid Architectures

Hybrid CNN–Transformer architectures aim to integrate the complementary strengths of convolutional neural networks and transformer models for ECG signal analysis. While CNNs are highly effective at extracting localized morphological features—such as QRS complexes, ST-segment deviations, and T-wave shapes—transformers excel at modelling long-range temporal dependencies and rhythm dynamics. By combining these two paradigms, hybrid architectures achieve both morphological sensitivity and contextual reasoning within a unified framework.

Typically, CNN layers are employed in the early stages of the model to capture low-level and mid-level waveform features through convolutional filtering and pooling operations. These extracted feature maps are then passed to transformer encoder layers, which apply self-attention mechanisms to model global

relationships across time. This design allows the network to understand how individual beats relate to broader rhythm patterns, improving detection of arrhythmias that depend on temporal irregularities rather than isolated waveform anomalies.

Hybrid architectures also improve computational efficiency compared to full transformer models. By reducing sequence dimensionality through convolutional pre-processing, the attention mechanism operates on compact feature representations rather than raw signals. Consequently, hybrid CNN–Transformer models provide a balanced solution for ECG foundation modelling, offering strong local feature extraction alongside robust global temporal modelling, making them well suited for scalable and clinically relevant cardiovascular AI systems.

6. Benchmarking and Evaluation Frameworks

A critical gap in self-supervised learning (SSL) research for ECG analysis is the absence of standardized benchmarking frameworks. While numerous studies report promising results, comparisons are often hindered by inconsistent datasets, pre-processing pipelines, evaluation metrics, and label availability assumptions. Without unified evaluation protocols, it becomes difficult to determine whether reported improvements stem from methodological innovation or dataset-specific optimization.

An effective benchmarking framework for ECG foundation models should incorporate several core characteristics. First, multi-dataset evaluation is essential to assess cross-cohort robustness and prevent overfitting to a single repository. Second, cross-device validation should test models on signals acquired using different electrode configurations, sampling frequencies, and hardware systems. Third, low-label fine-tuning protocols are necessary to evaluate representation quality under realistic annotation constraints. Fourth, robustness testing under noise perturbations, baseline wander, and motion artefacts is critical for real-world deployment. Finally, benchmarking should include fairness metrics, such as performance stratified by age, sex, and ethnicity, to ensure equitable generalization.

PhysioNet platforms (Goldberger et al., 2000) provide a strong foundation for reproducible research through open datasets, standardized formats, and challenge competitions. However, future SSL benchmarking efforts should expand toward unified splits, cross-dataset transfer tasks, and transparent reporting of computational cost and environmental impact. Establishing rigorous evaluation ecosystems will be pivotal in transitioning ECG foundation models from research prototypes to clinically reliable systems.

7. Domain Shift and Generalization

Domain shift remains one of the most significant obstacles in deploying ECG AI systems across real-world healthcare environments. ECG signals exhibit substantial variability due to differences in sampling rates, electrode configurations, signal pre-processing pipelines, demographic distributions, and clinical contexts. For example, hospital-grade 12-lead ECG recordings differ considerably from single-lead wearable recordings in signal resolution and noise characteristics. Similarly, age-related cardiac physiology and comorbidities introduce additional variability that can degrade model performance when applied outside the training domain.

Supervised deep learning models are particularly vulnerable to domain shift because they often learn dataset-specific patterns rather than generalizable physiological representations. SSL-based foundation models aim to mitigate this limitation by learning invariant and transferable embeddings from diverse unlabelled data. However, even SSL models require strategies to ensure robustness across domains.

Domain-invariant representation learning techniques seek to extract features that remain stable across acquisition conditions. Adversarial domain adaptation methods can encourage models to align latent distributions across datasets. Additionally, meta-learning strategies may enable rapid adaptation to new environments with minimal labelled data. Large-scale multi-institutional pretraining further reduces bias by exposing models to heterogeneous signals during representation learning.

Ultimately, overcoming domain shift requires not only algorithmic innovation but also diverse and representative training data. Robust cross-domain evaluation should become a central requirement in ECG foundation model development.

8. Explainability and Clinical Trust

Explainability is a central requirement for clinical deployment of ECG foundation models. While transformer architectures employ attention

mechanisms, attention weights do not inherently guarantee interpretability (Serrano & Smith, 2019). A model may attend to specific time segments without providing clinically meaningful reasoning aligned with electrophysiological principles.

Clinically trustworthy ECG models must satisfy several criteria. First, saliency maps and attribution techniques should align with known physiological structures such as P waves, QRS complexes, and T waves. When diagnosing arrhythmias, the model's highlighted regions should correspond to rhythm irregularities or conduction abnormalities identifiable by cardiologists. Second, models should provide calibrated uncertainty estimates, allowing clinicians to assess confidence levels in predictions. This is particularly important in high-risk scenarios such as myocardial infarction detection.

Beyond algorithmic transparency, cardiologist validation remains essential. Interpretability tools should be evaluated in collaboration with domain experts to ensure alignment with clinical reasoning. Additionally, explainability should extend beyond visualizations to include stability analysis—ensuring explanations remain consistent under small input perturbations.

Ultimately, explainability must evolve from a supplementary feature to an integrated design principle in ECG foundation models. Only through interpretable and reliable outputs can SSL-based systems gain clinician trust and regulatory approval.

9. Privacy, Federated Learning, and Ethical Considerations

Large-scale ECG pretraining requires substantial volumes of patient data, raising significant privacy and ethical concerns. Even when de-identified, ECG recordings may contain latent information linked to demographic attributes or rare medical conditions. Centralized data aggregation can conflict with privacy regulations such as HIPAA and GDPR.

Federated self-supervised learning offers a promising solution by enabling distributed model training without sharing raw patient data. In federated frameworks, participating institutions collaboratively update a shared model by exchanging encrypted parameter updates rather than signals themselves. This approach preserves data sovereignty while leveraging large-scale representation learning.

However, federated systems introduce new challenges, including communication overhead, heterogeneity across institutional datasets, and fairness across contributors. Ethical considerations extend beyond privacy to include demographic bias and data imbalance. Foundation models trained on skewed cohorts risk amplifying disparities in diagnosis and treatment recommendations.

Regulatory compliance further requires transparency, auditability, and accountability mechanisms. Bias detection, model monitoring, and explainability should be embedded throughout development pipelines. As ECG foundation models move toward clinical adoption, ethical AI principles must guide data collection, algorithm design, and deployment strategies to ensure equitable and responsible healthcare innovation.

10. Toward Multimodal ECG Foundation Models

Future ECG foundation models should expand beyond single-modality signal analysis to embrace multimodal physiological integration. The human cardiovascular system interacts dynamically with respiratory, neurological, and hemodynamic systems, suggesting that richer representations may emerge from cross-signal modelling.

Integrating ECG with photoplethysmography (PPG) can enhance pulse transit time estimation and blood pressure prediction. Combining ECG with electroencephalography (EEG) may support seizure detection or sleep-stage classification. Incorporating clinical notes alongside physiological signals enables contextual reasoning about patient history and comorbidities.

Cross-modal attention mechanisms within transformer architectures allow models to learn relationships between modalities, aligning temporal patterns across signals. Multimodal SSL objectives may include reconstructing one modality from another or enforcing cross-signal consistency.

Such integration moves toward holistic patient representation rather than isolated diagnostic tasks. Multimodal ECG foundation models could enable unified cardiovascular intelligence systems capable of supporting risk stratification, early disease detection, and personalized treatment planning.

11. Continual and Lifelong Learning

Cardiac physiology is not static; it evolves with aging, medication changes, stress levels, and comorbid conditions. Traditional deep learning models trained on static datasets struggle to adapt to evolving patient profiles. Continual and lifelong learning frameworks offer a pathway toward dynamic ECG intelligence systems.

Continual SSL allows models to update representations incrementally as new data becomes available. This capability supports patient-specific adaptation, enabling models to learn individual baseline rhythms and detect subtle deviations over time. Longitudinal monitoring becomes more effective when models incorporate historical context rather than relying solely on isolated recordings.

However, continual learning must address catastrophic forgetting, where adaptation to new data erases previously learned knowledge. Techniques such as regularization-based methods, replay buffers, and modular architectures can mitigate this risk.

Incorporating lifelong learning into ECG foundation models supports personalized risk prediction and proactive healthcare. Such systems could evolve alongside patients, providing adaptive diagnostics aligned with changing physiological states.

12. Computational Efficiency and Sustainability

Transformer-based foundation models often require significant computational resources for training and inference. Long ECG recordings and high-resolution multi-lead signals increase memory and energy demands. As healthcare AI scales globally, sustainability and accessibility become critical considerations.

Efficient attention mechanisms that reduce quadratic complexity to linear scaling offer promising solutions. Parameter-efficient fine-tuning strategies, such as adapter layers or low-rank updates, enable downstream adaptation without retraining full models. Knowledge distillation techniques can compress large foundation models into lightweight versions suitable for edge devices or wearables.

Energy consumption and carbon footprint reporting should become standard practice in ECG foundation model research. Balancing performance gains with computational sustainability ensures equitable access to advanced AI systems across diverse healthcare environments.

13. Roadmap for Next-Generation ECG Foundation Models

A mature ECG foundation model must satisfy several criteria. First, it should be pretrained on diverse, multi-institutional datasets to reduce demographic and device bias. Second, it should support multimodal fusion to integrate complementary physiological and clinical data. Third, uncertainty-aware prediction mechanisms must be embedded to support clinical decision-making.

Fourth, federated deployment capabilities should enable privacy-preserving collaboration across institutions. Finally, interpretability must be integrated into model design, ensuring alignment with physiological reasoning and regulatory standards.

Achieving this vision requires interdisciplinary collaboration among engineers, clinicians, ethicists, and policymakers. The roadmap toward next-generation ECG foundation models reflects a shift from narrow task optimization toward scalable, trustworthy, and adaptive cardiovascular intelligence systems capable of supporting precision medicine.

14. Conclusion

Self-supervised learning (SSL) and foundation-model paradigms represent a transformative shift in ECG signal processing, redefining how cardiovascular intelligence systems are designed, trained, and deployed. Traditional supervised approaches, while effective under controlled conditions, are fundamentally constrained by annotation scarcity, demographic bias, and limited generalization. In contrast, SSL enables models to extract physiologically meaningful representations directly from large-scale unlabelled ECG corpora, thereby reducing dependence on expert labelling and improving scalability across institutions and devices. By decoupling representation learning from narrowly defined tasks, foundation models promote transferable, reusable embeddings capable of supporting a broad spectrum of downstream clinical applications.

Till now, foundational advances in transformer architectures, contrastive learning frameworks, masked modelling strategies, and generative probabilistic modelling have collectively established the theoretical backbone for this transition. Transformers introduced global attention mechanisms capable of modelling long-range rhythm dependencies. Contrastive learning

demonstrated the power of instance-level discrimination for label-efficient representation learning. Masked modelling enabled contextual reconstruction of cardiac waveforms, reinforcing temporal coherence and morphology awareness. Generative approaches, including variational autoencoders and adversarial networks, provided probabilistic latent spaces for uncertainty estimation, anomaly detection, and data synthesis. Together, these innovations laid the groundwork for scalable ECG foundation models that extend beyond isolated classification tasks.

Looking forward, the next frontier in ECG foundation modelling lies in multimodal integration, continual learning, federated scalability, and rigorous clinical validation. Integrating ECG with complementary modalities such as PPG, EEG, and clinical text will enable holistic patient representations. Continual learning frameworks will support adaptation to evolving patient physiology, while federated SSL will facilitate privacy-preserving collaboration across healthcare systems. Equally important is the incorporation of interpretability, fairness assessment, and regulatory compliance into model development pipelines.

Ultimately, universal biomedical foundation models must balance scalability with trustworthiness, performance with transparency, and innovation with ethical responsibility. The convergence of SSL and ECG analytics signals not merely an incremental improvement, but a paradigm shift toward adaptive, generalizable, and clinically integrated cardiovascular AI systems capable of supporting precision medicine at scale.

References

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *ICML*.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers. *NAACL*.

Franceschi, J.-Y., Dieuleveut, A., & Jaggi, M. (2019). Unsupervised scalable representation learning for time series. *NeurIPS*.

Goldberger, A. L., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*.

Golany, T., & Radinsky, K. (2019). PGANs for ECG data synthesis. *AAAI*.

Hannun, A. Y., et al. (2019). Cardiologist-level arrhythmia detection. *Nature Medicine*.

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised representation learning. *CVPR*.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2021). Masked autoencoders are scalable vision learners. *CVPR*.

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. *ICLR*.

LeCun, Y., et al. (2021). Self-supervised learning: The dark matter of intelligence.

Moody, G. B., & Mark, R. G. (2001). The MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*.

Rajpurkar, P., et al. (2017). Cardiologist-level arrhythmia detection. *Nature Medicine*.

Serrano, S., & Smith, N. A. (2019). Is attention interpretable. *ACL*.

Vaswani, A., et al. (2017). Attention is all you need. *NeurIPS*.

Zhou, X., & Li, J. (2022). Temporal transformer networks for long-range ECG analysis. *IEEE TNSRE*.