Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Toward Standardized ECG Playback for Simulation-Based Interpretation Training: A Critical Narrative Review and Research Agenda

Fahmi Abu-Owaimer
Alfaisal University
Sadeq Abu-Dawas
saabudawas@alfaisal.edu

Alfaisal University
Loai Abu-Owaimer
Alfaisal University
Abdalrahman A. Thaher
Biruni University

Systematic Review

Keywords:

Posted Date: August 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10502616/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

ECG interpretation accuracy and inter-rater agreement stay limited across every level of training, and the interventions available today only nudge that number in the right direction. What could help is a practice stimulus that is reproducible and expert-verified; one that every institution can deliver in exactly the same way. Simulation platforms and commercial ECG testers already exist, and each solves part of the problem, but none combines an openly traceable, expert-verified waveform source with a playback pathway simple and cheap enough to install for reproducible training across multiple sites. This review asks whether standardized ECG repositories with documented clinical or expert-derived annotation processes, among them MIT-BIH, PTB-XL, and PTB Diagnostic, could supply that stimulus when paired with low-cost digital-to-analog playback hardware, and it lays out a staged framework for actually testing the idea. The underlying technology is not new: microcontroller- and DAC-based ECG signal generation has already been demonstrated in peer-reviewed proof-of-concept systems and commercial devices, though nobody has yet benchmarked how well it interoperates across different monitor brands. The educational case rests on deliberate practice, mastery learning, cognitive load theory, and functional task alignment, and together these suggest something reassuring: fidelity only needs to reach whatever threshold supports correct perceptual discrimination, not go beyond it. The evidence on learner stress is messier and does not support any tidy story where more realism simply calms people down. Taken together, database-sourced ECG playback looks technically feasible and educationally well motivated, but whether it outperforms existing training methods is still an open question. What comes next, then, is a three-phase validation agenda: bench-level fidelity testing, a randomized crossover trial, and a multisite pragmatic study offered here as the necessary step before anyone can claim an educational benefit.

1. Introduction

Few tools in medicine are as widely used, and as inconsistently taught, as the electrocardiogram. Reliable skill acquisition generally requires deliberate, repeated practice against a well-defined task with a known correct answer and outcome-referenced feedback, continued until a defined performance standard is met. Most current ECG teaching: lectures, textbook tracings, or commercial simulation mannequins does not consistently offer standardized, independently verified stimuli or a mastery-based progression rule. Inconsistent access to such stimuli may contribute to variability in ECG learning outcomes, though it is unlikely to be the sole contributor.

This review evaluates whether combining standardized ECG repositories with documented clinical or expert-derived annotation processes with low-cost microcontroller- and DAC-based playback hardware could supply that missing element affordably. Three technical paradigms are often conflated in this space: database-sourced playback, mathematical-model waveform synthesis, and commercial patient-simulator or tester architecture; and are treated separately in Section 5. The educational case (Section 3) draws on deliberate practice, mastery learning, cognitive load theory, experiential learning, and competency-based education. Learner stress (Section 7) is treated as a mixed and titratable variable rather than one that repeated exposure straightforwardly reduces. The remaining sections set out a

conceptual design brief, implementation and governance considerations, and a three-phase research agenda.

Existing ECG teaching tools already demonstrate the value of standardized case exposure, and commercial simulator or tester platforms (Section 5.3) already demonstrate that waveform output to monitoring equipment is feasible; neither claim originates here. Based on the systems and literature reviewed in Section 5, rather than an exhaustive patent or product search, the combination proposed here is openly citable, repository-derived waveform provenance paired with a low-cost, installable playback pathway for reproducible simulation-based education across sites was not identified among the closed commercial systems or screen-only computer-based tools surveyed. This novelty claim should be read as provisional and scoped to the literature and commercial landscape reviewed here, not as a freedom-to-operate or prior-art determination, and it is not itself load-bearing for the educational or technical arguments that follow. Section 3.7 addresses directly why physical playback, rather than an equivalent screen-based tool, is the appropriate solution for this specific gap.

2. Methods

We conducted a critical narrative review because the research question integrates heterogeneous evidence from medical education, ECG dataset curation, and biomedical signal engineering. Reporting was informed by the SANRA framework [26].

Literature was identified through targeted searches in PubMed/MEDLINE, Google Scholar, IEEE Xplore, and PhysioNet's database documentation, conducted between January and July 2026 (final search date: July 18, 2026), across five domains: ECG interpretation accuracy and inter-observer variability; standardized ECG signal repositories (MIT-BIH, PTB-XL, PTB Diagnostic, and KURIAS-ECG); electronics-based ECG signal generation and technology standards; simulation pedagogy (deliberate practice, mastery learning, cognitive load theory, experiential learning, competency-based education, and simulator fidelity); and learner stress during simulation. A representative search string for the ECG-accuracy domain in PubMed was: ("electrocardiogram" OR "ECG" OR "EKG") AND ("interpretation" OR "accuracy" OR "competence") AND ("education" OR "training" OR "simulation"); analogous strings, substituting domain-specific terms, were used for the remaining four domains. Searches were restricted to English-language sources. Search terms combined domain-specific keywords with Boolean operators, and reference lists of retrieved articles were hand-searched. Source identification, screening, and verification were performed by the author team; approximately 180 sources were screened and 38 retained in the final reference list. Peer-reviewed empirical studies, systematic reviews, foundational theoretical works, and official standards or database documentation were prioritized; non-peer-reviewed vendor material was retained only where labeled explicitly as illustrative. Where sources offered conflicting findings (e.g., the mixed learner-stress literature in Section 7), both directions of evidence were retained and reported rather than resolved by selective citation.

The technical specifications, cost estimates, and validation agenda that follow are proposals drawn from this literature, not findings from a built or tested system.

3. Educational Rationale for Standardized ECG Playback

3.1 Deliberate Practice and Mastery Learning

Ericsson's account of deliberate practice holds that expert performance comes from effortful, structured practice on well-defined tasks with immediate feedback, not from repetition or accumulated experience [15]. McGaghie and colleagues applied this to simulation-based mastery learning (SBML), in which every learner reaches a fixed passing standard before advancing, rather than a fixed-time model where outcomes are left to vary [16]. Issenberg and colleagues' BEME systematic review identified the features of high-fidelity simulation most associated with effective learning: feedback, repetitive practice, curriculum integration, range of task difficulty, multiple learning strategies, capture of clinical variation, a controlled environment, individualized learning, defined outcomes, and simulator validity [17]. A standardized, database-sourced ECG playback module would need to offer most of these features to be educationally useful.

The playback hardware proposed here (Section 8) supplies a standardized, difficulty-tagged stimulus, satisfying the well-defined-task requirement above, but does not itself score learner responses or enforce a mastery standard. Feedback and mastery enforcement would come from an instructor, a paired answer key, or software layered on top. The device is therefore necessary but not sufficient for a deliberate-practice or mastery-learning curriculum: it removes stimulus variability as a confound, while the surrounding educational program delivers the practice itself.

3.2 Cognitive Load Theory

Cognitive load theory distinguishes intrinsic load (task difficulty), extraneous load (added by poor instructional design), and germane load (effortful processing that contributes to learning) [18]. Lapierre and colleagues found that simulation design features: fidelity level, task complexity, instructional support measurably affect cognitive load in novice learners [19]. Unreliable equipment, such as a mannequin whose ECG display behaves inconsistently across sessions, likely adds extraneous load unrelated to the diagnostic task. A standardized, repeatable playback stimulus is one plausible way to reduce this extraneous load, though this mechanism has not been evaluated directly in ECG education.

3.3 Experiential Learning

Kolb's experiential learning cycle: concrete experience, reflective observation, abstract conceptualization, and active experimentation has been used in medical education to link real clinical experience with case discussion and simulation [20]. A standardized ECG playback module would function mainly as a controlled substitute for concrete experience, supporting repeated passes through the full cycle, including debriefing and reflection, more reliably than the variable and opportunistic ECG exposure available during clinical rotations.

3.4 Competency-Based Education

Competency-based education (CBE), as defined by Frank and colleagues, organizes training around the demonstrated achievement of specific, observable outcome abilities, defined in advance and assessed directly, rather than around completion of a prescribed curriculum [21]. A standardized ECG module fits CBE only when paired with an explicit competency framework, defined performance standards, and a way of tracking each learner's progress; used alone, it delivers ECG exposure without demonstrating competency. This pairing is a design requirement for the proposed module (Section 8), not something that follows automatically from the hardware.

3.5 Fidelity as Functional Task Alignment

Hamstra and colleagues went further than reframing simulator fidelity, recommending that the term be abandoned altogether in favor of two distinct concepts: physical resemblance (how closely a simulator matches its real-world counterpart) and functional task alignment (whether it supports practice of the actual cognitive or perceptual task being trained). They noted that across several empirical studies the degree of fidelity appeared largely independent of educational effectiveness [22]. A moderate-fidelity, database-sourced ECG signal may be pedagogically adequate provided it preserves the waveform features learners need to discriminate: ST-segment deviation, P-wave morphology, R-R interval timing. This hypothesis, motivated by the broader simulation-fidelity literature, has not been tested for ECG interpretation training specifically; Section 11 proposes to test it directly.

The technical specifications in Section 6.1 represent candidate minimum thresholds for preserving these curriculum-relevant features, drawn from existing engineering and clinical-standards literature. A device with too little fidelity would fail the functional-task-alignment test as surely as one with no standardized stimulus at all; the point is that fidelity should be pinned to the minimum sufficient for the task, not maximized as an end in itself.

3.6 Educational Outcomes Framework

The framework below distinguishes primary outcomes (the direct target of the intervention), secondary outcomes (related but insufficient alone to justify the intervention), and tertiary outcomes (implementation variables that determine adoption, Section 9).

Level	Outcome	Operational definition	Likely measure
Primary	Interpretation accuracy	Correct case-level diagnosis or recognition of the target ECG finding on a standardized case bank	Percentage correct against a predefined answer key
Primary	Diagnostic agreement	Agreement with the expert-derived key across learners	Kappa or percent agreement
Secondary	Time to diagnosis, confidence, retention	Fluency, self-reported confidence, and delayed-retest accuracy	Median time; Likert scale; retest at 2–6 weeks
Secondary	Cognitive workload	Perceived processing burden during interpretation (Section 3.2)	NASA-TLX or equivalent
Secondary	Transfer to physical monitor reading	Performance on monitor-displayed cases versus static tracings	Parallel-format comparison test
Tertiary	Faculty usability, technical reliability, workflow burden	Ease of setup, scenario-startup success, maintenance and setup time (Section 9)	Faculty survey; operational failure log; site timing log

Interpretation accuracy and diagnostic agreement are the only primary outcomes; secondary and tertiary measures are hypothesis-generating and should not, alone, establish benefit if the primary outcome is null (Section 11, Phase 2; Appendix A.2).

3.7 Why Physical Playback Rather Than a Screen-Based Interface

The educational case in Sections 3.1 to 3.6 rests on properties where standardized, expert-verified, difficulty-tagged waveforms that a screen-based application could in principle deliver at lower cost and without the calibration, licensing, and maintenance burden described in Section 9. For interpretation practice conducted in isolation, evaluating a scrollable or static waveform outside a broader clinical scenario, a screen interface is likely equivalent or superior on cost and logistical grounds, and this review does not claim otherwise. Curricula that only require isolated interpretation drills should use a screen-based tool, not the device proposed here.

The case for physical playback hardware rests on a narrower requirement. ECG interpretation is rarely trained in isolation within mastery-learning or team-based simulation curricula; it is typically embedded in mannequin-based or team scenarios (for example, ACLS or deteriorating-patient exercises) in which the ECG must appear on the same bedside monitor the learner will use in practice, integrated with other vital signs, and encountered through the physical workflow of lead placement and monitor operation. No screen-based tool or static tracing can inject a standardized, database-sourced signal into that physical monitor display. Existing mannequin platforms generate ECG signals internally from a limited, non-standardized library, and existing commercial testers (Section 5.3) are built to verify monitor function,

not to deliver open educational content. Physical playback hardware is proposed as the missing link for this specific use case, not as a general substitute for screen-based interpretation practice.

This reframes the functional-task-alignment argument in Section 3.5: the task that requires physical fidelity is not ECG interpretation per se, but ECG interpretation embedded in equipment-integrated simulation. This distinction is itself a hypothesis rather than a demonstrated finding; it has not been tested whether monitor-integrated delivery adds measurable educational value over an equally standardized screen-based tool, and the Phase 2 crossover trial (Section 11; Appendix A.2) is designed to test it directly by comparing playback hardware against both static tracings and a matched screen-based tool, rather than against static tracings alone. This also sharpens the distinction from a commercial tester (Section 5.3): a tester exists to verify monitor function against manufacturer specifications, whereas the module proposed here exists to deliver expert-verified educational content through monitor equipment as part of a training curriculum. That difference in intended use, not merely cost or openness, is what should determine the regulatory and validation pathway in Section 9.

4. The Problem: ECG Interpretation Variability

A meta-analysis published in JAMA Internal Medicine pooled diagnostic-accuracy data across dozens of studies. Accuracy rose with training level but stayed limited throughout: roughly 42.0% for medical students, 55.8% for residents, 68.5% for practicing physicians, and 74.9% for cardiologists at pretraining assessment [1]. The same review found that investigators rarely reused prior test instruments and often lacked an explicit procedure for establishing the correct answer against which learners were scored [1].

This inconsistency carries real clinical weight, though the evidence is strongest for challenging or equivocal tracings rather than for ECGs generally. A survey of 124 physicians producing more than 4,000 interpretations of 36 ECGs tied to possible ST-elevation myocardial infarction found inter-reader agreement that the authors characterized as poor, a kappa of 0.33; a value that falls in the “fair” range by Landis and Koch criteria on this specifically difficult category [2]. A case-vignette study of general practitioners and cardiologists found accuracy and agreement dropping sharply for subtle findings such as fascicular or bundle-branch blocks, while agreement held up better for clearer findings like atrial fibrillation [3]. A chart-review study found that physicians often missed incorrect computerized interpretations, and that the same cardiologists corrected these errors far more often when reading as the primary physician, who knew the patient, than when performing routine overreading [4]. Taken together, inconsistency is worst for challenging or equivocal cases, not spread evenly across any two ECGs; a distinction that matters because standardized stimuli would not be expected to resolve disagreement among experts themselves. The trainable share of interpretation error, not all of it, is what this proposal targets.

A companion systematic review by Oh and colleagues found that instructional modality, method, and interpretation approach each measurably shape ECG training outcomes [32]. Cook, Oh, and Pusic catalogued the features and validity evidence behind published ECG skill assessments and found

substantial heterogeneity in how ECG competence is measured across studies [33]. Two meta-analyses of computer-assisted ECG instruction reached differing conclusions: Oh and colleagues found a small but statistically significant advantage over face-to-face teaching (standardized mean difference 0.23, 95% CI 0.09–0.36) [32], whereas Viljoen and colleagues found no overall difference (standardized mean difference 0.32, 95% CI − 0.09 to 0.74), with benefit emerging only in blended designs offering unlimited access to materials or deliberate practice with feedback [34]. This literature establishes that some approaches already help; whether reproducible, expert-anchored waveform stimuli delivered through low-cost hardware add further benefit has not yet been tested.

5. Standardized ECG Sources: Three Technical Paradigms

Three related but non-equivalent technical paradigms are often discussed together in this space; they carry different strengths, limitations, and appropriate use cases, and are treated separately here.

5.1 Database-Sourced Playback

PhysioNet's MIT-BIH Arrhythmia Database is a landmark expert-annotated resource comprising 48 half-hour, two-channel ambulatory ECG recordings from 47 subjects, sampled at 360 samples per second per channel at 11-bit resolution, with disagreements between independent cardiologist annotators resolved into a single reference annotation covering approximately 110,000 heartbeats [5, 36]. This resource, and the broader PhysioNet/PhysioBank infrastructure that hosts it, was reported in 2001 as having been used at approximately 500 sites worldwide since 1980 [36, 37]. It is a strong source of validated ground truth for arrhythmia-focused waveform segments, but is a two-channel ambulatory dataset, not a general-purpose 12-lead educational standard, and claims of full 12-lead standardized playback should not rest on MIT-BIH alone.

PTB-XL [6, 35], hosted on PhysioNet (version 1.0.3), is a 12-lead dataset of 10-second clinical recordings with structured, multilabel diagnostic statements assigned by up to two cardiologists [6, 35]. Its 12-lead structure makes it more suitable than MIT-BIH for teaching tasks requiring complete diagnostic ECGs. The PTB Diagnostic ECG Database [7] is a related but distinct resource: 549 records from 290 subjects using a 15-lead configuration that includes the standard 12 leads plus three Frank XYZ leads. KURIAS-ECG offers a standardized diagnostic ontology across 12 leads, derived partly from machine-assisted diagnosis statements with quality-filtering during curation; including exclusion of records flagged as poor quality or as having reversed arm leads [38], but is currently unavailable for download until further notice, following an internal audit at the originating institution [8]; it is discussed here only as an illustration of ontology standardization, not as a near-term deployable source.

The datasets differ in how their labels were produced: MIT-BIH's annotations are independent cardiologist beat-by-beat markings with adjudicated disagreement; PTB-XL's are structured, multilabel diagnostic statements from up to two cardiologists; KURIAS-ECG's are derived from ontology mapping applied to machine-assisted statements with quality filtering. Any curriculum built on these sources

should state which dataset's labeling process underlies each case's answer key, rather than treating "expert-adjudicated" as a single undifferentiated property.

These repositories improve waveform provenance and repeatability relative to ad hoc or synthetic tracings, but do not reproduce the full realism of bedside ECG acquisition. MIT-BIH is two-channel ambulatory data from the late 1970s; PTB-XL and PTB Diagnostic provide structured clinical recordings but do not reproduce local monitor-filter behavior, lead-placement error, or simulation-center workflow; KURIAS removed poor-quality and reversed-lead data during curation. Curated repositories may also underrepresent motion artifact, lead misplacement, paced rhythms, and multi-pathology cases that learners encounter in practice. Educational claims here are scoped to waveform-source standardization, not to the full clinical acquisition environment (Section 12).

5.2 Mathematical-Model Waveform Synthesis

A separate paradigm generates ECG waveforms from mathematical or physiological models rather than played-back recordings. One peer-reviewed system used four coupled-oscillator and reaction-diffusion models on a 32-bit microcontroller with digital-to-analog converters to generate approximately 32 distinct cardiac rhythms and full 12-lead profiles [11]. This maximizes tunability and rhythm diversity, including rhythms underrepresented in any single recorded dataset, but the resulting waveforms are not patient-derived and require separate validation against real recorded morphology before use as a ground-truth training stimulus.

5.3 Commercial Patient-Simulator and Tester Architectures

A third paradigm is the commercial ECG simulator/tester, exemplified by the Fluke Biomedical ProSim family [28] and comparable commercial rhythm simulators, designed for biomedical engineering technicians to verify patient-monitor function rather than for open educational reproducibility. These systems optimize interoperability, amplitude accuracy, and documented lead-architecture behavior (Section 6), but are closed, proprietary, and not built around database-sourced, expert-labeled educational content. The distinction from the module proposed here (Section 8) is one of intended use rather than only of openness or cost: a tester verifies that a monitor correctly reproduces a known signal against manufacturer specifications, whereas an educational module exists to deliver expert-verified, curriculum-linked content through that monitor as part of a training program (Section 3.7). That distinction should determine the applicable regulatory and validation pathway (Section 9), and any implementation should state explicitly which purpose a given device serves.

5.4 Comparing the Three Paradigms

Paradigm	Provenance / ground truth	Strengths	Limitations
Database-sourced playback (MIT-BIH, PTB-XL, PTBDB)	Real recordings with documented clinical or expert-derived annotation processes	Traceable to validated ground truth; supports outcome-referenced feedback (Section 3.1)	Lead count and diagnostic diversity limited by dataset; licensing/availability varies (e.g., KURIAS)
Mathematical-model synthesis	Physiological/mathematical model, not patient-derived	High tunability; can generate rare or extreme rhythms on demand	Not independently expert-verified against real morphology; requires separate validation
Commercial simulator/tester	Manufacturer-specified, proprietary	Documented interoperability, amplitude accuracy, and lead-architecture behavior	Closed/proprietary; high cost; not built around open educational content

The proposal developed here (Section 8) draws primarily on database-sourced playback, because its documented, clinically or expert-derived annotation provenance is what allows it to function as the well-defined task with a known correct answer that Section 3.1 identifies as necessary, while acknowledging the lead-count and licensing limitations specific to that paradigm.

6. Technical Feasibility: From Digital Data to Analog Signal

Several proof-of-concept and commercial systems demonstrate that converting stored digital ECG data into a physical analog signal is technically feasible, although interoperability across monitor brands and formal educational validation remain limited. Published peer-reviewed examples include an Arduino-based educational ECG generator for biomedical engineering laboratory instruction [12], a vital-signs simulator design for patient-monitor testing [13], and a 12-lead patient simulator combining microcontrollers, DACs, and mathematical waveform models capable of reproducing roughly 32 distinct rhythms [11]. A separate digital-to-analog 12-lead reconstruction system reported grand-averaged root-mean-square voltage differences of 20.8 microvolts per channel when the reconstructed signal was re-digitized through the same manufacturer's analog-to-digital converter, and 28.4 microvolts when re-digitized through a different manufacturer's converter, demonstrating both that useful fidelity is achievable and that cross-manufacturer compatibility measurably affects signal accuracy [10].

Any device claiming diagnostic-style adult ECG playback should be evaluated against the American Heart Association's technical standard for ECG acquisition, which specifies preservation of low-frequency content to approximately 0.05 Hz and high-frequency content to at least 150 Hz, with a minimum digital sampling rate of about 500 samples per second per channel [9]; the standard permits the low-frequency requirement to be relaxed to 0.67 Hz or below for linear digital filters with zero phase

distortion [9]. This exceeds MIT-BIH's native 360 samples per second, so a system playing back MIT-BIH-derived waveforms while claiming full diagnostic-bandwidth fidelity would need appropriate interpolation or an explicit statement that it targets teaching-relevant morphology rather than diagnostic-grade bandwidth.

6.1 Conceptual Design Brief

The table below outlines feasibility and the key constraints a future device would need to satisfy; the figures given are candidate minimums derived from existing engineering and clinical-standards literature, not validated requirements.

Design element	Core constraint	Basis
Source waveform provenance	Named, citable dataset with known annotation provenance, stated explicitly per case (Section 5.1)	MIT-BIH for arrhythmia pedagogy; PTB-XL/PTBDB for 12-lead cases; KURIAS excluded pending open access
Electrode-output interface	Configurable interface supporting defined 3-, 5-, or 10-electrode configurations and their corresponding monitor-derived leads, not a literal "3-/5-/12-output" connector	A 12-lead display is computed by the monitor from limb and precordial electrode potentials via the Wilson central terminal, not delivered as 12 independent signals; conflating electrode count with lead count is a primary source of overclaiming (Section 10)
Signal-generation channel count	A synchronized multichannel D/A output stage, with channel count set by the intended electrode configuration	A single-channel DAC (e.g., MCP4921) can support single-limb-lead proof-of-concept playback only; simultaneous multi-electrode output requires a synchronized multichannel DAC arrangement or equivalent multichannel analog-output architecture
Playback fidelity target	Approximately 500 samples/s/channel and 0.05–150 Hz bandwidth per channel for adult diagnostic-style claims; 12-bit per-channel resolution as a starting point	AHA/ACC ECG technology standard [9]; MCP4921-class 12-bit per-channel DAC specifications [14]
Calibration and monitor testing	A bench calibration procedure and at least one named monitor family tested before any compatibility claim	Compatibility is empirical and monitor-specific, not assumed (Appendix A)

Finer-grained parameters: source/output impedance, noise and distortion targets, multi-vendor interoperability tiers should be finalized in protocol-specific engineering work. A standardized waveform library is not equivalent to a standardized ECG experience under every lead placement and monitor filter configuration; electrode placement and monitor-side filtering materially affect waveform appearance, so standardization claims are scoped to the waveform source specifically, not to the full downstream presentation (Section 5.1).

7. Learner Stress: A Titratable, Context-Dependent Variable

High-fidelity simulation can provoke genuine physiological and psychological stress, and the literature on repeated exposure is mixed rather than uniformly reassuring: one study found lower anxiety and higher confidence after repeated high-fidelity sessions [23], while others found that high-fidelity and procedural simulation increased self-perceived anxiety with modality-dependent cortisol responses [24], or produced complex, role-sensitive stress patterns rather than a simple reduction in one study, cortisol, alpha-amylase and testosterone responses differed according to whether a student was assigned the physician, nurse or assistant role within the same scenario [25]. Simulation-related stress is context-dependent and can be either beneficial, supporting engagement through germane load (Section 3.2), or harmful, adding extraneous load, depending on intensity, learner level, and how deliberately it is managed.

ECG playback standardization should not be expected to reduce learner stress directly: the proposed device is a waveform source, not a stress-management intervention. The specific, testable mechanism proposed here is that a standardized, repeatable stimulus could let instructors separate stress attributable to case difficulty, an intended, curriculum-relevant variable, from stress attributable to equipment inconsistency across sessions or sites, an extraneous, unintended variable, enabling more deliberate titration of the former (Sections 3.1–3.2). This is identified as a secondary, exploratory outcome for the Phase 2 trial (Section 11), not a primary justification for the device.

8. Toward a Standardized, Installable Module: Proposed Design

The evidence in Sections 3 through 7 motivates a specific design direction, offered here as a testable engineering and educational hypothesis rather than a validated solution: a compact electronics module built around a microcontroller and digital-to-analog converter, pre-loaded with a curated library of database-sourced ECG waveforms (primarily MIT-BIH for arrhythmia pedagogy and PTB-XL/PTBDB for 12-lead cases), designed to the specifications in Section 6.1 and targeting functional task alignment (Section 3.5) rather than maximal physical fidelity. Figure 1 summarizes the proposed architecture, and Fig. 2 summarizes the proposed timeline: three sequential validation phases (Section 11), followed by a fourth, non-validation deployment phase.

A minimal viable design, per Section 6.1, would include: a microcontroller programmed with a selectable, difficulty-tagged waveform library; a synchronized multichannel D/A output stage, with channel count set by the intended electrode configuration rather than a single-channel DAC feeding a generic multi-lead connector; a configurable electrode-output interface with an explicitly stated electrode configuration (3-, 5-, or 10-electrode) and the corresponding monitor-derived leads; and a selection interface allowing instructors to sequence cases by difficulty. Higher-resolution DAC architectures (e.g., 16-bit per channel) may be worth evaluating when subtle morphological discrimination is required, though their incremental educational value over a 12-bit baseline remains to be established rather than assumed. Tracking individual learners' progress toward a defined passing standard, consistent with competency-based

education (Section 3.4), would be a curricular or software-layer function built on top of the hardware, not a core property of the device itself. Monitor compatibility has been demonstrated in some published systems but should be verified empirically against named monitor families before any claim of general compatibility.

9. Implementation & Governance

Because the proposed module would be intended for simulation or test use rather than direct patient diagnosis or treatment, its regulatory status depends on intended use, labeling, distribution model, and jurisdiction, and should be determined prospectively rather than assumed from analogy; the Fluke ProSim family's reaching regulated markets [28] does not predetermine this device's classification. Relevant frameworks include risk-management principles from ISO 14971 [27] and, where applicable, the IEC 60601 series for electrical safety and electromagnetic compatibility [31]; reference to these standards is not a claim of compliance. Three practical requirements follow: a permanent label restricting the device to training use and prohibiting connection to a live patient; confirmed, per-dataset licensing for MIT-BIH, PTB-XL, and PTBDB, with KURIAS-ECG's unavailability following an internal audit (Section 5.1) treated as a near-term deployment barrier rather than a minor caveat; and version control over shared waveform libraries so institutions can verify they are running an identical, unmodified case set.

Adoption will likely depend as much on human factors; faculty willingness to use the device within a consistent prebriefing/feedback/debriefing structure (Section 3), onboarding burden, scenario setup and reset time, calibration ownership, and local biomedical-engineering or IT support; as on technical performance. A device with adequate fidelity but high setup burden or unclear maintenance ownership is unlikely to be adopted regardless of its educational rationale; this is an explicit tertiary outcome in Section 3.6 and an implementation endpoint in Appendix A (Phase 3).

10. Evidence Synthesis and Knowledge Gaps

The table below summarizes the review’s central claims, the evidence that supports them, their limitations, and the research implication that follows.

Domain	Available evidence	Evidence limitation	Research implication
ECG interpretation variability	Meta-analytic data show accuracy remains limited at every training level; inter-reader data show agreement is poorest for challenging or equivocal tracings, although pooled accuracy did not differ by test difficulty (Section 4)	Drawn largely from cross-sectional accuracy studies, not training-outcome trials	Motivates testing whether standardized stimuli reduce the trainable share of error
Database-sourced ground truth	MIT-BIH provides validated, adjudicated arrhythmia annotations; PTB-XL and PTBDB provide structured 12-lead diagnostic statements (Section 5.1)	MIT-BIH is two-channel and non-12-lead; KURIAS-ECG is not currently available for open deployment	12-lead curricula should draw on PTB-XL/PTBDB rather than MIT-BIH alone
Signal-generation feasibility	Peer-reviewed and commercial systems demonstrate microcontroller-based ECG playback (Section 6)	Interoperability and fidelity benchmarking across monitor manufacturers remain limited	Phase 1 bench testing (Appendix A) is required before any general compatibility claim
Educational effectiveness	Deliberate practice, mastery learning, and cognitive load theory provide a plausible rationale (Section 3)	No trial has tested database-sourced playback against existing ECG training methods, or against a screen-based equivalent (Section 3.7)	Phase 2 randomized crossover trial (Section 11) is the direct test
Learner stress	Some studies show reduced anxiety with repeated high-fidelity exposure; others show increased anxiety or workload (Section 7)	Evidence is heterogeneous and modality-dependent	Retained only as a secondary, exploratory endpoint (Section 11)
Cost and deployment	Mannequin-platform procurement can reach tens of thousands of pounds per contract [30]	Total ownership cost: assembly, calibration, maintenance, faculty support, and any regulatory pathway is unestimated	Cost should not be treated as the proposal's primary justification

11. Proposed Validation Research Agenda

Because this review makes technical and educational claims not yet empirically demonstrated for this application, three sequential studies are proposed, each gating the next (Fig. 2): bench validation of fidelity and interoperability (Phase 1 – Fig. 3), a randomized crossover educational trial comparing hardware, a matched screen-based tool, and static tracings (Phase 2, Section 3.7), and a multisite pragmatic implementation study (Phase 3). Only Phase 1 is a prerequisite for a technical fidelity claim; Phases 2 and 3 would be required before the educational and cross-site standardization claims in

Sections 3 and 8 could be considered demonstrated rather than hypothesized. Full endpoints, planning scales, and statistical approach for each phase are given in Appendix A; the table below summarizes the essential logic.

Phase	Purpose	Primary endpoint	Planning scale
1. Bench validation	Establish signal fidelity and monitor interoperability before any compatibility claim	RMS voltage error against source, evaluated against a margin defined prospectively from clinical standards and expert consensus (Appendix A), not from prior systems' observed error alone [10]	20–30 cases across 2–3 named monitor families
2. Randomized crossover trial	Test educational benefit of hardware vs. a matched screen-based tool vs. static tracings	ECG interpretation accuracy on a standardized test bank	40–60 learners (planning range only; not power-calculated)
3. Multisite pragmatic study	Test cross-site reproducibility and implementation feasibility	Between-site variability in performance on a common case set, operationalized via a multilevel model (e.g., intraclass correlation), combined with successful delivery of a predefined proportion of cases without technical deviation	4–8 sites, ~ 20–30 learners/site (naive figure; requires cluster-aware power planning)

This design follows the general logic of recent centrally designed, large-scale multisite simulation programs, which have shown that shared curricula can be delivered with reasonably consistent learner-reported outcomes across sites, even though such programs have not previously focused specifically on ECG playback hardware [29].

12. Limitations

No formal quality-appraisal instrument or PRISMA-compatible search log was used, consistent with the scope set out in Section 1. Two pedagogical claims central to this proposal remain untested rather than demonstrated, and Section 11 is designed specifically to test both. First, the claim in Section 3.5 that a moderate-fidelity, database-sourced signal is adequate for ECG interpretation training is extrapolated from the general simulation-fidelity literature, which addresses procedural and team-based skills more often than single-modality diagnostic pattern recognition. Second, the claim in Section 3.7 that physical, monitor-integrated delivery adds educational value beyond an equally standardized screen-based tool is a design rationale, not an empirical finding; if the Phase 2 trial finds no advantage over a matched screen-based comparator, the case for hardware over software would need to rest on integration into team-based or mannequin simulation workflows specifically, rather than on interpretation accuracy alone. The source datasets discussed in Section 5 are heterogeneous in lead count, licensing, and availability, and this heterogeneity materially affects what any given implementation can credibly claim.

No direct educational-effectiveness trial for this or a closely comparable device currently exists; the effect sizes cited from mastery-learning research (Section 3.1) are drawn predominantly from procedural skills (catheter insertion, thoracentesis, lumbar puncture) and may not transfer directly to a purely interpretive skill such as ECG reading. No costing of the proposed module is attempted here: the single procurement figure cited for an existing commercial platform is contextual only, and no component-level build cost, assembly and calibration cost, or engineering-labor estimate for the proposed device has been established. Finally, a standardized waveform source does not eliminate variability introduced by electrode placement, monitor-side filtering, instructor practice, or debriefing quality; this review's standardization claims are scoped specifically to waveform-source standardization, not to the full downstream simulation experience.

13. Conclusion

Database-derived ECG playback is technically feasible and could provide reproducible, provenance-traceable stimuli for simulation-based ECG education. However, its educational advantage over static tracings and existing simulators has not been demonstrated. Bench testing should first establish signal fidelity and monitor interoperability, followed by controlled educational evaluation and multisite implementation research. Until such evidence is available, the proposed system should be regarded as a testable educational technology rather than a validated training intervention.

Declarations

Ethics approval and consent to participate: Not applicable. This manuscript is a narrative review and design framework; it did not involve human participants, animal subjects, or identifiable personal data, and no new primary data were collected.

Data availability: No new datasets were generated or analyzed for this manuscript. Publicly available third-party datasets discussed (MIT-BIH, PTB-XL, PTB Diagnostic, KURIAS-ECG) are cited to their PhysioNet sources with version numbers where applicable; each carries its own license terms, which should be independently verified before reuse.

Code availability: No original software or firmware was developed or released with this manuscript.

Conflicts of interest: None declared.

Funding: None declared.

Author contributions (CRediT): F.A.-O. and S.A.-D. conceived and designed the review. F.A.-O. conducted the literature search, curated the literature, prepared the initial draft of the manuscript, and developed the figures. S.A.-D. supervised the project, contributed to the methodological framework and formal analysis, critically revised the manuscript, and coordinated the submission. L.A.-O. contributed to the methodology, literature review, figure development, and critical revision of the manuscript. A.A.T.

contributed to the methodological design, validated the conceptual framework, and critically revised the manuscript. F.A.-O. and S.A.-D. contributed equally and share first authorship. All authors reviewed, edited, and approved the final manuscript

Acknowledgments: None.

References

1. Cook DA, Oh SY, Pusic MV. Accuracy of physicians' electrocardiogram interpretations: a systematic review and meta-analysis. JAMA Intern Med. 2020;180(11):1461–1471. https://doi.org/10.1001/jamainternmed.2020.3989
2. McCabe JM, Armstrong EJ, Ku I, et al. Physician accuracy in interpreting potential ST-segment elevation myocardial infarction electrocardiograms. J Am Heart Assoc. 2013;2(5):e000268. https://doi.org/10.1161/JAHA.113.000268
3. Compiet SAM, Willemsen RTA, Konings KTS, Stoffers HEJH. Competence of general practitioners in requesting and interpreting ECGs: a case vignette study. Neth Heart J. 2018;26(7-8):377–384. https://doi.org/10.1007/s12471-018-1124-2
4. Anh D, Krishnan S, Bogun F. Accuracy of electrocardiogram interpretation by cardiologists in the setting of incorrect computer analysis. J Electrocardiol. 2006;39(3):343–345. https://doi.org/10.1016/j.jelectrocard.2006.02.002
5. MIT-BIH Arrhythmia Database v1.0.0. PhysioNet. https://doi.org/10.13026/C2F305
6. Wagner P, Strodthoff N, Bousseljot RD, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset, v1.0.3. PhysioNet; 2022. https://doi.org/10.13026/kfzx-aw45
7. PTB Diagnostic ECG Database v1.0.0. PhysioNet. https://physionet.org/content/ptbdb/1.0.0/
8. KURIAS-ECG: a 12-lead electrocardiogram database with standardized diagnosis ontology, v1.0. PhysioNet. Unavailable for download until further notice following an internal audit at Korea University Medical Center (notice posted Nov. 30, 2021). https://physionet.org/content/kurias-ecg/1.0/
9. Kligfield P, Gettes LS, Bailey JJ, et al. Recommendations for the standardization and interpretation of the electrocardiogram, Part I: the electrocardiogram and its technology. J Am Coll Cardiol. 2007;49(10):1109–1127. https://doi.org/10.1016/j.jacc.2007.01.024
10. Kothadia RB, Kulecz WB, Kofman IS, Black AJ, Grier JW, Schlegel TT. New system for digital to analog transformation and reconstruction of 12-lead ECGs. PLoS ONE. 2013;8(4):e61076. https://doi.org/10.1371/journal.pone.0061076
11. Quiroz-Juarez MA, et al. ECG patient simulator based on mathematical models. Sensors. 2022;22(15):5714. https://doi.org/10.3390/s22155714
12. Aqil M, Jbari A. ECG generator for educational biomedical engineering laboratory. Int J Online Biomed Eng (iJOE). 2022;18(2):181–199. https://doi.org/10.3991/ijoe.v18i02.25363

13. Narvaez JMN, Avendano GO. Design and implementation of vital signs simulator for patient monitor. J Telecommun Electron Comput Eng (JTEC). 2017;9(2-4):7–9. https://jtec.utem.edu.my/jtec/article/view/2350
14. Microchip. MCP4921 12-bit DAC product page and datasheet. https://www.microchip.com/en-us/product/MCP4921
15. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. Acad Emerg Med. 2008;15(11):988–994. https://doi.org/10.1111/j.1553-2712.2008.00227.x
16. McGaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. Med Educ. 2014;48(4):375–385. https://doi.org/10.1111/medu.12391
17. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach. 2005;27(1):10–28. https://doi.org/10.1080/01421590500046924
18. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive load theory: implications for medical education: AMEE Guide No. 86. Med Teach. 2014;36(5):371–384. https://doi.org/10.3109/0142159X.2014.889290
19. Lapierre A, Arbour C, Maheu-Cadotte MA, Vinette B, Fontaine G, Lavoie P. Association between clinical simulation design features and novice healthcare professionals' cognitive load: a systematic review and meta-analysis. Simul Gaming. 2022;53(5):538–563. https://doi.org/10.1177/10468781221120599
20. Wijnen-Meijer M, Brandhuber T, Schneider A, Berberat PO. Implementing Kolb's experiential learning cycle by linking real experience, case-based discussion and simulation. J Med Educ Curric Dev. 2022;9:23821205221091511. https://doi.org/10.1177/23821205221091511
21. Frank JR, Mungroo R, Ahmad Y, Wang M, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. Med Teach. 2010;32(8):631–637. https://doi.org/10.3109/0142159X.2010.500898
22. Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-based training. Acad Med. 2014;89(3):387–392. https://doi.org/10.1097/ACM.0000000000000130
23. Yu JH, Chang HJ, Kim SS, Park JE, Chung WY, Lee SK, Kim M, Lee JH, Jung YJ. Effects of high-fidelity simulation education on medical students' anxiety and confidence. PLoS ONE. 2021;16(5):e0251078. https://doi.org/10.1371/journal.pone.0251078
24. Barbadoro P, Brunzini A, Dolcini J, et al. Stress responses in high-fidelity simulation and standard simulation training among medical students. BMC Med Educ. 2023;23:116. https://doi.org/10.1186/s12909-023-04101-x
25. Stecz P, Makara-Studzińska M, Białka S, Misiołek H. Stress responses in high-fidelity simulation among anesthesiology students. Sci Rep. 2021;11(1):17073. https://doi.org/10.1038/s41598-021-96279-7

26. Baethge C, Goldbeck-Wood S, Mertens S. SANRA: a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019;4:5. https://doi.org/10.1186/s41073-019-0064-8
27. ISO 14971:2019. Medical devices — application of risk management to medical devices. https://www.iso.org/standard/72704.html
28. FDA. 510(k) Premarket Notification K110429 (ProSim 4/6/8 patient simulator). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K110429
29. Fung CY, Wraith C, Houghton N, Woods B, Brown C, Dharmarajah A, Sam AH. Effectiveness of a centrally designed, large-scale, multisite undergraduate medical simulation programme: student-reported confidence across eight London hospitals in a UK medical school setting. BMJ Open. 2025;15(9):e094222. https://doi.org/10.1136/bmjopen-2024-094222
30. Nursing simulator procurement notice, GBP 47,131 (Laerdal-class mannequins), unit composition unspecified. Find a Tender Service, Notice 008000-2026, UK. https://www.find-tender.service.gov.uk/Notice/008000-2026
31. IEC 60601-1:2005+AMD1:2012+AMD2:2020 CSV. Medical electrical equipment — Part 1: general requirements for basic safety and essential performance. International Electrotechnical Commission. https://webstore.iec.ch/en/publication/67497
32. Oh SY, Cook DA, Van Gerven PWM, Nicholson J, Fairbrother H, Smeenk FWJM, Pusic MV. Physician training for electrocardiogram interpretation: a systematic review and meta-analysis. Acad Med. 2022;97(4):593–602. https://doi.org/10.1097/ACM.0000000000004607
33. Cook DA, Oh SY, Pusic MV. Assessments of physicians' electrocardiogram interpretation skill: a systematic review. Acad Med. 2022;97(4):603–615. https://doi.org/10.1097/ACM.0000000000004140
34. Viljoen CA, Scott Millar R, Engel ME, Shelton M, Burch V. Is computer-assisted instruction more effective than other educational methods in achieving ECG competence amongst medical students and residents? A systematic review and meta-analysis. BMJ Open. 2019;9(11):e028800. https://doi.org/10.1136/bmjopen-2018-028800
35. Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Sci Data. 2020;7:154. https://doi.org/10.1038/s41597-020-0495-6
36. Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng Med Biol Mag. 2001;20(3):45–50. https://doi.org/10.1109/51.932724
37. Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215
38. Yoo H, Yum Y, Park SW, Lee JM, Jang M, Kim Y, Kim J-H, Park H-J, Han KS, Park JH, Joo HJ. Standardized database of 12-lead electrocardiograms with a common standard for the promotion of cardiovascular research: KURIAS-ECG. Healthc Inform Res. 2023;29(2):132–144. https://doi.org/10.4258/hir.2023.29.2.132

Figures

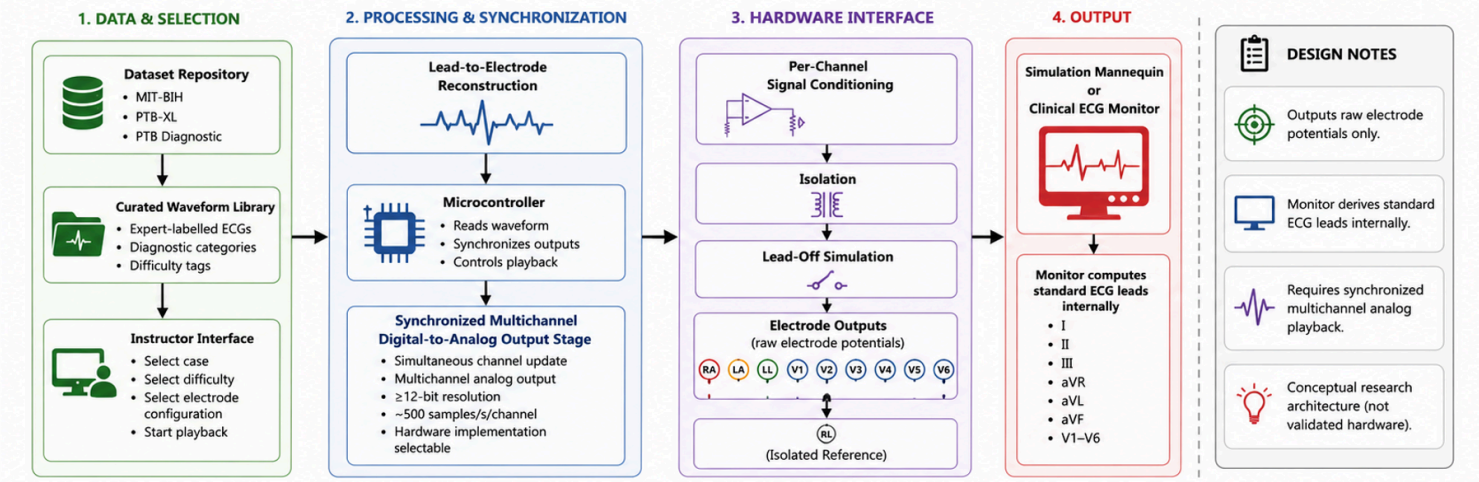


Figure 1

Proposed system architecture. The module is organized into four stages: dataset selection, processing and synchronization, per-channel hardware interface, and output to an unmodified monitor or mannequin. The central design principle is that the device outputs raw electrode potentials only (RA, LA, LL, V1–V6); it does not transmit lead views. A clinical 12-lead ECG is not 12 independent physical outputs; the receiving monitor computes the Wilson central terminal from the limb electrodes and derives leads I, II, III, aVR, aVL, aVF and V1–V6 internally. Reproducing this therefore requires a synchronized multichannel digital-to-analog output stage with simultaneous channel update, rather than a single-channel converter feeding a generic multi-lead connector; a single-channel device supports limited-lead proof-of-concept playback only (Section 6.1). Stored 12-lead records are first reconstructed to absolute electrode potentials, and identical per-channel conditioning, isolation and lead-off simulation are applied to every channel, with the right-leg connection provided as an isolated reference rather than a reconstructed sensing channel. The two numeric targets shown approximately 500 samples per second per channel and at least 12-bit per-channel resolution are the candidate minimums set out in Section 6.1 and are proposed for bench evaluation in Phase 1 (Section 11); they are not validated requirements. The instructor interface selects cases and difficulty but does not score learner responses or enforce a mastery standard, which remain curricular or software-layer functions external to the device (Section 3.1). KURIAS-ECG is omitted from the dataset-repository box because it is not currently available for open deployment (Section 5.1). This figure depicts a proposed design, not a built or validated system.

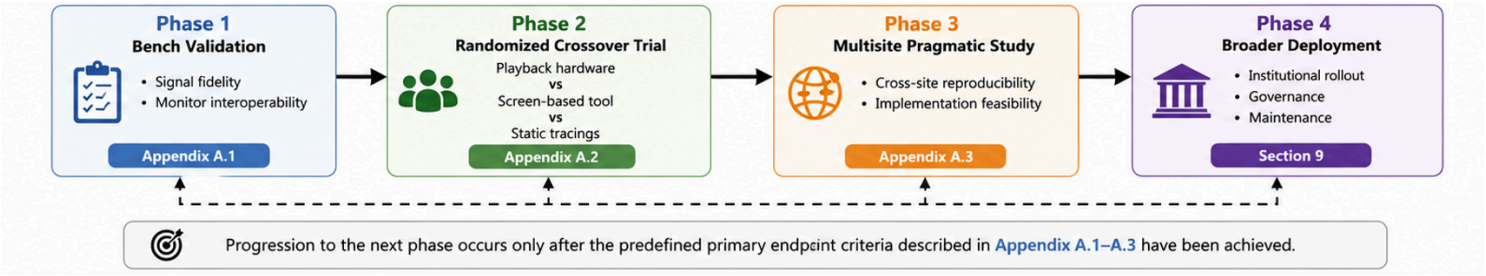


Figure 2

Proposed development and validation timeline. Phases 1–3 constitute the three-phase validation agenda detailed in Section 11; Phase 4 (broader deployment) is not part of that agenda and should not begin until Phase 1–3 endpoints are met, consistent with Section 9's governance requirements.

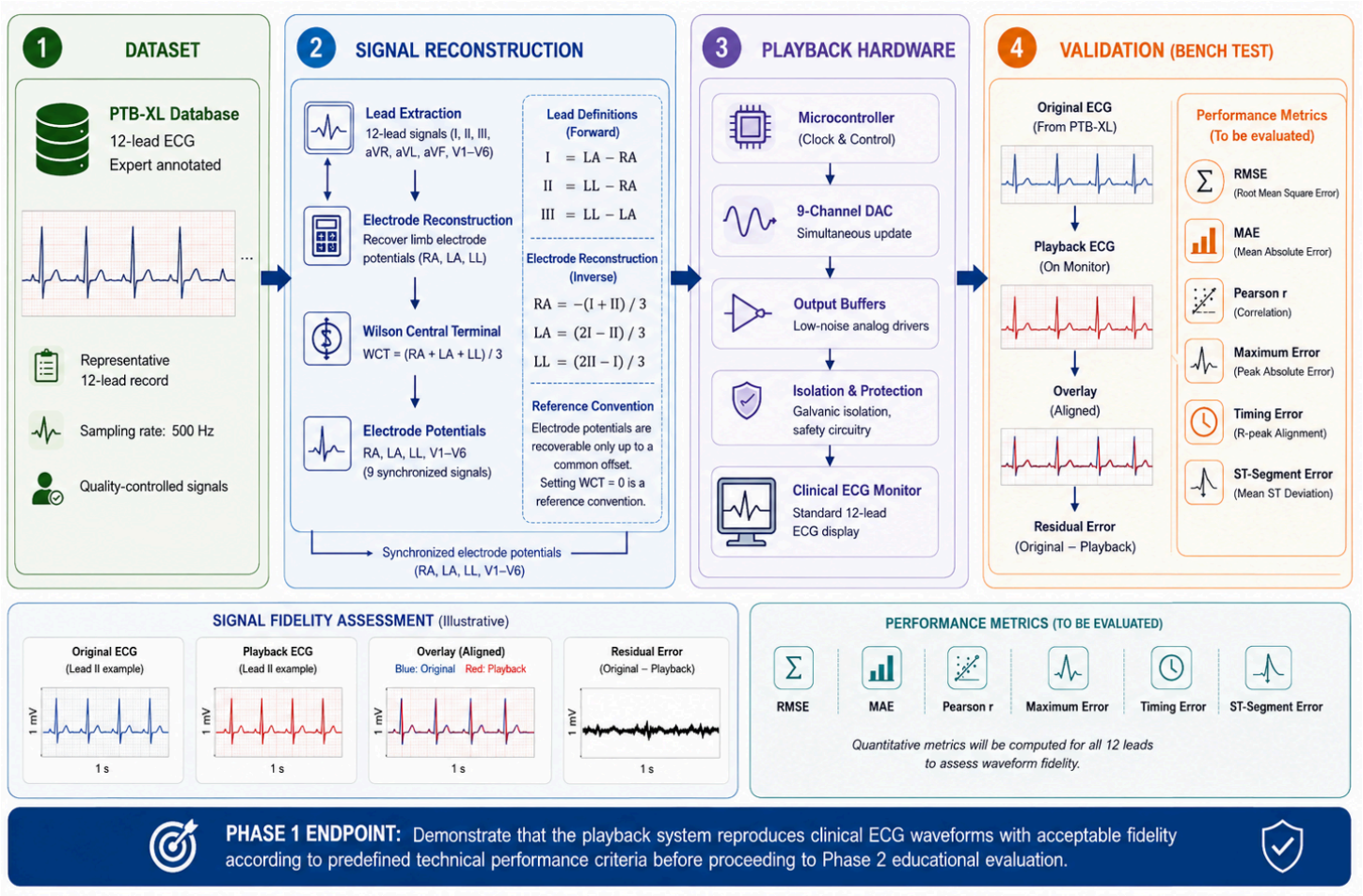


Figure 3

Proposed Phase 1 bench-validation workflow. A representative PTB-XL 12-lead record is decomposed into absolute electrode potentials, played back as nine synchronized analog channels, and compared with the source waveform on an unmodified clinical monitor. The forward lead definitions (I = LA − RA; II = LL − RA; III = LL − LA) and their inverse (RA = −(I + II)/3; LA = (2I − II)/3; LL = (2II − I)/3) are shown together because

the module performs the inverse transform before playback while the receiving monitor performs the forward transform on receipt. Electrode potentials are recoverable only up to a common offset; setting the Wilson central terminal to zero is a reference convention, not a measurement. Performance metrics are those specified in Appendix A.1, with root mean square voltage error as the primary endpoint and timing and ST-segment deviation error as secondary endpoints. Waveforms shown are illustrative. This figure specifies the proposed validation methodology and does not present experimental results; no bench testing has been performed.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixA.docx