Temporal signatures, not thresholds: The case for dynamic digital biomarkers

Madhur Mangalam[1][0000−0001−6369−0414], Yudai Fujimoto[3][0009−0005−2148−0255], and Ken Kiyono[3][0000−0001−5433−7002]

[1] Department of Biomechanics,
University of Nebraska at Omaha, Omaha, NE 68182, USA
mmangalam@unomaha.edu; tdeligiannis@unomaha.edu
[2] Graduate School of Engineering Science,
University of Osaka, Toyonaka, Osaka 560-8531, Japan
fujimoto@bpe.es.osaka-u.ac.jp; kiyono@bpe.es.osaka-u.ac.jp

Abstract. The promise of digital biomarkers lies not in the extraction of static point estimates but in the capture of dynamics—patterns of change that reveal the underlying state of physiological systems. Traditional *wet biomarker* paradigms have relied on thresholds, assuming that a single value can stand in for health or disease. Yet, this threshold-based logic overlooks the inherently dynamic and context-dependent nature of biological function. In contrast, digital health technologies generate dense, continuous streams of data that invite us to view biomarkers as temporal structures rather than static quantities. This chapter argues that disease is more faithfully reflected in the temporal organization, variability, and scaling properties of signals than in any isolated measurement. Three insurmountable challenges stand in the way—individualizability, generalizability, and reproducibility. Individualizability: Digital biomarkers must adapt to the unique physiological signatures of each person, rather than imposing uniform thresholds. Generalizability: they must maintain interpretive power across diverse populations, conditions, and environments. Reproducibility: they must deliver stable and consistent indicators of disease risk and progression across time, devices, and analytic frameworks. By reframing digital biomarkers as properties of dynamics rather than static metrics, we foresee a new generation of precision diagnostics—while recognizing that this demands methodological innovation, interdisciplinary collaboration, and—above all—epistemological humility.

Keywords: Digital biomarkers · Wet biomarkers · Threshold-based biomarkers · Dynamic biomarkers · Temporal dynamics · Physiological variability · Scaling properties · Individualizability · Generalizability · Reproducibility · Precision diagnostics · Nonlinear dynamics · Context-dependence · Multiscale analysis · Continuous monitoring.

1 Introduction: The threshold problem

For decades, medicine has trusted *thresholds* to define health and disease. These so-called *wet biomarkers*—derived from blood, urine, or tissue assays—were re-

duced to static point estimates, with clinical interpretation hinging on whether a measured value crossed a predefined cutoff [97, 110]. A blood pressure above 140/90 mmHg, a fasting glucose above 126 mg/dL, or a cholesterol level above 200 mg/dL could serve as decisive indicators of disease [61, 86]. This threshold-based logic offered clarity, simplicity, and broad clinical utility, particularly in an era dominated by laboratory-based, one-off measurements.

But thresholds are blind to *dynamics*. Human physiology is not static but inherently variable, adaptive, and context-dependent. A single value, however precise, cannot capture the temporal processes that sustain or erode health. Moreover, thresholds conceal the richness of individual variation. A blood pressure reading benign for one individual may be pathological for another, depending on age, genetics, lifestyle, or comorbid conditions [48]. By collapsing biological processes into cutoffs, traditional biomarkers risk misrepresenting the dynamical nature of physiological systems.

The rise of digital health technologies disrupts this threshold paradigm—wearables, smartphones, and other sensing devices generate dense, continuous streams of data, transforming biomarkers from static numbers into temporal trajectories [9, 28, 51]. Digital biomarkers are not snapshots but *films*, not cutoffs but *dynamics*. The central question is no longer whether a measurement crosses a line, but how its fluctuations, variability, and scaling properties reveal the state of the underlying system.

This chapter develops this argument through heart rate variability (HRV) as a paradigmatic case. HRV has long been studied through point-estimate indices such as SDNN, RMSSD, or LF/HF ratio, often interpreted against population-based cutoffs [91, 101]. Yet HRV is fundamentally a dynamic signal, one whose complexity, fractal properties, and non-Gaussian variability more faithfully index cardiovascular health than any single threshold can [23, 39, 40]. Through HRV and its relationship with other digital biomarkers, we argue that the future of precision diagnostics lies in embracing dynamics rather than point estimates. The age of thresholds is ending; the age of dynamics has begun.

1.1 AI and the promise of dynamic biomarkers

This shift from thresholds to dynamics is not only a matter of rethinking biomarkers but also of reconsidering how artificial intelligence (AI) can be used to study them. Digital biomarkers generate dense time series rather than isolated numbers, and interpreting these streams requires methods that can make sense of high-dimensional, continuous, and context-dependent data. Traditional statistical approaches, designed around averages and cutoffs, often struggle to capture such complexity. By contrast, AI systems have been developed precisely to detect structure in sequences and to learn patterns across scales of time and context [34, 81, 108]. Techniques ranging from recurrent neural networks to newer transformer-based architectures are well-suited to modeling temporal dependencies, interactions, and multiscale variability.

In this sense, digital biomarkers provide AI with a fertile domain, and AI in turn provides the computational infrastructure needed to uncover meaning

in dynamic physiological signals. Yet the same challenges that define dynamic biomarkers for human interpretation—individualizability, generalizability, and reproducibility—also shape the feasibility of AI-driven approaches. An AI model that performs well in one dataset but fails to transfer across individuals, devices, or contexts is no better than a flawed threshold. Acknowledging these parallel challenges at the outset situates this chapter squarely at the intersection of AI and health sciences, where advances in neuroscience, physiology, computation, and clinical practice must converge.

2 From wet to digital biomarkers

The distinction between wet and digital biomarkers is not merely technical; it is historical, infrastructural, and epistemological. Wet biomarkers emerged from a clinical–laboratory ecosystem optimized for intermittent, invasive sampling (blood, urine, tissue) and for decision rules that must be simple, auditable, and actionable. Digital biomarkers emerge from a distinct ecosystem: Continuous sensing, inexpensive storage, ubiquitous computation, and the potential to model physiology as a process rather than a point in time. Understanding why the former did *not* pursue dynamics is essential to motivating the latter.

2.1 Why the wet era did not pursue dynamics

Measurement constraints. Wet assays consume the sample and are invasive, time-limited, and expensive. In practice, this enforces sparse sampling (clinic visits, fasting labs) and precludes the dense, longitudinal streams necessary to characterize dynamics. When only a few observations are feasible, cross-sectional interpretation and population thresholds are the only tractable options.

Clinical culture and workflows. Clinical decisions must often be made under time pressure with clear triage. Thresholds furnish deterministic rules (e.g., cholesterol >200 mg/dL), facilitating communication, guideline writing, and medico-legal defensibility. Dynamic descriptors (e.g., scaling exponents, entropy) did not fit the *"if above/below, then do X"* logic embedded in clinical pathways.

Statistical and computational methods. Twentieth-century biostatistics prioritized cross-sectional comparisons, normal ranges, and hypothesis tests applied to point estimates. Although time-series methods were available, they remained peripheral to the development of clinical biomarkers. Limited computational resources further restricted nonlinear and multiscale analyses in everyday practice.

Infrastructure and reimbursement. Laboratory information systems, electronic health records, and billing codes were built around discrete results and threshold-based alerts. Regulatory frameworks rewarded prespecified cutoffs and hard endpoints. Reimbursement favored reportable single values over computationally intensive summaries of dynamics.

Reproducibility by averaging. To ensure analytical reproducibility across labs and platforms, wet biomarkers were engineered for stability, with averaging, calibration, and quality control rules suppressing high-frequency fluctuations as "noise." This design goal conflicts with capturing meaningful variability as a signal.

2.2 The transitional compromise: Holter and other ambulatory measures

Holter electrocardiography and other ambulatory monitors (e.g., 24h blood pressure monitoring) appear to break the constraint of sparsity by recording continuously outside the lab. Historically, however, they preserved the threshold paradigm in three ways:

Event counting under fixed rules. Holter workflows emphasized the detection and counting of discrete events (PVCs, pauses, episodes of tachycardia/bradycardia, ST deviations) relative to fixed criteria. The analytic backbone remained *"how many events exceed a threshold in 24 hours?"*—a dynamic acquisition collapsed back into a static summary. In effect, Holter turned streams into spreadsheets.

Compression into point estimates. For heart rhythm, Holter-derived summaries (e.g., 24h mean heart rate, SDNN, RMSSD, LF/HF ratio) were typically interpreted against population norms or percentiles, not modeled as evolving structures across scales and contexts. For blood pressure, 24h means and "dipping" categories substituted for explicit temporal modeling of variability and coupling to behavior, posture, or sleep.

Practical constraints of the era. Storage, compute, and physician time limited what could be extracted. Nonstationarity, artifacts, posture changes, and circadian structure were acknowledged but handled via filtering and averaging rather than formal dynamic modeling. Reimbursement and report templates further nudged results toward a small set of thresholds and counts. In short, Holter monitoring expanded *data*, but workflows reimposed *thresholds*.

2.3 State vs. process: The epistemic fork

Wet-era biomarkers operationalize *state*: a single measurement situated on a reference scale of normality or pathology, abstracted from its temporal context. Digital biomarkers, in contrast, foreground *process*: The regulation, adaptation, and breakdown of physiology as they unfold across time and circumstance. This distinction is not merely a matter of higher resolution or denser sampling. It represents a fundamental shift in representation and, therefore, in the kinds of inferences that can be drawn. Thresholds answer sufficiency questions (*is it above or below the line?*); dynamics answer questions of organization and integrity (*how does it fluctuate, coordinate, and scale across systems and timescales?*).

Thresholds answer sufficiency questions (*is it above or below the line?*); dynamics answer questions of organization and integrity (*how does it fluctuate, coordinate, and scale across systems and timescales?*). **Table 1** summarizes the key dimensions along which wet and digital biomarker paradigms diverge.

Table 1. Fundamental differences between wet and digital biomarker paradigms.

Dimension	**Wet biomarkers**	**Digital biomarkers**
Sampling	Sparse, invasive, intermittent	Dense, continuous, non-invasive
Representation	State (point estimate)	Process (temporal trajectory)
Interpretation	Threshold-based (above/below cutoff)	Dynamics-based (variability, scaling, coupling)
Variability	Suppressed as noise	Signal of adaptive capacity
Context	Controlled (fasting, supine)	Embedded (free-living, multimodal)
Analytics	Cross-sectional statistics	Time-series, nonlinear, multiscale methods
Clinical logic	If-then rules (deterministic)	Pattern recognition (probabilistic)

2.4 What digital biomarkers add (and demand)

Digital health technologies now routinely yield dense, longitudinal streams of data, including inter-beat intervals, gait cycles, sleep–wake microstructure, activity bursts, and contextual signals (such as light, posture, and temperature). These enable:

- **Temporal organization:** analysis of rhythms, drift, and recovery, including nonstationarity and regime shifts.
- **Variability-as-information:** distinguishing structured variability (adaptive complexity) from deregulated randomness.
- **Multiscale structure:** scaling laws, long-memory processes, and cross-scale interactions that cannot be seen in snapshots.
- **Relational meaning:** coupling between signals (e.g., heart–gait, heart–sleep) as indicators of system integration or fracture.

These gains, however, impose obligations:

Methods fit for streams, not snapshots. Time-series modeling, state-space and switching models, fractal and multifractal analysis, entropy/complexity measures, recurrence, and network approaches are required to respect temporal dependence and cross-signal interactions.

Inference within persons, then across persons. Dynamic biomarkers must separate *idiographic* baselines and trajectories from *nomothetic* regularities. Hierarchical modeling and adaptive priors are needed so that population information stabilizes—but does not overwrite—individual structure.

Context modeling. Posture, sleep stage, medication timing, stress, and environment shape dynamics. Context must be measured (or inferred) and explicitly modeled to avoid conflating physiology with circumstance.

Device and pipeline heterogeneity. Sampling rates, filtering, beat detection, and artifact handling alter the dynamics estimates. Transparent pipelines, signal quality indices, and cross-device calibration are prerequisites for reproducibility in the dynamic regime.

2.5 Revisiting the canonical triad under dynamics

The shift from thresholds to dynamics redefines how we should think about the three canonical challenges of biomarkers: individualizability, generalizability, and reproducibility. Each acquires new meaning when physiology is treated as a process rather than a point.

Individualizability. Static thresholds erase inter-individual differences, forcing diverse physiologies into a single cutoff. Dynamic models, by contrast, can accommodate individuality by estimating person-specific baselines, reactivity, and recovery profiles. In this framework, N-of-1 trajectories are elevated from statistical nuisances to primary objects of inference.

Generalizability. Models must retain validity across devices, demographics, comorbidities, and contexts. This argues for mechanism-aware features (e.g., scaling exponents, coupling measures) that travel better than raw cutoffs, and for explicit domain shift evaluation.

Reproducibility. In dynamics, reproducibility does not mean the constancy of a single number; it means the stability of *relationships* (e.g., scaling slopes, coupling strengths) under controlled changes, and the predictable change under biological perturbations. Pre-registration of time-series pipelines, reporting of parameter choices, and sensitivity analyses become essential.

2.6 From thresholds to dynamics: The logical argument

1. Wet biomarkers were shaped by constraints that made dynamics unobservable at scale; thresholds were a rational solution to sparse, invasive sampling.
2. Transitional technologies (e.g., Holter, ambulatory BP) captured continuous data but, for practical and institutional reasons, collapsed streams back into counts and cutoffs.

3. Digital sensing removes the acquisition barrier and makes dynamics observable; therefore, the *representational target* of biomarkers should shift from state to process.
4. Once the process is the target, appropriate analytics must follow (temporal, relational, multiscale), and validation must be reframed around stability of patterns rather than fixed thresholds.
5. This shift directly addresses the triad of individualizability, generalizability, and reproducibility by modeling within-person dynamics, evaluating portability across contexts, and defining reproducibility as invariance of structure under controlled variation.

2.7 Foreshadowing the HRV case

Heart rate variability (HRV) exemplifies this trajectory. Holter-era practice summarized rich inter-beat interval streams with event counts and a handful of point estimates, interpreted via thresholds. In contrast, a dynamic view treats HRV as a structured process whose scaling, non-Gaussianity, and coupling with other digital biomarkers convey cardiovascular integrity in ways no single cutoff can. Section 3 develops this case in detail.

We highlight HRV as a case study because it sits at the interface of neuroscience and physiology: it reflects central autonomic network function, stress reactivity, and sleep–wake dynamics. Its wide use in both clinical and consumer settings, coupled with its rich dynamical structure, makes it a unique candidate to exemplify the broader argument of this chapter. In addition, our own expertise in HRV research enables us to provide an in-depth analysis that can serve as a template for other biomarkers.

3 Heart rate variability: A case in point

HRV—the fluctuation in intervals between successive heartbeats—has become one of the most widely studied digital biomarkers of autonomic nervous system function [101, 105]. Derived from electrocardiogram (ECG) or photoplethysmography (PPG) signals, HRV reflects the dynamic interplay between sympathetic and parasympathetic influences on cardiac rhythm [91]. Because it reflects autonomic responsiveness and adaptability, HRV has been linked to outcomes such as stress reactivity, cardiovascular morbidity and mortality, sleep quality, and resilience to psychiatric disorders [101, 105]. In this sense, HRV epitomizes the promise of digital biomarkers: sensitive to physiological state, non-invasive, and readily measurable with both clinical and consumer-grade devices.

3.1 Traditional use: point estimates

Historically, HRV has been operationalized through point-estimate indices derived from fixed-length recordings [101]. The most common measures include time-domain indices—for example, the standard deviation of normal-to-normal

intervals (SDNN, a time-domain measure of overall variability) and the root mean square of successive differences (RMSSD, a time-domain measure of short-term variability)—as well as frequency-domain indices such as low-frequency (LF) and high-frequency (HF) power, and their ratio (LF/HF, a frequency-domain marker of sympathovagal balance) [92]. Nonlinear complexity indices—such as sample entropy, approximate entropy, and multiscale entropy—though less widely adopted in clinical practice, have also been proposed to capture aspects of signal regularity and complexity [29, 30, 85]. However, despite their theoretical advantages, these measures have remained less widely adopted in clinical practice and, critically, have typically been operationalized in the same threshold-based manner: computed as point estimates from fixed-length recordings and compared against population-based reference values. These measures are typically averaged over standardized intervals, most often 5-min short-term recordings in laboratory settings or 24-h Holter recordings in ambulatory monitoring, and in some protocols, up to 72-h recording [101]. The resulting averages are then compared against population-based reference values to classify autonomic function. Clinical practice has sometimes attempted to formalize cutoffs: for instance, SDNN below 50 ms has been associated with high mortality risk in cardiac patients, values between 50–100 ms with moderate risk, and values above 100 ms with normal or healthy function [59, 60]. Similar thresholds, including those for RMSSD and LF/HF ratio, have been proposed in both research and clinical guidelines to stratify health and disease [101, 60]. These population-referenced summaries capture coarse autonomic tone but, by construction, compress the temporal structure into static surrogates.

3.2 The threshold problem

Despite their appeal, such thresholds face the same limitations as other cutoff-based biomarkers. Physiological individuality ensures that a value deemed "abnormal" in one person may be entirely normal in another. For example, an SDNN of 50 ms may place a sedentary 70-year-old with diabetes into a high-risk category [55, 109], yet the same value may be well within the expected range for a 30-year-old endurance athlete whose autonomic system adapts differently to sustained cardiovascular training [8]. Attempts to impose universal cutoffs therefore risk collapsing diverse physiologies into a single scale, obscuring clinically meaningful differences and generating both false positives and false negatives. In effect, thresholds sacrifice biological specificity for simplicity, and in doing so, can misrepresent the very variability they are meant to capture.

3.3 Limitations of static indices

The limitations of static HRV indices are well documented. First, they are strongly influenced by demographic factors such as age and sex: HRV generally declines with age and often differs between males and females, complicating attempts at universal reference ranges [109, 117]. Second, HRV exhibits circadian rhythms, with higher values during sleep and lower values during waking

hours [19, 47]. A cutoff that appears valid in one context may misclassify physiology in another. Third, device heterogeneity (sampling rate, detection, artifact handling) introduces additional variability: indices computed from short-term ECG, long-term Holter monitoring, or wrist-worn PPG may differ in magnitude and reliability, complicating cross-study comparisons [14, 77]. Finally, point estimates collapse temporal richness into a single number, discarding fluctuations, scaling properties, and nonlinear dynamics that better reflect system integrity [39, 40].

3.4 From static indices to dynamics

If traditional HRV metrics highlight the limitations of point estimates, newer approaches illustrate what can be gained by treating HRV as a dynamic process. Rather than collapsing variability into a single number, these methods examine the temporal structure of inter-beat intervals across multiple scales. Entropy-based measures, such as approximate entropy, sample entropy, and multiscale entropy, capture the irregularity and complexity of heart rhythms, distinguishing structured variability from random noise [29, 30, 85]. Detrended fluctuation analysis (DFA) assesses fractal scaling in heartbeat fluctuations, providing insights into long-range correlations and the integrity of physiological control systems [46, 78]. Multifractal analyses extend this logic further, revealing how fluctuations of different magnitudes may follow distinct scaling laws and how pathological states can erode this multiscale richness [4, 49].

These dynamics-based approaches demonstrate that HRV is not simply "more" or "less" variability but a structured signal whose temporal organization carries meaning [39, 40]. For instance, reduced entropy and diminished fractal scaling have been linked to heightened mortality risk, even when conventional indices fall within normal ranges [112]. Conversely, preserved complexity is associated with resilience and adaptive capacity [39, 40]. Such findings strongly suggest that disease is not best understood as a departure from a universal threshold, but rather as a breakdown in the underlying dynamics of variability itself. **Table 2** contrasts the traditional threshold-based approach to HRV with the emerging dynamics-based paradigm across key analytical and interpretive dimensions. In this way, HRV vividly exemplifies the broader argument of this chapter: the future of digital biomarkers lies not in static cutoffs but in the rich patterns of change that reveal the deeper organizing principles of living systems and their capacity for adaptation.

4 Physiological coupling and multibiomarker integration as an AI challenge

If HRV illustrates the promise and pitfalls of treating a single dynamic signal as a biomarker, the next frontier lies in examining how multiple dynamic signals interact with each other. Physiological systems rarely operate in isolation: cardiac

Table 2. Contrasting traditional and dynamic approaches to heart rate variability analysis.

Aspect	**Traditional approach**	**Dynamic approach**
Measures	SDNN, RMSSD, LF/HF ratio (point estimates)	Multiscale entropy, DFA scaling exponents, multifractal spectra
Interpretation	Compare to population thresholds (e.g., SDNN $<$ 50 ms = high risk)	Assess temporal organization, complexity, and scaling properties
Variability	Single number (higher = better)	Structured signal revealing regulatory integrity
Time scale	Single (5-min or 24-h average)	Multiple (short-term to long-range correlations)
Context	Ignored or controlled	Explicitly modeled (posture, sleep stage, circadian phase)
Diagnostic target	Autonomic tone	Autonomic adaptability and system integrity

dynamics interact with locomotor activity, sleep influences autonomic tone, neural oscillations couple with motor performance, and immune processes feed back into behavior and affect [12, 13, 50]. The clinical meaning of a single biomarker often emerges only when interpreted in relation to others [11]. For instance, HRV during deep sleep carries different implications than HRV during upright posture [26, 95]; gait variability is benign in youth but predictive of falls in older adults; and joint consideration of HRV and EEG microstates during sleep may reveal signatures of stress recovery that neither signal alone could capture [53]. Likewise, cardiolocomotor coupling—the synchronization of heart rhythms with gait cycles—has been shown to index cardiovascular fitness, adaptability, and resilience, illustrating how coupling across systems can be more diagnostic than either system in isolation [72, 73].

AI is particularly well positioned to address this challenge, as contemporary machine learning methods are expressly designed for multimodal integration and can uncover latent structure across diverse physiological signals that traditional approaches struggle to reconcile [10]. Deep learning architectures can combine heterogeneous time series—cardiac, motor, neural, behavioral, and metabolic—and discover coordinated patterns that would be invisible if each signal were analyzed in isolation. Methods such as multimodal transformers, graph neural networks, and cross-attention can discover structured relationships across heterogeneous inputs [57, 116]. For example, a multimodal model might learn reductions in HRV complexity signal pathology only when they coincide with increased gait asymmetry, disrupted sleep microstructure, or contextual markers of dysregulation, whereas the same reductions in isolation could reflect transient or adaptive variation. This kind of relational inference could move biomarker

science beyond threshold-based classification into a space where patterns of integration and disintegration become the diagnostic target [12, 13, 50].

But this opportunity is also a challenge. The complexity of modeling the "network of networks" increases dramatically when dynamic biomarkers are considered in concert. A model that captures HRV well may fail when asked to integrate HRV with sleep staging, locomotor activity, and cognitive performance, unless it is explicitly designed for relational inference across modalities [10]. Multimodal models risk overfitting to spurious correlations if the training data are small, biased, or contextually narrow [24]. They also raise issues of interpretability: when a model predicts deterioration based on a combination of HRV, gait, and sleep features, it is not always clear which relationships drove the prediction or whether those relationships correspond to known physiology [45, 88]. Without careful design, multimodal integration models risk becoming accurate yet clinically uninterpretable black boxes.

The critical AI challenge, therefore, is not only to model dynamics within a single time series, but to capture the structure of interactions across biomarkers and across systems. Doing so will require architectures that treat multimodal signals as coupled processes rather than parallel inputs, as well as evaluation frameworks that assess not only accuracy within one modality but coherence across many. It will also require explicit tests of generalizability and reproducibility across contexts, devices, and populations, since coupling patterns that hold in one setting may fail in another. Dynamic biomarkers thus demand an AI of integration, one capable of moving from isolated measures to the networks that truly define health and disease. In practice, this means that the most informative biomarkers of the future may not be the dynamics of HRV, gait, or sleep alone, but the stability or breakdown of their couplings across scales and contexts.

5 Three insurmountable challenges

Reframing digital biomarkers as dynamics rather than thresholds opens new possibilities for precision diagnostics, but it also exposes deep challenges that cannot be solved by technology alone. These challenges—individualizability, generalizability, and reproducibility—mark the fault lines between the promise of digital biomarkers and the practical demands of clinical translation. Labeled "insurmountable," these challenges do not signal impossibility of progress but mark structural tensions that must be continually negotiated rather than permanently resolved. Left unaddressed, they threaten to turn digital biomarkers into a new source of noise, bias, and inequity rather than a foundation for precision medicine.

5.1 Interpretability as a constraint on AI-driven biomarkers

The use of AI to analyze dynamic biomarkers raises a fundamental question: not only whether models can predict outcomes, but whether their predictions can be interpreted in biologically meaningful ways [45, 88]. Dynamic biomarkers already

challenge human intuition by foregrounding temporal variability and nonlinear structure. If AI adds an additional layer of opacity, we risk substituting one kind of black box (thresholds without context) for another (predictions without explanation) [2, 7].

Explainable AI (XAI) techniques—such as saliency maps [93], feature attribution [66, 84], and attention weights [111]—offer potential insights into what models use to make predictions. Yet their reliability in biomedical contexts remains uncertain [37, 56]. An attribution heatmap may highlight features of a signal, but whether those features correspond to real physiological processes or are artifacts of model training is often unclear [3, 106]. The attention mechanisms widely used in modern AI may not even represent true explanations [52, 114].

Interpretability thus becomes a dual challenge: making sense of dynamic biomarkers themselves and ensuring that AI models respect rather than obscure their physiological basis. In the context of human physiology and neuroscience, where mechanisms are as important as outcomes, this issue is particularly acute. The risk is that we may achieve predictive accuracy while losing the mechanistic insights that make biomarkers scientifically valuable.

5.2 Individualizability

Each person's physiology is unique. Heart rate variability (HRV), gait dynamics, and sleep rhythms are shaped by genetics [94], developmental history [104], environment [35], lifestyle [33], and comorbidities [43]. Even within the same individual, dynamics vary across circadian cycles [19, 47], activity levels [54], and psychological states [6, 102]. Static thresholds collapse this heterogeneity into a single line of demarcation, ignoring meaningful differences in baseline and adaptability. For example, an SDNN of 50 ms may indicate heightened cardiac risk in a sedentary 70-year-old with diabetes [55, 109] but be entirely normal for a 30-year-old endurance athlete [8]. Similarly, stride-to-stride gait variability that predicts falls in frail older adults [41, 42] can instead reflect adaptive flexibility in younger adults [67, 115]. Thresholds misclassify both groups.

Dynamic biomarkers can instead model N-of-1 baselines [63], reactivity to stress [90], and recovery trajectories [79]. A person's *pattern of change* becomes the diagnostic signal [118], not a comparison to a population cutoff. But here lies the central threat: without careful design and methodological discipline, such individualization risks collapsing into idiosyncratic models that cannot be compared or communicated. What begins as precision may quickly turn into fragmentation, with diagnostics tailored so narrowly that they lose broader clinical transferability. The enduring challenge is to honor individuality without sacrificing interpretability or shared medical consensus.

5.3 Generalizability

If individualizability is the promise of dynamic biomarkers, generalizability is their counterweight. A model that works beautifully in one population may fail in another. HRV entropy measures, for instance, can distinguish stress in a lab

setting but are highly sensitive to posture (supine vs. upright) [26, 95], leading to false alarms in free-living conditions [92]. Sleep staging algorithms trained on healthy young adults often misclassify stages in older adults or in patients with sleep apnea [18]. Even device placement changes the story: gait variability measured by a wrist-worn accelerometer may differ markedly from that measured by a hip-worn sensor [68, 96, 119, 100], threatening comparability across studies.

The risk is clear and pressing: without generalizability, dynamic biomarkers inevitably deepen existing health disparities. Statistical models validated in privileged, homogeneous samples may systematically misclassify risk in underrepresented groups, leading to false reassurance for some and unnecessary interventions for others [22, 27, 74]. Mechanism-aware features (e.g., scaling exponents, coupling indices) can often travel better across contexts than raw cutoffs [30], but even these require rigorous and systematic testing under domain shift [17, 75]. Unless explicitly acknowledged and addressed, generalizability threatens to turn digital biomarkers into yet another form of algorithmic bias [25, 82], steadily undermining trust among patients, clinicians, and regulators.

5.4 Reproducibility

The third challenge is reproducibility. In the threshold era, reproducibility meant that the same blood sample yielded the same cholesterol level in different labs. In the dynamic era, reproducibility must be redefined as the stability of *relationships*: scaling slopes, coupling strengths, synchrony patterns, or temporal correlations. These features are not fixed constants but should exhibit predictable modulation under perturbation. For instance, multiscale entropy of HRV reliably decreases during anesthesia and increases during recovery [62, 87, 98], but the absolute values vary widely depending on artifact handling, segment length, or software implementation. Likewise, gait variability can appear "abnormal" or "healthy" depending on the algorithm used to detect step boundaries—even from the same raw signal [68, 96, 119, 100]. **Table 3** catalogs the major sources of variability that threaten reproducibility in dynamic biomarker research, distinguishing between technical factors (device, preprocessing, analysis parameters) and physiological context.

The threat here is credibility. If results cannot be reproduced across analytic pipelines, devices, or labs, dynamic biomarkers risk being dismissed as fragile artifacts rather than robust clinical signals. The very complexity that gives them promise becomes a liability, fueling skepticism among clinicians, statisticians, and regulators. Avoiding this fate requires transparent, pre-registered pipelines; reporting of parameter choices; systematic sensitivity analyses; and calibration standards that allow results to be compared across platforms. Without such rigor, reproducibility becomes the Achilles' heel of dynamic biomarker science.

5.5 A structural triad

These challenges delineate a structural triad that maps the fundamental terrain of dynamic digital biomarkers. Individualizability emphasizes the uniqueness of

Table 3. Major sources of variability affecting reproducibility of dynamic biomarkers.

Source	**Example**	**Impact on dynamic measures**
Device heterogeneity	Sampling rate (ECG: 250 Hz vs. 1000 Hz; PPG: wrist vs. chest)	Alters detection of high-frequency variability and scaling exponents
Preprocessing	Beat detection algorithms, artifact removal thresholds	Changes time-series structure, affects entropy and DFA estimates
Segmentation	Window length (5 min vs. 24 h), overlap	Influences multiscale measures and long-range correlations
Parameter choices	Embedding dimension (m), tolerance (r) for entropy; scale range for DFA	Directly determines numerical values of complexity indices
Context variation	Posture, circadian phase, sleep stage, activity level	Modulates physiological state, not device or algorithm

each person and resists the flattening effect of universal thresholds. Generalizability demands that models retain validity across diverse populations, devices, and contexts rather than collapsing under domain shift. Reproducibility ensures that dynamic relationships—scaling slopes, coupling strengths, synchrony patterns—remain stable across time and analytic frameworks.

Each challenge guards against a distinct threat: the loss of interpretability that arises when individual variation is collapsed into a single cutoff; the inequitable misclassification that follows when models trained on one population are applied uncritically to others; and the collapse of credibility that occurs when analytic pipelines, devices, or laboratories cannot reproduce the same dynamic signatures. **Table 4** summarizes the structural triad of challenges, their associated threats, and potential pathways toward solutions. These threats are not peripheral inconveniences but fundamental obstacles that determine whether digital biomarkers will serve as trustworthy clinical devices or devolve into sources of confusion, bias, and mistrust. To acknowledge the triad of individualizability, generalizability, and reproducibility as structural rather than incidental is therefore more than an abstract theoretical exercise. It is the necessary starting point for developing a science of dynamic digital biomarkers that is both realistic in its expectations and robust in its clinical, regulatory, and scientific applications.

6 Toward a science of dynamic digital biomarkers

The turn from thresholds to dynamics is not a matter of upgrading resolution; it is a reorientation of the science of biomarkers. To take dynamics seriously means abandoning the assumption that pathology can be captured by a single cutoff and

Table 4. Three core challenges for dynamic digital biomarkers: associated risks and potential solutions.

Challenge	Risk	Solution
Individualizability	Universal thresholds collapse diverse physiologies into single cutoffs, misclassifying individuals with different baselines (e.g., athletes vs. sedentary patients)	Model N-of-1 baselines, reactivity patterns, and recovery trajectories; treat individual's pattern of change as the diagnostic signal
Generalizability	Models validated on narrow populations fail across demographics, devices, and contexts; algorithmic bias deepens health disparities	Use mechanism-aware features (scaling exponents, coupling indices); test under domain shift; validate across diverse populations and devices
Reproducibility	Results vary across analytic pipelines, devices, and labs; dynamic signatures appear fragile, undermining clinical credibility	Pre-register pipelines; report parameter choices; conduct sensitivity analyses; develop calibration standards; define reproducibility as stability of relationships, not fixed values

instead developing diagnostics that recognize patterns, rhythms, and interactions as the primary signals of health and disease. Pattern-based diagnostics shift attention from the static to the evolving, from the isolated to the relational, and from the singular to the multiscale.

A key implication is that biomarkers cannot be understood in isolation. Just as physiological systems are interdependent, so too must biomarkers be interpreted through their relationships. A single signal, such as heart rate variability, is informative, but its real meaning often emerges when studied in concert with gait dynamics, sleep patterns, or activity rhythms. Network-based approaches capture these interdependencies, treating the body as a system of coupled processes rather than a collection of independent parts [12, 13, 50]. In this perspective, the biomarker is not a number but a node in a network, and its significance resides in the structure of connections.

Developing such a science demands methodological pluralism. Nonlinear dynamics provide methods to detect critical transitions, scaling properties, and fractal organization. Machine learning offers flexible strategies to identify patterns in dense, high-dimensional data streams, but must be disciplined by interpretability and physiological plausibility [16, 31]. Multiscale analysis ensures that processes unfolding over milliseconds, hours, and years can be integrated into a coherent framework rather than collapsed into a single timescale [38, 70]. Together, these approaches reconfigure biomarker science from one of static classification to one of dynamic characterization.

Yet methodological sophistication alone is insufficient; a science of dynamic biomarkers must also be ethically and epistemologically grounded. The abundance of data and the power of algorithms should not be mistaken for certainty. The field must embrace complexity without fetishizing it, acknowledge uncertainty without paralyzing action, and design diagnostics that are both physiologically faithful and clinically usable [9, 28]. Above all, it must resist the temptation to reimpose the false clarity of cutoffs under new guises. The challenge is to build a science that honors the dynamics of living systems without sacrificing the pragmatics of medical decision-making.

6.1 Data provenance and context embedding

Dynamic biomarkers are deeply context-dependent [12, 13, 50]. For instance, heart rate variability, gait dynamics, and sleep rhythms shift not only with pathology but also with posture [26, 95], circadian phase [19, 47], medication timing [71, 103], stress [44, 89], and environmental conditions. A value recorded in the morning may carry a different meaning than the same value at night [47]; a signal from a wrist-worn sensor may not be directly comparable to one from a chest strap [14, 96, 100]. Sleep stage further modulates these relationships, with HRV patterns during REM sleep differing markedly from those during deep sleep [107, 53]. Without explicit accounting for context, digital biomarkers risk conflating ordinary variability with pathological change.

AI systems trained on these data will only ever be as robust as the metadata they are given [15, 32]. Building models that explicitly encode context—through multimodal inputs, contextual covariates, or architectures designed to track environmental states—is therefore a critical requirement [10, 83]. Data provenance systems that track not only the biomarker values but also their collection context become essential for meaningful interpretation [21, 69]. Absent such considerations, we risk mistaking ordinary context effects (e.g., time of day, posture, sensor placement, or sleep stage) for genuine pathology, thereby undermining not only analytic accuracy but also clinical trustworthiness. Context embedding is therefore not a refinement but a foundational requirement if AI is to make meaningful sense of dynamic biomarkers in real-world clinical settings.

6.2 Longitudinal adaptation and drift

Dynamic biomarkers are not static traits but evolving processes. HRV, gait, and sleep dynamics all change over time due to aging [109, 1, 5], training [79, 20, 58], recovery, stress, medication [71, 103], and environmental conditions. An AI model trained on data collected at one moment may silently degrade in accuracy as the underlying physiology shifts. This phenomenon, known as *concept drift*, refers to the gradual or abrupt change in the statistical properties of data over time, such that previously learned models no longer remain valid [36, 65, 113]. It is an especially serious risk for longitudinal biomarkers intended for continuous physiological monitoring [14, 80, 99].

Without robust mechanisms to detect and adapt to drift, AI systems may easily misclassify normal adaptation as pathology or completely overlook the onset of disease. For instance, gradual age-related changes in HRV scaling exponents might be mistaken for progressive cardiovascular decline, while training-induced increases in variability could be falsely flagged as pathological instability. Addressing drift requires methods that can recalibrate baselines, update models incrementally [64, 76], and distinguish between adaptive change and pathological change. In practice, this means that AI for dynamic biomarkers must be engineered not only for pattern recognition but also for *temporal resilience*: the ability to remain valid as the signal itself evolves.

7 Conclusion

Disease is not a number but a pattern unfolding in time. The future of biomarker science depends on recognizing that digital biomarkers must be reframed as dynamic and relational, not static and isolated. Health and disease are systemic properties, best understood through variability, coordination, and adaptation rather than thresholds and cutoffs.

A new generation of diagnostics will therefore require methodological innovation, interdisciplinary collaboration, and epistemological humility. Innovation is needed to develop analytic methods that can interpret dense temporal streams. Collaboration across medicine, engineering, mathematics, and the social sciences is needed to integrate methods and perspectives. Humility is needed to ensure that the quest for clarity does not collapse back into the illusion of fixed cutoffs. By embracing dynamics, we can move toward a more faithful science of biomarkers, one that honors the complexity of living systems while serving the practical aims of precision medicine.

7.1 AI-specific challenges for dynamic biomarkers

Placing dynamic biomarkers in the context of AI also highlights new challenges that go beyond physiology alone. AI methods risk overfitting to individual idiosyncrasies, mistaking transient fluctuations for stable traits, and thereby undermining individualizability. Generalizability is equally precarious: models trained on narrow, homogeneous datasets may fail under domain shift, producing algorithmic bias that mirrors and amplifies health disparities. Reproducibility, already difficult in dynamic physiology, becomes even more fragile in AI when black-box architectures obscure the features that drive predictions.

These concerns point to a central paradox. AI has the capacity to model nonlinear, multiscale, and networked dynamics at a scale far beyond human cognition. But without transparent pipelines, interpretable models, and rigorous standards for validation, AI risks reintroducing the very opacity that dynamic biomarkers are meant to overcome. The task, then, is not only to embrace dynamics in physiology but also to design AI systems that respect those dynamics

while remaining interpretable, portable, and reproducible. In this way, the convergence of AI and dynamic biomarkers offers not a solution but a call for new epistemic standards at the interface of neuroscience, physiology, computation, and clinical practice.

References

1. Abhishekh, H.A., Nisarga, P., Kisan, R., Meghana, A., Chandran, S., Raju, T., Sathyaprabha, T.N.: Influence of age and gender on autonomic regulation of heart. Journal of Clinical Monitoring and Computing **27**, 259–264 (2013), https://doi.org/10.1007/s10877-012-9424-3
2. Adadi, A., Berrada, M.: Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access **6**, 52138–52160 (2018), https://doi.org/10.1109/ACCESS.2018.2870052
3. Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V.I., Consortium, P.: Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making **20**, 310 (2020), https://doi.org/10.1186/s12911-020-01332-6
4. Amaral, L.A.N., Ivanov, P.C., Aoyagi, N., Hidaka, I., Tomono, S., Goldberger, A.L., Stanley, H.E., Yamamoto, Y.: Behavioral-independent features of complex heartbeat dynamics. Physical Review Letters **86**(26), 6026 (2001), https://doi.org/10.1103/PhysRevLett.86.6026
5. Antelmi, I., De Paula, R.S., Shinzato, A.R., Peres, C.A., Mansur, A.J., Grupi, C.J.: Influence of age, gender, body mass index, and functional capacity on heart rate variability in a cohort of subjects without heart disease. American Journal of Cardiology **93**(3), 381–385 (2004), https://doi.org/10.1016/j.amjcard.2003.09.065
6. Appelhans, B.M., Luecken, L.J.: Heart rate variability as an index of regulated emotional responding. Review of General Psychology **10**(3), 229–240 (2006), https://doi.org/10.1037/1089-2680.10.3.229
7. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion **58**, 82–115 (2020), https://doi.org/10.1016/j.inffus.2019.12.012
8. Aubert, A.E., Seps, B., Beckers, F.: Heart rate variability in athletes. Sports Medicine **33**, 889–919 (2003), https://doi.org/10.2165/00007256-200333120-00003
9. Babrak, L.M., Menetski, J., Rebhan, M., Nisato, G., Zinggeler, M., Brasier, N., Baerenfaller, K., Brenzikofer, T., Baltzer, L., Vogler, C., et al.: Traditional and digital biomarkers: Two worlds apart? Digital Biomarkers **3**(2), 92–102 (2019), https://doi.org/10.1159/000502000
10. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(2), 423–443 (2018), https://doi.org/10.1109/TPAMI.2018.2798607
11. Barabási, A.L., Gulbahce, N., Loscalzo, J.: Network medicine: A network-based approach to human disease. Nature Reviews Genetics **12**, 56–68 (2011), https://doi.org/10.1038/nrg2918
12. Bartsch, R.P., Liu, K.K., Bashan, A., Ivanov, P.C.: Network physiology: How organ systems dynamically interact. PloS One **10**(11), e0142143 (2015), https://doi.org/10.1371/journal.pone.0142143

13. Bashan, A., Bartsch, R.P., Kantelhardt, J.W., Havlin, S., Ivanov, P.C.: Network physiology reveals relations between network topology and physiological function. Nature Communications **3**, 702 (2012), https://doi.org/10.1038/ncomms1705
14. Bent, B., Goldstein, B.A., Kibbe, W.A., Dunn, J.P.: Investigating sources of inaccuracy in wearable optical heart rate sensors. npj Digital Medicine **3**, 18 (2020), https://doi.org/10.1038/s41746-020-0226-6
15. Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., Riboni, D.: A survey of context modelling and reasoning techniques. Pervasive and Mobile Computing **6**(2), 161–180 (2010), https://doi.org/10.1016/j.pmcj.2009.06.002
16. Bienefeld, N., Boss, J.M., Lüthy, R., Brodbeck, D., Azzati, J., Blaser, M., Willms, J., Keller, E.: Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. npj Digital Medicine **6**, 94 (2023), https://doi.org/10.1038/s41746-023-00837-4
17. Bird, J.J., Kobylarz, J., Faria, D.R., Ekárt, A., Ribeiro, E.P.: Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG. IEEE Access **8**, 54789–54801 (2020), https://doi.org/10.1109/ACCESS.2020.2979074
18. Birrer, V., Elgendi, M., Lambercy, O., Menon, C.: Evaluating reliability in wearable devices for sleep staging. npj Digital Medicine **7**, 74 (2024), https://doi.org/10.1038/s41746-024-01016-9
19. Bonnemeier, H., Wiegand, U.K., Brandes, A., Kluge, N., Katus, H.A., Richardt, G., Potratz, J.: Circadian profile of cardiac autonomic nervous modulation in healthy subjects: Differing effects of aging and gender on heart rate variability. Journal of Cardiovascular Electrophysiology **14**(8), 791–799 (2003), https://doi.org/10.1046/j.1540-8167.2003.03078.x
20. Buchheit, M., Gindre, C.: Cardiac parasympathetic regulation: Respective associations with cardiorespiratory fitness and training load. American Journal of Physiology-Heart and Circulatory Physiology **291**(1), H451–H458 (2006), https://doi.org/10.1152/ajpheart.00008.2006
21. Buneman, P., Khanna, S., Wang-Chiew, T.: Why and where: A characterization of data provenance. In: Bussche, J., Vianu, V. (eds.) Database Theory — ICDT 2001. pp. 316–330. Springer, Berlin, Heidelberg, Germany (2001), https://doi.org/10.1007/3-540-44503-X_20
22. Canali, S., Schiaffonati, V., Aliverti, A.: Challenges and recommendations for wearable devices in digital health: Data quality, interoperability, health equity, fairness. PLoS Digital Health **1**(10), e0000104 (2022), https://doi.org/10.1371/journal.pdig.0000104
23. Castiglioni, P., Parati, G., Civijian, A., Quintin, L., Di Rienzo, M.: Local scale exponents of blood pressure and heart rate variability by detrended fluctuation analysis: Effects of posture, exercise, and aging. IEEE Transactions on Biomedical Engineering **56**(3), 675–684 (2009), https://doi.org/10.1109/TBME.2008.2005949
24. Chen, I., Johansson, F.D., Sontag, D.: Why is my classifier discriminatory? arXiv preprint arXiv:1805.12002 (2018), https://doi.org/10.48550/arXiv.1805.12002
25. Chen, I.Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., Ghassemi, M.: Ethical machine learning in healthcare. Annual Review of Biomedical Data Science **4**, 123–144 (2021), https://doi.org/10.1146/annurev-biodatasci-092820-114757
26. Chuangchai, W., Pothisiri, W.: Postural changes on heart rate variability among older population: A preliminary study. Current Gerontology and Geriatrics Research **2021**(1), 6611479 (2021), https://doi.org/10.1155/2021/6611479

27. Colvonen, P.J., DeYoung, P.N., Bosompra, N.O.A., Owens, R.L.: Limiting racial disparities and bias for wearable devices in health science research. Sleep **43**(10), zsaa159 (2020), https://doi.org/10.1093/sleep/zsaa159
28. Coravos, A., Khozin, S., Mandl, K.D.: Developing and adopting safe and effective digital biomarkers to improve patient outcomes. npj Digital Medicine **2**, 14 (2019), https://doi.org/10.1038/s41746-019-0090-4
29. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale entropy analysis of complex physiologic time series. Physical Review Letters **89**(6), 068102 (2002), https://doi.org/10.1103/PhysRevLett.89.068102
30. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale entropy analysis of biological signals. Physical Review E **71**(2), 021906 (2005), https://doi.org/10.1103/PhysRevE.71.021906
31. De Cannière, H., Corradi, F., Smeets, C.J., Schoutteten, M., Varon, C., Van Hoof, C., Van Huffel, S., Groenendaal, W., Vandervoort, P.: Wearable monitoring and interpretable machine learning can objectively track progression in patients during cardiac rehabilitation. Sensors **20**(12), 3601 (2020), https://doi.org/10.3390/s20123601
32. Dey, A.K.: Understanding and using context. Personal and Ubiquitous Computing **5**, 4–7 (2001), https://doi.org/10.1007/s007790170019
33. Earnest, C.P., Lavie, C.J., Blair, S.N., Church, T.S.: Heart rate variability characteristics in sedentary postmenopausal women following six months of exercise training: The DREW study. PloS One **3**(6), e2288 (2008), https://doi.org/10.1371/journal.pone.0002288
34. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., Dean, J.: A guide to deep learning in healthcare. Nature Medicine **25**, 24–29 (2019), https://doi.org/10.1038/s41591-018-0316-z
35. Fatisson, J., Oswald, V., Lalonde, F.: Influence diagram of physiological and environmental factors affecting heart rate variability: An extended literature overview. Heart International **11**(1), e32–e40 (2016), https://doi.org/10.5301/heartint.5000232
36. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Computing Surveys **46**(4), 1–37 (2014), https://doi.org/10.1145/2523813
37. Ghorbani, A., Abid, A., Zou, J.: Interpretation of neural networks is fragile. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 3681–3688 (2019), https://doi.org/10.1609/aaai.v33i01.33013681
38. Gierałtowski, J., Żebrowski, J., Baranowski, R.: Multiscale multifractal analysis of heart rate variability recordings with a large number of occurrences of arrhythmia. Physical Review E **85**(2), 021915 (2012), https://doi.org/10.1103/PhysRevE.85.021915
39. Goldberger, A.L., Amaral, L.A., Hausdorff, J.M., Ivanov, P.C., Peng, C.K., Stanley, H.E.: Fractal dynamics in physiology: Alterations with disease and aging. Proceedings of the National Academy of Sciences **99**(suppl_1), 2466–2472 (2002), https://doi.org/10.1073/pnas.012579499
40. Goldberger, A.L., Peng, C.K., Lipsitz, L.A.: What is physiologic complexity and how does it change with aging and disease? Neurobiology of Aging **23**(1), 23–26 (2002), https://doi.org/10.1016/S0197-4580(01)00266-4
41. Hausdorff, J.M.: Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking. Human Movement Science **26**(4), 555–589 (2007), https://doi.org/10.1016/j.humov.2007.05.003

42. Herman, T., Giladi, N., Gurevich, T., Hausdorff, J.: Gait instability and fractal dynamics of older adults with a "cautious" gait: Why do certain older adults walk fearfully? Gait & Posture **21**(2), 178–185 (2005), https://doi.org/10.1016/j.gaitpost.2004.01.014
43. Hillebrand, S., Gast, K.B., de Mutsert, R., Swenne, C.A., Jukema, J.W., Middeldorp, S., Rosendaal, F.R., Dekkers, O.M.: Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: Meta-analysis and dose–response meta-regression. Europace **15**(5), 742–749 (2013), https://doi.org/10.1093/europace/eus341
44. Hjortskov, N., Rissén, D., Blangsted, A.K., Fallentin, N., Lundberg, U., Søgaard, K.: The effect of mental stress on heart rate variability and blood pressure during computer work. European Journal of Applied Physiology **92**, 84–89 (2004), https://doi.org/10.1007/s00421-004-1055-z
45. Holzinger, A., Biemann, C., Pattichis, C.S., Kell, D.B.: What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923 (2017), https://doi.org/10.48550/arXiv.1712.09923
46. Huikuri, H.V., M"akikallio, T.H., Peng, C.K., Goldberger, A.L., Hintze, U., Maøller, M.: Fractal correlation properties of RR interval dynamics and mortality in patients with depressed left ventricular function after an acute myocardial infarction. Circulation **101**(1), 47–53 (2000), https://doi.org/10.1161/01.CIR.101.1.47
47. Huikuri, H.V., Niemelä, M., Ojala, S., Rantala, A., Ikäheimo, M., Airaksinen, K.: Circadian rhythms of frequency domain measures of heart rate variability in healthy subjects and patients with coronary artery disease. effects of arousal and upright posture. Circulation **90**(1), 121–126 (1994), https://doi.org/10.1161/01.CIR.90.1.121
48. Ioannidis, J.P., Bossuyt, P.M.: Waste, leaks, and failures in the biomarker pipeline. Clinical Chemistry **63**(5), 963–972 (2017), https://doi.org/10.1373/clinchem.2016.254649
49. Ivanov, P.C., Amaral, L.A.N., Goldberger, A.L., Havlin, S., Rosenblum, M.G., Struzik, Z.R., Stanley, H.E.: Multifractality in human heartbeat dynamics. Nature **399**, 461–465 (1999), https://doi.org/10.1038/20924
50. Ivanov, P.C., Liu, K.K., Bartsch, R.P.: Focus on the emerging new fields of network physiology and network medicine. New Journal of Physics **18**(10), 100201 (2016), https://doi.org/10.1088/1367-2630/18/10/100201
51. Izmailova, E.S., Wagner, J.A., Perakslis, E.D.: Wearable devices in clinical trials: Hype and hypothesis. Clinical Pharmacology & Therapeutics **104**(1), 42–52 (2018), https://doi.org/10.1002/cpt.966
52. Jain, S., Wallace, B.C.: Attention is not explanation. arXiv preprint arXiv:1902.10186 (2019), https://doi.org/10.48550/arXiv.1902.10186
53. Jurysta, F., Van De Borne, P., Migeotte, P.F., Dumont, M., Lanquart, J.P., Degaute, J.P., Linkowski, P.: A study of the dynamic interactions between sleep EEG and heart rate variability in healthy young men. Clinical Neurophysiology **114**(11), 2146–2155 (2003), https://doi.org/10.1016/S1388-2457(03)00215-3
54. Kamath, M.V., Fallen, E.L.: Power spectral analysis of heart rate variability: A noninvasive signature of cardiac autonomic function. Critical Reviews in Biomedical Engineering **21**(3), 245–311 (1993), https://pubmed.ncbi.nlm.nih.gov/8243093/
55. Kaze, A.D., Yuyun, M.F., Erqou, S., Fonarow, G.C., Echouffo-Tcheugui, J.B.: Cardiac autonomic neuropathy and risk of incident heart failure among adults

with type 2 diabetes. European Journal of Heart Failure **24**(4), 634–641 (2022), https://doi.org/10.1002/ejhf.2432

56. Kindermans, P.J., Hooker, S., Adebayo, J., Alber, M., Schütt, K.T., Dähne, S., Erhan, D., Kim, B.: The (un) reliability of saliency methods. In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.R. (eds.) Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, pp. 267–280. Springer, Cham, Switzerland (2019), https://doi.org/10.1007/978-3-030-28954-6_14
57. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016), https://doi.org/10.48550/arXiv.1609.02907
58. Kiviniemi, A.M., Hautala, A.J., Kinnunen, H., Tulppo, M.P.: Endurance training guided individually by daily heart rate variability measurements. European Journal of Applied Physiology **101**, 743–751 (2007), https://doi.org/10.1007/s00421-007-0552-2
59. Kleiger, R.E., Miller, J.P., Bigger Jr, J.T., Moss, A.J.: Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. American Journal of Cardiology **59**(4), 256–262 (1987), https://doi.org/10.1016/0002-9149(87)90795-8
60. Kleiger, R.E., Stein, P.K., Bigger Jr, J.T.: Heart rate variability: Measurement and clinical utility. Annals of Noninvasive Electrocardiology **10**(1), 88–101 (2005), https://doi.org/10.1111/j.1542-474X.2005.10101.x
61. Levin, A., Stevens, P.E., Bilous, R.W., Coresh, J., De Francisco, A.L., De Jong, P.E., Griffith, K.E., Hemmelgarn, B.R., Iseki, K., Lamb, E.J., et al.: Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney International Supplements **3**(1), 1–150 (2013), https://doi.org/10.1038/kisup.2012.73
62. Liang, Z., Wang, Y., Sun, X., Li, D., Voss, L.J., Sleigh, J.W., Hagihira, S., Li, X.: EEG entropy measures in anesthesia. Frontiers in Computational Neuroscience **9**, 16 (2015), https://doi.org/10.3389/fncom.2015.00016
63. Lillie, E.O., Patay, B., Diamant, J., Issell, B., Topol, E.J., Schork, N.J.: The n-of-1 clinical trial: The ultimate strategy for individualizing medicine? Personalized Medicine **8**(2), 161–173 (2011), https://doi.org/10.2217/pme.11.7
64. Losing, V., Hammer, B., Wersing, H.: Incremental on-line learning: A review and comparison of state of the art algorithms. Neurocomputing **275**, 1261–1274 (2018), https://doi.org/10.1016/j.neucom.2017.06.084
65. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., Zhang, G.: Learning under concept drift: A review. IEEE Transactions on Knowledge and Data Engineering **31**(12), 2346–2363 (2018), https://doi.org/10.1109/TKDE.2018.2876857
66. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems **30** (2017)
67. Mangalam, M., Kelty-Stephen, D.G., Sommerfeld, J.H., Stergiou, N., Likens, A.D.: Temporal organization of stride-to-stride variations contradicts predictive models for sensorimotor control of footfalls during walking. PloS One **18**(8), e0290324 (2023), https://doi.org/10.1371/journal.pone.0290324
68. McCamley, J., Donati, M., Grimpampi, E., Mazzà, C.: An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data. Gait & Posture **36**(2), 316–318 (2012), https://doi.org/10.1016/j.gaitpost.2012.02.019

69. Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., et al.: The open provenance model core specification (v1.1). Future Generation Computer Systems **27**(6), 743–756 (2011), https://doi.org/10.1016/j.future.2010.07.005
70. Nakamura, T., Kiyono, K., Wendt, H., Abry, P., Yamamoto, Y.: Multiscale analysis of intensive longitudinal biomedical signals and its clinical applications. Proceedings of the IEEE **104**(2), 242–261 (2016), https://doi.org/10.1109/JPROC.2015.2491979
71. Niemelä, K., Palatsi, I., Ikäheimo, M., Takkunen, J., Vuori, J.: Evidence of impaired left ventricular performance after an uninterrupted competitive 24 hour run. Circulation **70**(3), 350–356 (1984), https://doi.org/10.1161/01.CIR.70.3.350
72. Niizeki, K., Kawahara, K., Miyamoto, Y.: Interaction among cardiac, respiratory, and locomotor rhythms during cardiolocomotor synchronization. Journal of Applied Physiology **75**(4), 1815–1821 (1993), https://doi.org/10.1152/jappl.1993.75.4.1815
73. Niizeki, K., Saitoh, T.: Cardiolocomotor phase synchronization during rhythmic exercise. Journal of Physical Fitness and Sports Medicine **3**(1), 11–20 (2014), https://doi.org/10.7600/jpfsm.3.11
74. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.: Dissecting racial bias in an algorithm used to manage the health of populations. Science **366**(6464), 447–453 (2019), https://doi.org/10.1126/science.aax2342
75. Otović, E., Njirjak, M., Jozinović, D., Mauša, G., Michelini, A., Stajduhar, I.: Intra-domain and cross-domain transfer learning for time series data—How transferable are the features? Knowledge-Based Systems **239**, 107976 (2022), https://doi.org/10.1016/j.knosys.2021.107976
76. Parisi, G.I., Kemker, R., Part, J.L., Kanan, C., Wermter, S.: Continual lifelong learning with neural networks: A review. Neural Networks **113**, 54–71 (2019), https://doi.org/10.1016/j.neunet.2019.01.012
77. Peltola, M.A.: Role of editing of R–R intervals in the analysis of heart rate variability. Frontiers in Physiology **3**, 148 (2012), https://doi.org/10.3389/fphys.2012.00148
78. Peng, C.K., Havlin, S., Stanley, H.E., Goldberger, A.L.: Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos **5**(1), 82–87 (1995), https://doi.org/10.1063/1.166141
79. Plews, D.J., Laursen, P.B., Stanley, J., Kilding, A.E., Buchheit, M.: Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring. Sports Medicine **43**, 773–781 (2013), https://doi.org/10.1007/s40279-013-0071-8
80. Pontes, E.D., Pinto, M., Lopes, F., Teixeira, C.: Concept-drifts adaptation for machine learning EEG epilepsy seizure prediction. Scientific Reports **14**, 8204 (2024), https://doi.org/10.1038/s41598-024-57744-1
81. Rajkomar, A., Dean, J., Kohane, I.: Machine learning in medicine. New England Journal of Medicine **380**(14), 1347–1358 (2019), https://doi.org/10.1056/NEJMra1814259
82. Rajkomar, A., Hardt, M., Howell, M.D., Corrado, G., Chin, M.H.: Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine **169**(12), 866–872 (2018), https://doi.org/10.7326/M18-1990
83. Ramachandram, D., Taylor, G.W.: Deep multimodal learning: A survey on recent advances and trends. IEEE Signal Processing Magazine **34**(6), 96–108 (2017), https://doi.org/10.1109/MSP.2017.2738401

84. Ribeiro, M.T., Singh, S., Guestrin, C.: "why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1135–1144 (2016), https://doi.org/10.1145/2939672.2939778
85. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology **278**(6), H2039–H2049 (2000), https://doi.org/10.1152/ajpheart.2000.278.6.H2039
86. Rifai, N., Gillette, M.A., Carr, S.A.: Protein biomarker discovery and validation: The long and uncertain path to clinical utility. Nature Biotechnology **24**, 971–983 (2006), https://doi.org/10.1038/nbt1235
87. Riganello, F., Larroque, S.K., Bahri, M.A., Heine, L., Martial, C., Carrière, M., Charland-Verville, V., Aubinet, C., Vanhaudenhuyse, A., Chatelle, C., et al.: A heartbeat away from consciousness: Heart rate variability entropy can discriminate disorders of consciousness and is correlated with resting-state fMRI brain connectivity of the central autonomic network. Frontiers in Neurology **9**, 769 (2018), https://doi.org/10.3389/fneur.2018.00769
88. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence **1**, 206–215 (2019), https://doi.org/10.1038/s42256-019-0048-x
89. Schubert, C., Lambertz, M., Nelesen, R., Bardwell, W., Choi, J.B., Dimsdale, J.: Effects of stress on heart rate complexity—A comparison between short-term and chronic stress. Biological Psychology **80**(3), 325–332 (2009), https://doi.org/10.1016/j.biopsycho.2008.11.005
90. Segerstrom, S.C., Nes, L.S.: Heart rate variability reflects self-regulatory strength, effort, and fatigue. Psychological Science **18**(3), 275–281 (2007), https://doi.org/10.1111/j.1467-9280.2007.01888.x
91. Shaffer, F., Ginsberg, J.P.: An overview of heart rate variability metrics and norms. Frontiers in Public Health **5**, 258 (2017), https://doi.org/10.3389/fpubh.2017.00258
92. Shaffer, F., McCraty, R., Zerr, C.L.: A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability. Frontiers in Psychology **5**, 1040 (2014), https://doi.org/10.3389/fpsyg.2014.01040
93. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034 (2013), https://doi.org/10.48550/arXiv.1312.6034
94. Singh, J.P., Larson, M.G., O'Donnell, C.J., Tsuji, H., Evans, J.C., Levy, D.: Heritability of heart rate variability: The Framingham Heart Study. Circulation **99**(17), 2251–2254 (1999), https://doi.org/10.1161/01.CIR.99.17.2251
95. Soni, A., Rawal, K.: Analyzing the effect of postural change on heart rate variability using multi-distance sample entropy (MDSE). Biomedical Signal Processing and Control **87**, 105476 (2024), https://doi.org/10.1016/j.bspc.2023.105476
96. Storm, F.A., Buckley, C.J., Mazzà, C.: Gait event detection in laboratory and real life settings: Accuracy of ankle and waist sensor based methods. Gait & Posture **50**, 42–46 (2016), https://doi.org/10.1016/j.gaitpost.2016.08.012
97. Strimbu, K., Tavel, J.A.: What are biomarkers? Current Opinion in HIV and AIDS **5**(6), 463–466 (2010), https://doi.org/10.1097/COH.0b013e32833ed177
98. Su, C., Liang, Z., Li, X., Li, D., Li, Y., Ursino, M.: A comparison of multiscale permutation entropy measures in on-line depth of anesthesia monitoring. PLoS One **11**(10), e0164104 (2016), https://doi.org/10.1371/journal.pone.0164104

99. Takeda, R., Lisco, G., Fujisawa, T., Gastaldi, L., Tohyama, H., Tadano, S.: Drift removal for improving the accuracy of gait parameters using wearable sensor systems. Sensors **14**(12), 23230–23247 (2014), https://doi.org/10.3390/s141223230
100. Tamura, T., Maeda, Y., Sekine, M., Yoshida, M.: Wearable photoplethysmographic sensors—past and present. Electronics **3**(2), 282–302 (2014), https://doi.org/10.3390/electronics3020282
101. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology: Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Circulation **93**(5), 1043–1065 (1996), https://doi.org/10.1161/01.CIR.93.5.1043
102. Thayer, J.F., Hansen, A.L., Saus-Rose, E., Johnsen, B.H.: Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. Annals of Behavioral Medicine **37**(2), 141–153 (2009), https://doi.org/10.1007/s12160-009-9101-z
103. Thayer, J.F., Lane, R.D.: The role of vagal function in the risk for cardiovascular disease and mortality. Biological Psychology **74**(2), 224–242 (2007), https://doi.org/10.1016/j.biopsycho.2005.11.013
104. Thayer, J.F., Lane, R.D.: Claude Bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration. Neuroscience & Biobehavioral Reviews **33**(2), 81–88 (2009), https://doi.org/10.1016/j.neubiorev.2008.08.004
105. Thayer, J.F., Yamamoto, S.S., Brosschot, J.F.: The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. International Journal of Cardiology **141**(2), 122–131 (2010), https://doi.org/10.1016/j.ijcard.2009.09.543
106. Tjoa, E., Guan, C.: A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE Transactions on Neural Networks and Learning Systems **32**(11), 4793–4813 (2021), https://doi.org/10.1109/TNNLS.2020.3027314
107. Tobaldini, E., Nobili, L., Strada, S., Casali, K.R., Braghiroli, A., Montano, N.: Heart rate variability in normal and pathological sleep. Frontiers in Physiology **4**, 294 (2013), https://doi.org/10.3389/fphys.2013.00294
108. Topol, E.J.: High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine **25**, 44–56 (2019), https://doi.org/10.1038/s41591-018-0300-7
109. Umetani, K., Singer, D.H., McCraty, R., Atkinson, M.: Twenty-four hour time domain heart rate variability and heart rate: Relations to age and gender over nine decades. Journal of the American College of Cardiology **31**(3), 593–601 (1998), https://doi.org/10.1016/s0735-1097(97)00554-8
110. Vasan, R.S.: Biomarkers of cardiovascular disease: Molecular basis and practical considerations. Circulation **113**(19), 2335–2362 (2006), https://doi.org/10.1161/CIRCULATIONAHA.104.482570
111. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. arXiv preprint arXiv:1805.12002 (2018), https://doi.org/10.48550/arXiv.1805.12002
112. Voss, A., Kurths, J., Kleiner, H., Witt, A., Wessel, N., Saparin, P., Osterziel, K., Schurath, R., Dietz, R.: The application of methods of nonlinear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death. Cardiovascular Research **31**(3), 419–433 (1996), https://doi.org/10.1016/S0008-6363(96)00008-9

113. Widmer, G., Kubat, M.: Learning in the presence of concept drift and hidden contexts. Machine Learning **23**, 69–101 (1996), https://doi.org/10.1023/A:1018046501280
114. Wiegreffe, S., Pinter, Y.: Attention is not not explanation. arXiv preprint arXiv:1908.04626 (2019), https://doi.org/10.48550/arXiv.1908.04626
115. Wilson, T.J., Mangalam, M., Stergiou, N., Likens, A.D.: Multifractality in stride-to-stride variations reveals that walking involves more movement tuning and adjusting than running. Frontiers in Network Physiology **3**, 1294545 (2023), https://doi.org/10.3389/fnetp.2023.1294545
116. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems **32**(1), 4–24 (2021), https://doi.org/10.1109/TNNLS.2020.2978386
117. Zhang, J.: Effect of age and sex on heart rate variability in healthy subjects. Journal of Manipulative and Physiological Therapeutics **30**(5), 374–379 (2007), https://doi.org/10.1016/j.jmpt.2007.04.001
118. Zhu, Y., Huang, X., Li, L.: Dynamic prediction of time to a clinical event with sparse and irregularly measured longitudinal biomarkers. Biometrical Journal **62**(6), 1371–1393 (2020), https://doi.org/10.1002/bimj.201900112
119. Zijlstra, W., Hof, A.L.: Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait & Posture **18**(2), 1–10 (2003), https://doi.org/10.1016/S0966-6362(02)00190-X