

Machine Learning-Based Continuous Blood Pressure Estimation Using Noisy Photoplethysmography Signals

Deminik Neal

Department of Computer Science, University of Central Florida, Orlando, FL, USA.
nealdominik@ucf.edu

Dustin Koskinen

Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
dkoskinen@colostate.edu

Florian Watson

Department of Electrical Engineering and Computer Science, University of Missouri,
Columbia, MO, USA.
florianw@missouri.edu

Abstract

Continuous blood pressure monitoring is central to cardiovascular risk management, yet conventional cuff-based devices remain episodic and intrusive. Photoplethysmography provides a low-cost optical signal that can be embedded in wearable devices, but ambulatory recordings are degraded by motion, sensor displacement, ambient light, and tissue heterogeneity. Machine learning has been proposed to map noisy photoplethysmographic waveforms to continuous blood pressure estimates. However, reliable translation requires a systems perspective that extends beyond model architecture. This article examines the structural trade-offs involved in signal conditioning, feature extraction, learning strategies, data governance, deployment, and regulatory alignment. It argues that robust systems must combine adaptive preprocessing with uncertainty-aware models, representative datasets, validation against clinical standards, and explicit fairness considerations. A layered design is discussed in which local signal restoration, population-level calibration, and institutional governance operate together. The analysis integrates evidence from physiological sensing, machine learning, and health policy to identify barriers and pathways toward sustainable continuous blood pressure estimation. Emphasis is placed on cross-domain lessons such as pipeline governance, auditability, and lifecycle maintenance. The conclusion highlights the need for coordinated technical and regulatory evolution rather than isolated algorithmic optimization.

Keywords

continuous blood pressure estimation; photoplethysmography; machine learning; noise robustness; clinical deployment; data governance; fairness.

1. Introduction

Hypertension remains a leading modifiable risk factor for cardiovascular disease, stroke, and renal failure, yet its detection and management are limited by the episodic nature of cuff-based measurement. Continuous noninvasive blood pressure monitoring could support early

detection, treatment titration, and lifestyle intervention. Photoplethysmography has attracted sustained attention because it can be acquired with inexpensive optical sensors that are already embedded in watches, rings, and bedside monitors [1]. Unlike oscillometric cuffs, photoplethysmographic sensors do not interrupt the user and can provide a continuous stream of pulse-related information. However, the path from a raw photoplethysmographic waveform to a clinically meaningful blood pressure value involves multiple system layers, including sensor hardware, signal conditioning, feature representation, machine learning, calibration, and clinical validation. The central challenge is not simply to fit a model to laboratory data but to design a robust measurement infrastructure that remains trustworthy under ambulatory noise.

The photoplethysmographic waveform reflects blood volume changes in the microvascular bed and is modulated by cardiac ejection, vascular tone, respiration, and autonomic state [2]. Morphological characteristics of the pulse wave, including its systolic upstroke, dicrotic notch, and diastolic decay, contain information that correlates with arterial stiffness and pressure. Nevertheless, the physiological relationship is indirect, and the relation can vary across individuals, measurement sites, and hemodynamic states. Because photoplethysmography is susceptible to motion artifact, sensor pressure, skin pigmentation, and ambient light, any system intended for continuous blood pressure estimation must treat noise as a first-order design constraint. International standards for wearable cuffless blood pressure devices have defined expectations for static accuracy, environmental tolerance, and labeling, but they do not yet fully resolve how machine learning models should be evaluated under real-world conditions [3]. The validation of blood pressure devices has been formalized through consensus protocols that emphasize reference measurements, subject diversity, and statistical reporting [4]. These protocols provide an essential governance layer for clinical acceptance, but their application to continuously adaptive learning systems remains unsettled.

Early algorithmic work demonstrated that pulse features and regression models can estimate blood pressure with promising laboratory results, and subsequent studies have applied deeper neural architectures to raw signal segments [5]. Yet translation into clinical and consumer settings has been slower than initial results suggested. One reason is that laboratory datasets frequently contain clean recordings, controlled posture, and narrow demographic distributions, while field-deployed devices encounter heterogeneous users, intermittent connectivity, firmware updates, sensor drift, and uncontrolled motion. A system-level analysis therefore must examine not only the estimator architecture but also the data supply chain, failure modes, and institutional mechanisms that sustain performance over time. This article addresses these dimensions through an interdisciplinary lens, drawing on signal processing, machine learning, health policy, and infrastructure studies. The following sections examine the photoplethysmographic signal chain, modeling trade-offs, data governance, robustness, fairness, deployment, and sustainability.

2. The PPG Signal Chain and Noise Landscape

The photoplethysmographic signal chain begins with optical transduction. A light source, typically in the green, red, or infrared spectrum, illuminates the skin, and a photodetector measures either transmitted or reflected light that varies with pulsatile blood volume. The choice of wavelength affects depth of penetration and susceptibility to motion and ambient light. Green light is often preferred in wearable reflectance sensors because it is strongly absorbed by hemoglobin and relatively robust to certain motion artifacts, while red and infrared configurations are common in transmission-mode pulse oximeters [1]. The analog

front-end must reject out-of-band interference, avoid saturation, and preserve the small alternating-current component of the signal against a much larger quasi-static baseline. Digitization then introduces quantization noise and places demands on sampling rate, dynamic range, and power consumption. Each decision in this chain creates a trade-off between signal fidelity, sensor cost, battery life, form factor, and data volume.

Noise in ambulatory photoplethysmography is not a monolithic distortion. Motion artifact arises from relative movement between the sensor and skin, changing the optical path and producing large, structured fluctuations that can resemble pulse waveforms. Ambient light can saturate the photodetector or introduce periodic interference from artificial lighting. Respiration and venous pulsation contribute low-frequency baseline wander. Sensor pressure can alter local perfusion and waveform morphology. Skin pigmentation and tissue scattering influence signal amplitude and signal-to-noise ratio, creating both technical and equity concerns. These noise sources are nonstationary and often correlated with the physiological state being estimated. For example, physical activity simultaneously increases motion artifact and changes blood pressure through sympathetic activation. This coupling makes simple filtering dangerous because the removal of artifact can also remove diagnostically meaningful hemodynamic variation. A robust system therefore requires adaptive signal restoration that distinguishes artifact from true physiological dynamics.

Signal conditioning architectures range from fixed band-pass filters and wavelet denoising to adaptive notch filters and motion reference fusion. Fixed filters are computationally inexpensive but assume stationary noise spectra. Wavelet methods can isolate transient features but may require careful threshold selection and introduce latency. Motion reference approaches, using accelerometer or gyroscope signals, can guide the suppression of movement-related components, but they require tight synchronization and may not capture all artifact mechanisms. Adaptive filtering is attractive because it can adjust to changing conditions, but adaptation introduces state memory and potential instability if the reference signal is poorly correlated with artifact. The system-level implication is that preprocessing cannot be treated as an isolated cleanup stage; it is part of the overall estimation pipeline and must be designed jointly with downstream feature extraction and model calibration. Choices made here propagate to data distribution, model complexity, update frequency, and regulatory explainability.

3. Machine Learning Architectures and Their System-Level Trade-Offs

Machine learning approaches for photoplethysmography-based blood pressure estimation can be organized into two broad families. The first extracts handcrafted morphological features from the pulse wave and fits a regression model. These features may include systolic upstroke time, pulse width, reflected wave timing, area ratios, and derivatives. The second family feeds raw or minimally processed signal segments into deep neural networks, allowing the model to discover representations. Spectro-temporal neural architectures have been applied to photoplethysmographic segments and have reported strong in-domain performance on benchmark datasets [6]. Deep models can exploit subtle interactions across time and frequency, but they also introduce challenges related to data hunger, computational cost, interpretability, and sensitivity to distribution shift. Handcrafted feature pipelines tend to be more transparent and less data intensive, but they may discard information that is not captured by the chosen features.

Tree-based ensembles such as gradient boosting remain important for tabular features because they are robust to heterogeneous scales, missing data, and nonlinear interactions [7]. They can

be trained on smaller datasets and can be deployed on modest hardware, making them attractive for embedded devices. However, they rely heavily on the quality of the upstream feature extraction and may fail when the feature set does not encode the relevant hemodynamic information. Alternative physiological signals have been combined with photoplethysmography to improve estimation. Pulse transit time, often derived from the electrocardiogram and a peripheral optical signal, provides a more direct surrogate for arterial stiffness and has been extensively studied for cuffless blood pressure monitoring [8]. Multimodal fusion adds hardware complexity and synchronization burdens, but it can reduce the ambiguity of photoplethysmography-only estimates. The choice between single-sensor and multisensor architectures is therefore not purely a matter of accuracy; it involves user burden, power, regulatory scope, and deployment infrastructure.

Adaptive preprocessing can be coupled with machine learning to improve robustness. Adaptive notch-filtration has been shown to recover photoplethysmographic signals during physical activity, thereby reducing the artifact burden before model training and inference [9]. This kind of front-end restoration changes the downstream learning problem: the model no longer has to memorize artifact patterns, which may differ across devices and populations. At the same time, adaptive filtering introduces additional state variables, such as filter coefficients and reference signal quality, that must be monitored. If the adaptive stage diverges, the model receives distorted inputs and may produce confident but erroneous estimates. A system-level design should therefore include anomaly detectors that assess signal quality before estimation and suppress outputs when confidence is low. Many fielded systems combine a signal quality index, a physiological plausibility check, and a model confidence gate. These safeguards are often more important for clinical safety than incremental improvements in mean absolute error.

Another architectural boundary is the degree of personalization. Population models trained across many subjects must contend with diverse vascular physiology, medication effects, and sensor placement. Personalized calibration can improve accuracy by adapting to individual physiology, but it requires reference measurements and creates a maintenance burden. This trade-off resembles the cold-start problem in recommendation systems and the calibration problem in continuous glucose monitoring. A global model may be immediately usable but less precise for some users, while a personalized model may be precise but dependent on periodic cuff measurements and vulnerable to drift. Hybrid architectures attempt to combine a global representation with lightweight user-specific adaptation, yet this raises questions about data ownership, update authority, and how to validate continuously changing models.

4. Data Infrastructure, Labeling, and Governance

The reliability of any machine learning system is bounded by the quality and representativeness of its training data. In continuous blood pressure estimation, ground truth labels are usually obtained from oscillometric cuffs, arterial catheters, or auscultatory references, each of which has distinct error characteristics and temporal resolution. The reference device may be discontinuous, whereas the photoplethysmogram is continuous, producing an alignment problem that can introduce label noise. Multimodal features derived from photoplethysmography and electrocardiography have been explored to capture blood pressure changes, but their value depends on consistent beat detection, synchronization, and artifact rejection [10]. Deep architectures have also been developed to estimate blood pressure and heart rate jointly from photoplethysmographic and derivative signals, illustrating the

tendency of modern models to consolidate multiple physiological tasks [11]. These modeling advances, however, depend on data pipelines that are rarely discussed in algorithmic papers.

Data infrastructure for training must address collection protocols, participant selection, sensor placement, environmental conditions, and reference measurement timing. If training data are collected in quiet laboratory conditions, the resulting model may fail during walking, typing, or sleep. If only healthy adults are included, the model may not generalize to older adults, hypertensive patients, or individuals with arrhythmias. Data labeling is also complicated by the fact that blood pressure changes continuously with respiration, emotion, and activity, so a single cuff value may not represent the average pressure over the window used for model input. The temporal mismatch creates target ambiguity that can mislead even well-regularized learners. Furthermore, labels from different reference devices are not interchangeable without a calibration offset and a defined uncertainty model. Oscillometric cuffs, for example, estimate mean arterial pressure and then derive systolic and diastolic values through proprietary algorithms, whereas intra-arterial catheters measure a direct waveform that may be influenced by transducer location and damping. These discrepancies add systematic bias to the training target and complicate cross-study comparisons. A governance framework must therefore require detailed reference device metadata, including measurement standard, calibration interval, and synchronization precision. Without such metadata, the data pipeline cannot support the reproducibility expectations of clinical machine learning.

Data governance for ambulatory sensing must also address consent, data ownership, and secondary use. Photoplethysmographic waveforms contain information about cardiac activity, respiratory patterns, and possibly emotional state, even when the user perceives the sensor as a simple heart rate monitor. The aggregation of longitudinal physiological data across large populations raises privacy concerns that conventional anonymization cannot fully resolve, because waveform morphology can be individually distinctive. Infrastructure design should minimize raw data retention, favor on-device feature extraction, and apply differential privacy or federated learning where centralization is not needed. Federated approaches distribute model updates rather than personal waveforms, but they introduce new coordination costs and may still leak information through gradients. These governance decisions are not secondary to the machine learning task; they shape the dataset that can be legally and ethically assembled, and they determine the trust that patients and clinicians place in the system.

A further challenge is the temporal drift of both sensors and physiology. Optical sensors degrade over time due to encapsulation wear, light source aging, and detector sensitivity changes. Physiological calibration can shift with medication, hydration, arterial remodeling, and disease progression. A fixed model trained on a static snapshot may therefore become inaccurate months after deployment, even if it was valid at installation. Lifecycle management requires scheduled recalibration, model monitoring, and incident response. In this sense, continuous blood pressure estimation is not a product but a service, with an operational model that includes data collection, performance surveillance, and feedback loops. This perspective is familiar in large-scale systems engineering, where reliability is achieved through redundancy, monitoring, and planned maintenance rather than through a single static algorithm. Dataset documentation and model cards should record the intended use, demographic composition, and known limitations of training data [14].

5. Robustness, Fairness, and Cross-Population Generalization

Robustness in photoplethysmography-based blood pressure estimation has two dimensions: robustness to sensor-level noise and robustness to distributional shift across populations and

environments. Motion artifact and signal dropout are the most visible noise sources, but equally important are changes in skin perfusion, ambient temperature, sensor placement, and user behavior. A model that performs well on clean laboratory segments may fail when the optical path is altered by wrist flexion or when the user has cold extremities. Signal quality indicators can be used to gate predictions, but they must themselves be validated against reference outcomes. If the quality index is too conservative, the system produces frequent gaps and becomes clinically useless; if it is too permissive, erroneous estimates erode trust. The thresholding decision is therefore a risk management trade-off rather than a purely technical parameter. Research on wearable optical heart rate sensors has demonstrated that skin pigmentation, motion, and sensor coupling can affect accuracy, underscoring the need for inclusive evaluation [15].

Fairness concerns arise because photoplethysmographic signal amplitude and morphology vary with skin pigmentation, vascular density, and tissue optical properties. Optical sensors have historically been calibrated predominantly on lighter skin tones, and pulse oximetry has shown well-documented racial bias under certain conditions. Similar risks exist for blood pressure estimation, where attenuation of the optical signal in darker skin may reduce signal-to-noise ratio and alter waveform morphology. If training datasets underrepresent darker skin tones or certain body mass distributions, the model may exhibit systematically larger errors in those groups. This is not merely a data sampling problem; it is a structural flaw in the measurement pipeline that requires inclusive data collection, skin tone reporting, optical sensor design, and algorithmic bias audits. Ethical machine learning in healthcare demands that model developers anticipate these disparities and test for them explicitly [16].

Cross-population generalization also involves age, vascular stiffness, medication effects, and comorbidity. Elderly patients may have less compliant arteries and different pulse wave reflections, while hypertensive patients may show altered pulse contours. A model trained on healthy students cannot be expected to capture the hemodynamic states of patients with heart failure or diabetes. Domain adaptation and transfer learning offer partial remedies, but they require representative target data and careful regularization. Another lesson from large-scale machine learning is that data diversity should be treated as an engineering requirement with measurable targets, not as an optional enhancement. Dataset documentation should include demographic distributions, health status, sensor types, and environmental conditions. Without such documentation, clinicians cannot judge whether a model is appropriate for their patient population.

The interaction between fairness and noise is particularly important because disadvantaged groups may also experience more challenging measurement conditions. For example, users in physically demanding occupations may have higher motion artifact and limited time for stable measurement. A system that dismisses noisy segments may disproportionately reduce availability for those users, creating a form of algorithmic denial. Fairness metrics should therefore include not only error magnitude but also data completeness, prediction frequency, and the distribution of confidence outputs across groups. Governance mechanisms can mandate disaggregated evaluation before deployment and post-market surveillance to detect emerging disparities. Explainability methods can help auditors understand which waveform features drive predictions, but they are not a substitute for robust performance auditing [17].

6. Deployment, Calibration, and Lifecycle Management

Deployment of continuous blood pressure estimation systems involves a series of decisions about computation placement, data flow, and intervention protocols. On-device inference

reduces latency and privacy exposure but limits model complexity and update frequency. Cloud-based inference enables larger models and centralized monitoring but requires reliable connectivity and may introduce unacceptable latency for real-time alerts. Hybrid architectures often place a lightweight model on the device and a more complex model in the cloud, with synchronization and conflict resolution between the two. This layered pattern is well known in distributed systems and has direct implications for medical safety: if a device and cloud model disagree, the system must have a defined arbitration rule, and the user must understand which estimate is authoritative. Personalized cuffless blood pressure estimation often uses a global model with individualized calibration to balance these trade-offs [18].

Calibration strategy is a key lifecycle decision. Population models can be deployed with no user-specific training, but they may not meet clinical accuracy thresholds for all users. Personalized calibration requires one or more reference blood pressure measurements to adjust model parameters or select a user-specific mapping. The frequency of recalibration depends on physiological drift and sensor stability. Some systems use periodic cuff measurements as anchors, while others estimate drift from internal consistency checks. A critical risk is silent miscalibration, in which the model continues to output plausible values that have slowly diverged from the true pressure. This failure mode is more dangerous than a complete outage because it may not trigger an alarm. Continuous monitoring systems in other safety-critical domains address this by combining redundant measurements, plausibility bounds, and explicit uncertainty reporting. Cuffless blood pressure systems should adopt similar principles, displaying confidence intervals or qualitative reliability indicators. The convergence of artificial intelligence and medicine offers opportunities for improvement, but it also raises fundamental questions about clinical accountability when models are continuously updated [19].

Model updates over time raise additional regulatory and governance questions. A model may be retrained as new data become available, but each update can change behavior for individual users in ways that are difficult to predict. Versioning, rollback, and shadow deployment are standard practices in software engineering but are underdeveloped in medical machine learning. A blood pressure estimation system should maintain an audit trail of model versions, calibration events, sensor replacements, and reference measurements. This traceability supports post-market surveillance, failure investigation, and clinical review. Without it, clinicians and regulators cannot distinguish a sensor fault from a model error or a physiological change. Wearable photoplethysmography for cardiovascular monitoring has shown promise, but its clinical utility depends on sustained performance across real-world use [20].

Energy consumption and sustainability are often overlooked in the machine learning literature but are central to wearable deployment. A continuous photoplethysmographic sensor with frequent neural network inference can drain a small battery within hours, reducing user adherence. Lower-power modes may sacrifice sampling rate or model complexity, introducing a trade-off between accuracy and battery life. At the population scale, the energy footprint of cloud retraining and data storage contributes to the environmental burden of digital health infrastructure. A sustainable design should minimize unnecessary data transfer, use efficient architectures, and consider the full lifecycle of devices and cloud resources. The evaluation of cuffless blood pressure devices has revealed significant challenges in calibration and validation, emphasizing the need for standardized protocols that account for device learning and adaptation [21].

7. Policy, Regulation, and Institutional Governance

Regulatory frameworks for cuffless blood pressure devices are evolving. Traditional sphygmomanometers are evaluated against international protocols that specify sample size, reference method, and error thresholds [3][4]. The IEEE standard for wearable cuffless blood pressure measuring devices defines technical requirements for safety, labeling, and performance, but it does not fully address continuously learning machine learning components [13]. Regulatory agencies have begun to provide guidance for software as a medical device and artificial intelligence, emphasizing risk-based classification, clinical validation, and change control. However, the continuous adaptation of models creates tension with traditional premarket approval, because a model that changes after approval may no longer be the same device that was validated. A possible governance model is a confined learning loop with explicit change control, where updates are evaluated against predefined acceptance criteria and released through a controlled process. This mirrors the concept of continuous integration in software systems but with clinical safety gates. Legal and ethical analyses of artificial intelligence in healthcare stress the need for accountability mechanisms that can keep pace with model evolution [22].

Institutional governance also includes procurement, clinical workflow integration, and professional accountability. Hospitals and health systems must decide whether a cuffless device is a screening tool, a diagnostic device, or a monitoring adjunct, because each role carries different regulatory and liability implications. Clinicians need clear guidance on how to interpret continuous blood pressure trends, which may fluctuate with activity and stress. Alerts based on noisy estimates can lead to alarm fatigue, while missing true hypertensive episodes can produce false reassurance. Protocols must define escalation thresholds, confirmation measurements with cuffs, and documentation standards. The involvement of clinical engineers, data stewards, and patient safety officers is necessary to manage the device beyond the initial validation study. Reviews of artificial intelligence implementation in medicine emphasize that clinical adoption depends on workflow integration and evidence of patient benefit, not on algorithmic novelty alone [23].

Fairness policies must be embedded in procurement and post-market surveillance. A healthcare system that adopts a model without evidence of performance across its patient population may inadvertently institutionalize bias. Disaggregated evaluation should be a condition of purchase, and real-world performance should be audited periodically. If disparities emerge, the system operator must have the authority to suspend the model, recalibrate, or require vendor remediation. This governance layer transforms machine learning from a purely technical artifact into a sociotechnical infrastructure whose safety depends on organizational routines. International harmonization is complicated by differences in data protection law, medical device regulation, and clinical practice. A model developed in one region may not easily transfer to another because reference standards, patient demographics, and healthcare workflows differ. Cross-border validation requires careful mapping of regulatory categories and collaborative governance. The global nature of wearable platforms also raises questions about data sovereignty, because personal physiological data may traverse multiple jurisdictions. A responsible system should localize sensitive processing where possible and provide transparent data flow documentation to users and regulators.

8. Conclusion

Machine learning-based continuous blood pressure estimation from noisy photoplethysmographic signals is a systems problem, not solely an algorithmic one. The path

from optical waveform to clinical insight requires adaptive signal conditioning, robust feature extraction, representative training data, uncertainty management, and institutional governance. This article has argued that each layer of the pipeline creates structural trade-offs that cannot be resolved by optimizing a single performance metric. Preprocessing must balance artifact removal against preservation of hemodynamic information, model architectures must balance accuracy against interpretability and computational cost, and data governance must reconcile privacy protection with the need for population diversity. Adaptive signal restoration can significantly improve robustness under physical activity, but it also introduces stateful components that require monitoring and quality control. Similarly, personalization improves individual accuracy but complicates lifecycle management and regulatory certainty.

The deployment of these systems should be guided by the principles of safety-critical infrastructure: redundancy, traceability, planned maintenance, and explicit failure modes. Cuffless blood pressure monitors that silently provide inaccurate values pose a greater risk than those that refuse to measure under uncertain conditions. Fairness must be treated as an operational requirement rather than a post hoc audit. This means defining measurable equity targets during dataset construction, model selection, and deployment planning. Reference data should include skin tone assessments using validated scales such as the Fitzpatrick or Monk scale, and sensor hardware should be tested under varied tissue optical properties. When such diversity is not achieved, the model's intended use should be explicitly limited. During post-market surveillance, aggregated performance should be monitored across demographic subgroups, and any statistically significant disparity should trigger a review process. Confidence outputs should also be examined for bias, because a system that reports lower confidence for certain groups may lead clinicians to distrust or ignore those estimates disproportionately. The fairness mandate extends beyond the model to the entire measurement infrastructure, including sensor design, preprocessing, and user interface.

The deployment of a continuous blood pressure estimation system also requires careful attention to model update governance. A static model may become inaccurate over time due to sensor drift, population changes, or evolving clinical standards. A continuously updated model creates a moving target for regulators and clinicians. A middle path is to separate the fixed feature extractor from the adaptive calibration layer. The feature extractor, which has undergone extensive validation, remains stable, while the calibration layer can be adjusted within a predefined range. This design allows personalization without changing the core algorithm, simplifying regulatory review. The adaptive calibration should be constrained by plausibility bounds derived from physiological knowledge, such as the expected relationship between pulse transit time and blood pressure. When the adaptation would push parameters outside these bounds, the system should request a reference cuff measurement rather than silently adjusting. This is analogous to safety interlocks in industrial control systems, where autonomous adjustments are allowed only within certified envelopes.

Post-market surveillance must become a standard part of medical machine learning. Regulators increasingly expect manufacturers to monitor real-world performance after approval, but the infrastructure for such monitoring is still immature. Health systems can contribute by establishing registries of wearable device data linked to clinical outcomes, but this raises additional privacy and data governance challenges. Federated approaches can enable performance monitoring without centralizing personal waveforms, and they have been proposed for digital health applications [23]. In a federated framework, local devices or

hospital servers compute aggregate error statistics and share only model updates or summary metrics, reducing the risk of re-identification. However, federated architectures require careful synchronization, communication efficiency, and protection against malicious participants. The choice between centralized and federated monitoring is not purely technical; it depends on the trust model between patients, providers, and device manufacturers.

International policy alignment remains a significant obstacle. The IEEE standard for wearable cuffless blood pressure measuring devices provides a technical foundation, but it does not specify how machine learning models should be validated under continuous adaptation [13]. Regional regulators have adopted different risk classifications, leading to a fragmented market where the same device may be considered a wellness product in one jurisdiction and a medical device in another. This disparity undermines clinical trust and complicates cross-border data sharing. A harmonized approach would define minimum requirements for reference data, demographic reporting, noise stress testing, and uncertainty communication. It would also clarify the conditions under which a model update constitutes a new device. Legal scholarship has argued that privacy and safety regulations must evolve in tandem with machine learning capacity, particularly when data are generated continuously and outside clinical settings [21]. The convergence of high-performance medicine and artificial intelligence has raised expectations, but it has also revealed the limits of traditional evidence hierarchies [18].

Ultimately, the success of machine learning-based blood pressure estimation will depend on the establishment of trust among patients, clinicians, and regulators. Trust is built not only by low mean absolute error but also by transparent reporting of limitations, consistent behavior under noise, and accountability when failures occur. A system that occasionally produces erroneous but confident blood pressure values can erode trust more quickly than a system that refuses to measure under uncertain conditions. Therefore, the user interface should communicate reliability clearly, using qualitative indicators such as low, moderate, or high confidence rather than a raw probability that may be misinterpreted. This is especially important for home users who do not have clinical training. The design of alerts and trend displays should be informed by human factors research, ensuring that critical hypertension events are not buried among low-priority notifications.

In academic research, the field must move beyond benchmark reporting to include system-level evaluations. Papers should report not only error metrics on clean segments but also performance on noisy segments, missing data rates, and subgroup analyses. Open datasets should document the conditions under which data were collected, including sensor placement, activity, skin tone distribution, and reference device type. This level of transparency is necessary for reproducibility and for assessing generalizability. The machine learning community has developed tools for dataset documentation that could be adopted by the physiological sensing community [12]. Without such documentation, it is impossible to compare models fairly or to predict their behavior in new populations. The governance of data and models should be integrated into the research workflow rather than treated as an afterthought.

The path from noisy photoplethysmography to reliable continuous blood pressure estimation is long and requires sustained interdisciplinary effort. Signal processing researchers must design adaptive filters that preserve hemodynamic information while suppressing artifact. Machine learning practitioners must move beyond in-distribution accuracy to consider robustness, fairness, and uncertainty. Clinicians must articulate the scenarios in which

continuous cuffless monitoring adds real value, such as nocturnal hypertension, medication titration, or perioperative care. Health system administrators must invest in the data infrastructure and governance needed to monitor these systems over time. Regulators must create adaptive frameworks that balance innovation with patient safety. Only through such coordinated action can the promise of continuous blood pressure monitoring be realized without repeating the pattern of algorithmic bias that has affected other medical sensing technologies. The goal is not to replace the cuff but to extend its reach, providing a continuous, low-burden complement that is trustworthy for all users.

References

1. Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), R1–R39. <https://doi.org/10.1088/0967-3334/28/3/R01>
2. Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. *Current Cardiology Reviews*, 8(1), 14–25. <https://doi.org/10.2174/157340312801215782>
3. O'Brien, E., Atkins, N., Stergiou, G., Karpettas, N., Parati, G., Asmar, R., ... & Shennan, A. (2010). European Society of Hypertension International Protocol revision 2010 for the validation of blood pressure measuring devices in adults. *Blood Pressure Monitoring*, 15(1), 23–38. <https://doi.org/10.1097/MBP.0b013e3283360e98>
4. Stergiou, G. S., Alpert, B. S., Mieke, S., Asmar, R., Atkins, N., O'Brien, E., ... & Shennan, A. (2018). A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) universal standard (ISO 81060-2:2018) and the AAMI standard. *Hypertension*, 71(3), 368–374. <https://doi.org/10.1161/HYPERTENSIONAHA.117.10237>
5. Kurylyak, Y., Lamonaca, F., & Grimaldi, D. (2013). A neural network-based method for continuous blood pressure estimation from a PPG signal. In 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) (pp. 280–283). IEEE. <https://doi.org/10.1109/I2MTC.2013.6555424>
6. Slapničar, G., Mlakar, N., & Luštrek, M. (2019). Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. *Sensors*, 19(15), 3420. <https://doi.org/10.3390/s19153420>
7. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
8. Mukkamala, R., Hahn, J. O., Inan, O. T., Mestha, L. K., Kim, C. S., Töreyn, H., & Kyal, S. (2015). Toward ubiquitous blood pressure monitoring via pulse transit time: Theory and practice. *IEEE Transactions on Biomedical Engineering*, 62(8), 1879–1901. <https://doi.org/10.1109/TBME.2015.2441951>
9. Zheng, X., Dwyer, V. M., Barrett, L. A., Derakhshani, M., & Hu, S. (2022). Adaptive notch-filtration to effectively recover photoplethysmographic signals during physical activity. *Biomedical Signal Processing and Control*, 72, 103303.
10. Mousavi, S. S., Firouzmand, M., Charmi, M., Hemmati, M., Moghadam, M., & Ghorbani, M. (2019). Blood pressure estimation from appropriate and inappropriate PPG signals

- using a whole-based method. *Biomedical Signal Processing and Control*, 47, 196–206.
<https://doi.org/10.1016/j.bspc.2018.08.022>
11. Esmaelpoor, J., Moradi, M. H., & Kadkhodamohammadi, A. (2022). A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals. *Computers in Biology and Medicine*, 141, 105114.
<https://doi.org/10.1016/j.combiomed.2021.105114>
 12. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92. <https://doi.org/10.1145/3458723>
 13. Institute of Electrical and Electronics Engineers. (2014). IEEE standard for wearable cuffless blood pressure measuring devices (IEEE Std 1708-2014). IEEE.
<https://doi.org/10.1109/IEEESTD.2014.6882122>
 14. Bent, B., Goldstein, B. A., Kibbe, W. A., & Dunn, J. P. (2020). Investigating sources of inaccuracy in wearable optical heart rate sensors. *npj Digital Medicine*, 3, 18.
<https://doi.org/10.1038/s41746-020-0226-6>
 15. Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *New England Journal of Medicine*, 378(11), 981–983. <https://doi.org/10.1056/NEJMp1714229>
 16. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). <https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
 17. Watanabe, N., Bando, Y. K., Kawachi, T., Yamakita, H., Futatsuyama, K., Honda, Y., ... & Kondo, T. (2020). Development and validation of a novel cuff-less blood pressure monitoring device. *JACC: Basic to Translational Science*, 5(9), 797–808.
<https://doi.org/10.1016/j.jacbts.2020.06.009>
 18. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
 19. Castaneda, D., Esparza, A., Ghamari, M., Soltanpur, C., & Nazeran, H. (2018). A review on wearable photoplethysmography sensors and their potential future applications in health care. *International Journal of Biosensors & Bioelectronics*, 4(4), 195–202.
<https://doi.org/10.15406/ijbsbe.2018.04.00125>
 20. Solà, J., & Delgado-Gonzalo, R. (2019). *The Handbook of Cuffless Blood Pressure Monitoring: A Practical Guide for Clinicians, Researchers, and Engineers*. Springer.
<https://doi.org/10.1007/978-3-030-24701-0>
 21. Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 25(1), 37–43. <https://doi.org/10.1038/s41591-018-0272-7>
 22. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMr1814259>
 23. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3, 119. <https://doi.org/10.1038/s41746-020-00323-1>