Dean&Francis
ISSN 2959-409X

A Review of Consumer-Grade Wrist-Worn Wearables for AI-Based Health Monitoring and Early Warning

Mingbo Wang[1, *]

[1]RCF Experimental School, Beijing, China
*Corresponding author: l1227530921@outlook.com

Abstract:

Consumer-grade wrist-worn wearables have recently arisen as a promising tool for daily life health detection; however, their accuracy and bias remain under debate. In this review, the author discussed the latest development of consumer-grade wrist-worn wearables in detecting OSA (obstructive sleep apnea) and AF (atrial fibrillation), by using PPG (Photoplethysmography), SpO_2 (oxyhemoglobin saturation), and IMU (inertial measurement units). Based on these two illnesses, the author evaluates how skin color may influence the OSA and AF detection. The author pointed out the findings for OSA and AF detection in different sensors and approaches. For example, in OSA, the author analyzes the method of detecting changes in SpO_2 and changes in posture by using the IMU. In AF detection, the author analyzes the two-step approach of AF capturing, analyzes the statistical data for On-demand ECG confirmation, and studies experiments from Apple Inc. and Huawei Inc. Based on that, the author shows three future perspectives, such as improvements that could be made in the device and users, and the supervision department, then summarizes the entire text.

Keywords: Wearables; Artificial Intelligence; Obstructive Sleep Apnea; Atrial Fibrillation; Skin-tone bias.

1. Introduction

Over the past decade, consumer-grade wearables such as smartwatches and smart bands transformed from a step-counting tool to a health detection platform. Compared with traditional medical-grade monitoring equipment, their key advantages are high accessibility, easy wearability, and the ability to support long-term daily use: Users can passively accumulate a large amount of data in real-life situations, thereby increasing the probability of detecting intermittent or nocturnal health issues. Traditional diagnostic approaches, for example, a 12-Lead electrocardiogram (ECG), provide only a short-term snapshot of the disease within medical institutions, making it hard to capture this momentary event.

Sensors in a wrist-wearable system consist of: PPG (Photoplethysmography), responsible for pulse mon-

itoring and can further deduce the heart rate, irregularity of rhythm, and SpO_2 (oxyhemoglobin saturation) trend; Single-lead ECG (Electrocardiography), responsible for an on-demand 30-second strip for rapid confirmation of suspected rhythm events; IMU (Inertial measurement units) is responsible for posture capturing and subtle cardio-respiratory micromotions; Skin temperature and environmental sensors provide context for sleep and illness trends.

Across these signals, recent studies are following a similar engineering pipeline: Signal detection, cleaning, and quality control → Key evidence extraction → lightweight criteria / ML / simple deep models → personalized thresholds → Alarm to record ECG for confirmation. To meet privacy and battery constraints, many systems run inference on the device side --- edge AI --- using model compression, such as quantization and distillation, to meet compute and energy costs. Since it delivers noise-resistant, personalized alerts and timely ECG trigger functions through lightweight device end inference, while also meeting the requirements of privacy protection and battery usage limitations.

However, opportunities come with some challenges. Noise and data quality are typical issues for optical signals, since optical signals for detection require extremely high accuracy. In this case, if signal accuracy is poor, the tolerance of the algorithm can only be increased, which leads to false positives or false negatives. For example, movement, sweat, ambient light, and even sleep position changes can cause inconclusive readings and false alerts. Also, bias is very crucial for PPG and pulse oximetry since they depend on optical reflection, which can create a systematic bias across user groups. Additionally, clinical integration is immature, so consumer-grade devices are best for daily monitoring and early warning, and then medical confirmation is required later, rather than being used as a substitute for medical equipment. Besides that, some new evaluations must be recorded, such as alert latency, incorrect rate, false alerts per given time, and battery impact.

In addition, previous reviews typically focus on a single condition (e.g., AF or OSA) or a single modality (e.g., PPG or IMU), and they barely consider day-to-day usability and diagnostic performance. Against this backdrop, this review mainly focuses on two representative, wearable-friendly applications: (1) OSA screening, emphasizing overnight use of PPG/SpO_2/IMU, and (2) AF and arrhythmia screening. A dedicated section then examines skin-tone effects on PPG and engineering mitigations, as a cross-cutting constraint for both applications. Together, these parts connect application performance and bias, offering a structured view of what consumer wrist-worn wearables can and cannot do, for everyday screening and early warning.

2. Current Evidence on OSA, AF, and PPG Bias

2.1 Obstructive sleep apnea (OSA)

Wearable sensors, including PPG for SpO_2, IMU for position detection, and even ECG combined with AI algorithms, those sensors were able to capture the occurrence of OSA. By distinguishing alternative indicators such as a reduction in SpO_2 or sudden pauses in chest motion, those wearable sensors could simulate PSG (Polysomnography) [1], [2]. Recent studies applied machine learning and deep learning to those indicators, allowing large-scale automated OSA screening during the night. In this section, this review is going to evaluate the representative studies that meet the criteria, mainly focusing on AI models, datasets, results, key findings, and limitations.

2.1.1 Method Based on PPG and SpO2

First, a representative approach is to utilize the PPG on a smartwatch or smart band to estimate AHI (Apnea–Hypopnea Index) or to detect OSA events. For instance, Papini et al. recorded 250 wrist PPG records, then trained a Deep Learning model to predict AHI[1]. Compared with PSG, their algorithm has good consistency, with an AHI estimation that has a relative index of 0.61, 3±10 events per hour on average[1]. Importantly, this algorithm can classify severity with medium correctness (Cohen's $\kappa \approx 0.51$) and shows high performance in screening. (In mild, moderate, and severe OSA, has approximately 0.84-0.86 ROC AUC) [1].

Other studies were focused on utilizing the SpO_2 derived from PPG in smartwatches (Galaxy Watch 4 and Apple Watch 7). Kim et al. compared PSG with a pulse oximeter built into consumer-grade smartwatches in 133 patients[2]. They found that SpO_2 based on PPG was able to indicate OSA[2]. For detecting any OSA (AHI $\geq$ 5) using the Galaxy Watch, its sensitivity and specificity were about 82.9% and 75.8%, respectively (AUC = 0.81)[2]. Using the lowest SpO_2 could improve the AUC = 0.85[2]. Apple Watch performs worse than the Galaxy Watch. By using the lowest SpO_2, sensitivity and specificity were 71.0% and 63.0%, respectively[2]. It is important to know that both tend to miss some severe events; In this study, the higher the OSA severity, the lower the diagnostic accuracy. For instance, diagnostic accuracy in mild OSA events has 70% accuracy, but in severe OSA events, it only has 63% diagnostic accuracy. This shows that although the algorithms were able to "flag" OSA via blood-oxygen patterns, sensor differences, and extreme OSA events may

ISSN 2959-409X

affect the trustworthiness in very severe cases.

More advanced PPG analysis enhances the performance. Wu et al. proposed using information-based similarity indices from wearable PPG pulse-rate data[3]. According to 92 bracelet records, their algorithm combines static and dynamic pattern matrices, achieving 84.7% accuracy, 76.7% sensitivity, and 89.6% specificity in OSA classification[3]. Those results indicate that ML can extract subtle cardiopulmonary features, therefore detect OSA with a specificity of approximately 90%.

2.1.2 Method Based on IMU

Another approach is to capture the pulse of breath by using motion sensors. Modern smartwatches have triaxial accelerometers and gyroscopes, which are sensitive enough to distinguish motion in the chest and wrist caused by the pulse of breath[4]. Hayano et al. show an IMU to quantify apneic episodes[4]. In their study of 122 adults taking PSG, they developed an algorithm to filter accelerometer/gyro signals between 0.13 to 0.7 Hz "breathing" band, and sense drops in signal amplitude which can last 10 to 90 seconds (assumed apneas)[4]. As shown in Fig. 1, the frequency of these detected signals is strongly correlated with the PSG AHI (r≈0.84 in the test set). For severe OSA events (AHI ≥ 30), the sensitivity and specificity are 90% and 88% respectively, as shown in Table 1. Given that only movement was measured, these performance levels are comparable to Type III home sleep tests[4].

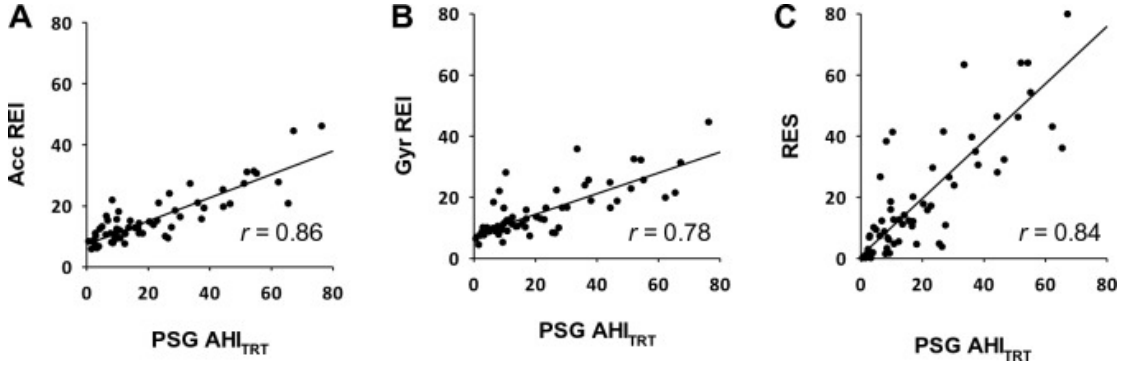

Fig. 1 Correlation between wearable-derived REI and PSG AHI. Adapted from Hayano et al.[4]. These strong correlations demonstrate that wrist-worn IMU sensors can reliably approximate clinical sleep study results.

Table 1. Confusion table for identification of severe SA (AHITST ≥ 30) by RES in the test group. [a]cutoff value of 20 was predetermined in the training group. Adapted from Hayano et al., 2024[4].

	AHI_{TST} < 30	AHI_{TST} ≥ 30	
RES < 20[a]	37	2	NPA = 95%
RES ≥ 20[a]	5	17	PPA = 77%
	Specificity = 88%	Sensitivity = 90%	Accuracy = 89%

This potential is not only in experiments, but also in the development of the technology industries. Apple Inc. in 2024 announced a Sleep Apnea Notification feature on Apple Watch, by using the accelerometer to trace the "Breathing Disturbances" during the night[5]. The algorithm is trained in a large multi-set dataset with PSG simultaneously. As shown in Table 2, in validation, the alert recognizes approximately 89% of severe OSA cases (sensitivity = 89.1% for AHI ≥ 30), and maintains 98.5% specificity for individuals who have no significant apnea[5]. However, for moderate OSA at this high specificity, the sensitivity is 43.4%, which is relatively low[5]. This shows that the algorithm on wearables must find a balance between high sensitivity (capture most cases) and high specificity (minimize false positives), depending on the usage[5]. Apple tends towards the latter, reflecting the prudent attitude towards consumer-oriented notifications[5].

Overall, these studies show that wrist-worn wearables can detect moderate and severe OSA with high accuracy. Many studies have 0.80-0.95 AUC in distinguishing OSA and healthy sleepers[1], [4]. At the clinical threshold of AHI ≥ 15 (moderate OSA), reported sensitivity has a range from 75% to over 90%, and specificity has a range from 70% to 90%, depending on different algorithms and sensors[2], [3].

A systematic review and meta-analysis in 2024 of 38 studies agreed that "wearable AI shows potential" for OSA identification. Overall sensitivity is about 93.8%, but the overall specificity is poor, about 75.2%[6]. In fact, this means that in the current stage, wearable algorithms tend to "over-mark" OSA; while capturing most of the real cases, some false positives are sacrificed[6]. Therefore, many authors recommend using wearable detections as an early warning rather than an independent diagnostic tool[6].

Table 2. Primary Endpoints: Sensitivity and Specificity. Reproduced from *Apple Inc.* **[5]**.

Algorithm Result	Value	Two-Sided 95% Confidence Interval
Sensitivity for moderate category	89/205 (43.4%)	(36.5%, 50.5%)
Sensitivity for severe category	164/184 (89.1%)	(83.7%, 93.2%)
Weighted overall sensitivity	66.3%	(62.2%, 70.3%)
Specificity for normal category	543/543 (100.0%)	(99.3%, 100.0%)
Specificity for mild category	315/346 (91.0%)	(87.5%, 93.8%)
Weighted overall specificity	98.5%	(98.0%, 99.0%)

Note: Weighted average sensitivity = (1/2)(moderate sensitivity) + (1/2)(severe sensitivity)
Weighted average specificity = (5/6)(normal specificity) + (1/6)(mild specificity)

2.1.3 Limitations

Although the result shows great potential, wearables are now facing many challenges. Data quality and noise are the main problems. For example, motion artifacts, poor sensor contact, or even low perfusion can degrade PPG / SpO_2 accuracy, leading to missing and false positives[2], [4]. In some cases, the wearables show bad performance, such as a change in sleep posture or potential cardiovascular diseases, etc., which could break the dependence of the original algorithm, such as heart rate and SpO_2[4].
Additionally, battery and computing power limit the algorithm in complexity. Some simple algorithms could omit details, but sophisticated algorithms must be highly optimized for running on the device side[1].
Finally, personalization is also a challenge since the fixed threshold does not apply to everyone. Some researchers advise using a self-learning strategy to adjust the algorithm to reduce individual differences[1]. Besides that, the supervision department is still conducting tests on various devices to ensure compliance with the equipment[2].

2.2 Atrial fibrillation (AF)

2.2.1 Two-Step Detection Approach in Wrist-worn Devices

Typically, modern smartwatches and bands use a two-step strategy to detect AF. First, the PPG sensors passively monitor the pulse for irregularities that could indicate the occurrence of AF. Once the algorithm detects a continuous arrhythmia, it alerts to inform the user. Second, the ECG confirmation is required to determine the AF. This method makes use of the convenience and wide time coverage of PPG, and the accuracy and high specificity of ECG[7]. By using PPG, only the suspected AF user may require an ECG confirmation, which leads to low false-positive and unnecessary results in healthy people.

2.2.2 PPG Screening for Irregular Pulse

By analyzing pulse interval data, PPG-based AF screening may detect irregular heart rhythm patterns. For example, in the Apple Heart Study, the irregular pulse tachograms were checked by the PPG algorithm of the watch, and if multiple irregular pulse was detected, the alarm was triggered[7]. Similarly, A large-scale study conducted by Fitbit uses an algorithm, which requires a continuous 30-minute irregular pulse under a resting situation to trigger the alarm[8]. Those strict standards ensure the temporary artifact and occasional premature contraction. In the real-world data show that the alarm rate was low, only 0.5-1% of users per observation period, which is reassuring in avoiding over-notification to healthy users, therefore indicating that PPG screening is reasonably selective[7], [8].

2.2.3 On-Demand Single-Lead ECG confirmation

While the device detects a possible AF, the user can initiate an ECG record, such as putting a finger on the electrode, and then the wearable's analysis follows up with a 30-second ECG to determine if AF occurs or not. This step is crucial for diagnostic accuracy since ECG is more accurate than PPG. Studies have found that ECG records have very high sensitivity and specificity to distinguish AF, normally in the range of 90-95%, compared with 12-Lead ECG[9]. But a big challenge is the inconclusive ECG result, which the algorithm cannot classify. A study of multiple devices found that about 15-20% first-time ECG record was marked as "inconclusive"[9]. For

ISSN 2959-409X

example, Apple Watch 6 has about 18% results that were "inconclusive", Fitbit Sense has about 21% results that were "inconclusive"[9]. In most cases, those inconclusive tracings could be interpreted by doctors, which indicates that the major reasons were noise, artifacts, and algorithm constraints. Usually, doing another ECG immediately could fix the problem[9]. Therefore, improvements in the algorithm and software are needed for reducing inconclusive rates, without sacrificing sensitivity[9]. It is worth noting that the built-in ECG allows the user to capture the heart rhythm immediately after an alert, which is a major advantage compared with the old confirmatory method, which had a longer delay in capturing the rhythm.

2.2.4 Recent Studies

From 2019 to 2025, multiple studies in real-world settings will evaluate this two-step approach.

2.2.4 .1 Apple Heart Study

This study (N = 419,000) showed the safety and value of the Apple Watch. Only 0.52% of candidates received the AF notifications, which shows a low probability of false positives[7]. Significantly, the PPV (positive predictive value) for the AF alert is 84%, which implies that most of the alerts were true[7]. The method they used for confirmation is to deliver some ECG patches to users. However, the confirmation of the AF has a delay from days to two weeks. During the delay, some short AF have already disappeared. This explained that only one-third of the AF was recaptured and highlighted the importance of instant confirmation[7].

2.2.4 .2 Huawei Heart Study

Large-scale AF screening research (N > 200,000) was conducted by using PPG in Huawei wearables and then confirmed in medical agents[10]. The result indicates that the PPV for the AF is 92%[10]. This is due to frequent sampling, which records once every ten minutes. But due to a lack of a built-in ECG, the delay also exists[10].

As summarized in Table 3, these studies indicate that a two-step approach can achieve high PPV, normally higher than 80%, even higher than 90%, which is an important index of acceptance[7], [8], [10]. Other than that, the inconclusive rate is also an important index, which depends on the environment and devices[9]. In the latest devices and algorithms, the inconclusive rate is less than 10%, but some independent test reports have reported a 15% inconclusive rate[9].

Table 3. Key Performance Metrics from Representative Studies (2020–2025)

Study (Year)	Device(s)	Population	PPV of PPG Alert	Inconclusive ECG Rate	Confirmation Latency
Apple Heart Study (2019) [7]	Apple Watch (Series 1–3, PPG only)	419,000 app users	84% for irregular pulse alert	n/a (no on-watch ECG in study)	~14 days (ECG patch sent)
Huawei Heart Study (2020) [10]	Huawei Watch (PPG)	187,000 telecom users	~92% for PPG alert	n/a (ECG not built-in)	Days (clinic visit/ Holter)

Notes: PPV = positive predictive value; PPG = photoplethysmography; ECG = electrocardiogram.

2.3 Skin-Tone and PPG Bias

2.3.1 Mechanism

PPG in wearables uses a light-emitting diode to illuminate the skin, and a photodetector to receive the reflected light; hence, this system could capture the volume change due to a pulse. Other than hemoglobin, melanin in skin also receives or scatters the visible light, causing a change in light intensity; this effect is even more obvious for green/red light[11]. This light mechanism indicates that dark-complexioned users are more likely to experience low-perfusion like PPG and low amplitude, therefore increasing the sensitivity to noise and artifact in the algorithm[11].

2.3.2 Evidence

Research on medical-grade PPG first proposed the systematic bias in skin color. Dark-skinned inpatients have 3 times higher Occult hypoxemia (SpO_2 > 92% and arterial blood SaO_2 < 88%) amounts compared with light-skinned inpatients, indicating that the results of the optical approach in dark-complexioned patients will be overestimated[12]. This bias could similarly influence the wearables based on PPG / SpO_2, such as OSA screening and AF screening based on an irregular time interval. In dark-complexioned people, a low PPG range and higher noise are easier to trigger "inconclusive", or lower the probability of reaching the algorithm threshold, showing an increase in false positives, and fewer true readings[12].

2.3.3 Influences on OSA and AF

For OSA screening, which is based on SpO_2 fluctuation

and desaturation events, the influence could result in an underestimation of the disease. For example, if a true desaturation of 4% misinterpreted as 2%, then the respiratory event count could be systematically underestimated, and cause a delay for moderate patients[12]. AF screening, which is based on a two-step approach, has a major in step one[12]. Small range and susceptible to interference, PPG in dark-complexioned users is harder to meet the condition of valid and continuous irregular readings, causing the lower sensitivity and higher inconclusive rate; therefore, instant ECG confirmation is important[12]. In the study, multi-wavelength, SQI gating, personalized threshold, and proper wearing can significantly reduce the color bias. However, in the reporting and validation process, it is necessary to have a clear hierarchical structure to highlight the skin color bias, and during the clinical interpretation, higher vigilance should be maintained for different individuals[11], [12].

3. Future Discussion

Consumer-grade wrist-worn wearables already have two mature approaches for health detection. The first one is OSA screening, and the other one is two-step PPG screening. Future improvement and optimization should focus on the actual usability in the real world rather than some indication, such as accuracy or inconclusive rate. Below, the author mentioned several enhancing measures.

Complete the loop from the alarm to the result. Most users saw the alarm but failed to do ECG confirmation, or too late to do so. For this situation, the alarm needs to vibrate and include a notification sound. If the user still does not respond, it will be sent to the family members' mobile phone, or a dedicated person will make a phone call to remind them. Besides that, educational guidance to users is also important. For the first time of use, it is crucial to inform the user that if an alarm occurs, an ECG confirmation must be conducted.

Pigmentation-related Bias. Dark-complexioned users are more likely to have a low signal range and a higher inconclusive rate, causing sensitivity decrease and a delay. Future improvement could be multi-wavelength and adaptive reflective index control in the sensing end; At the data processing end, perform weighted calculations based on the objective reflectivity; At the algorithm end, use multiple algorithms tailored for different skin tones, and the routing selection is made based on the actual situation of the user.

The future role of the supervision department is to define a risk-based framework for consumer-grade wrist-worn wearables. This could require:

a. A minimum reporting set which includes the inconclusive rate, false alerts per day, alert-to-ECG latency, and battery usage.

b. To conduct a mandatory 12-lead ECG experiment for key research

c. Considering the different skin tones of users, and adapting and adjusting the algorithm accordingly based on the skin color

4. Conclusion

In this review, by conducting a systematic summary and analysis, the author draws the following conclusions.

(1) Wrist-worn sensors such as reflective PPG, SpO_2, and IMU have demonstrated a high effectiveness in detecting and capturing OSA and AF. Particularly, PPG-based AF detection achieved high accuracy and a low false positive rate, while the ECG confirmation is performed promptly.

(2) In several studies, a common engineering pipeline has already emerged, including signal sampling, quality control, feature engineering, a lightweight model, an adaptive algorithm, and a personalized threshold. Nowadays, many systems make use of model compression technology to allow models to be inferred at the device end. This could reduce battery cost and ensure data privacy.

(3) However, there do exist some challenges. Optical signals are easily interfered with by artifacts, sweat, and low perfusion. Reliability and diagnostic accuracy of the signals are influenced by skin colors, so that dark-complexioned users have a smaller range of signals and a higher inconclusive rate.

Despite these findings, this review is limited by the diversity of datasets, limited multi-modal sensor integration, and uneven representation of skin tone groups. Future research may benefit from a larger dataset, more diverse study populations, and higher-quality data. This review integrates the existing detection and screening methods for OSA and AF and summarizes and explains the problems faced by wearables in real life. Hopefully, this review will provide a useful reference for future OSA and AF detection for consumer-grade wrist-worn wearables.

References

[1] G. B. Papini, P. Fonseca, M. M. van Gilst, J. W. M. Bergmans, R. Vullings, and S. Overeem, "Wearable monitoring of sleep-disordered breathing: estimation of the apnea–hypopnea index using wrist-worn reflective photoplethysmography," Sci Rep, vol. 10, no. 1, p. 13512, Aug. 2020, doi: 10.1038/s41598-020-69935-7.

[2] M. W. Kim, S. H. Park, and M. S. Choi, "Diagnostic Performance of Photoplethysmography-Based Smartwatch for Obstructive Sleep Apnea," J Rhinol, vol. 29, no. 3, pp. 155–162,

ISSN 2959-409X

Nov. 2022, doi: 10.18787/jr.2022.00424.
[3] S. Wu, M. Chen, K. Wei, and G. Liu, "Sleep apnea screening based on Photoplethysmography data from wearable bracelets using an information-based similarity approach," Computer Methods and Programs in Biomedicine, vol. 211, p. 106442, Nov. 2021, doi: 10.1016/j.cmpb.2021.106442.
[4] J. Hayano, M. Adachi, F. Sasaki, and E. Yuda, "Quantitative detection of sleep apnea in adults using inertial measurement unit embedded in wristwatch wearable devices," Sci Rep, vol. 14, no. 1, p. 4050, Feb. 2024, doi: 10.1038/s41598-024-54817-z.
[5] "Sleep_Apnea_Notifications_On_Apple_Watch_September_2024." Accessed: Aug. 25, 2025. [Online]. Available: https://images.apple.com/mideast/health/b/pdf/sleep-apnea/Sleep_Apnea_Notifications_on_Apple_Watch_September_2024.pdf
[6] A. Abd-alrazaq et al., "Detection of Sleep Apnea Using Wearable AI: Systematic Review and Meta-Analysis," J Med Internet Res, vol. 26, p. e58187, Sept. 2024, doi: 10.2196/58187.
[7] M. V. Perez et al., "Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation," N Engl J Med, vol. 381, no. 20, pp. 1909–1917, Nov. 2019, doi: 10.1056/NEJMoa1901183.
[8] S. A. Lubitz et al., "Detection of Atrial Fibrillation in a Large Population Using Wearable Devices: The Fitbit Heart Study," Circulation, vol. 146, no. 19, pp. 1415–1424, Nov. 2022, doi: 10.1161/CIRCULATIONAHA.122.060291.
[9] D. Mannhart et al., "Clinical Validation of 5 Direct-to-Consumer Wearable Smart Devices to Detect Atrial Fibrillation: BASEL Wearable Study," JACC Clin Electrophysiol, vol. 9, no. 2, pp. 232–242, Feb. 2023, doi: 10.1016/j.jacep.2022.09.011.
[10] Y. Guo, H. Wang, H. Zhang, Y. Chen, and G. Y. H. Lip, "Population-Based Screening or Targeted Screening Based on Initial Clinical Risk Assessment for Atrial Fibrillation: A Report from the Huawei Heart Study," Journal of Clinical Medicine, vol. 9, no. 5, Art. no. 5, May 2020, doi: 10.3390/jcm9051493.
[11] R. Al-Halawani, M. Qassem, and P. A. Kyriacou, "Monte Carlo Simulation of the Effect of Melanin Concentration on Light–Tissue Interactions in Reflectance Pulse Oximetry," Sensors, vol. 25, no. 2, p. 559, Jan. 2025, doi: 10.3390/s25020559.
[12] M. W. Sjoding, R. P. Dickson, T. J. Iwashyna, S. E. Gay, and T. S. Valley, "Racial Bias in Pulse Oximetry Measurement," N Engl J Med, vol. 383, no. 25, pp. 2477–2478, Dec. 2020, doi: 10.1056/NEJMc2029240.