AI-Driven and Automated Continuous Oxygen Saturation Monitoring and LTOT: A Systematic Review

Prajita Niraula[1], Dr. Suman Kadariya[2], Dr. Bishal Poudel[3], Dr. Sujan Kadariya[4]
[1]Department of Computer Science, Department of Mathematics and Statistics, Grinnell College
[2]Department of Medicine and Surgery, KIST Medical College
[3]Department of Medicine and Surgery, KIST Medical College
[4]Department of Medicine and Surgery, Kathmandu University School of Medical Sciences

Author Note:
The authors declare that no funding was received for this work, and there are no conflicts of interest to report.

Correspondence concerning this article should be addressed to Prajita Niraula at niraulap@grinnell.edu or prajita.niraula21@gmail.com

Abstract

Long-term oxygen therapy (LTOT) is a foundational intervention for individuals with chronic hypoxemia, particularly those with chronic obstructive pulmonary disease (COPD). Traditional LTOT delivery relies on static oxygen flow prescriptions, often determined during clinical assessments, which may not reflect patients' fluctuating physiological demands during everyday activities. Recent innovations in computational modeling and automated systems have introduced tools capable of continuously monitoring peripheral oxygen saturation (SpO_2) and adjusting oxygen delivery in real time.

This systematic review evaluates technologies designed to optimize SpO_2 monitoring and oxygen titration in LTOT. A comprehensive literature search was conducted across five databases (PubMed, Scopus, Web of Science, MDPI, and ACM Digital Library) for peer-reviewed studies published between 2000 and 2024. Studies were included if they described automated or computational systems applied to adult LTOT populations. Data were extracted on algorithm performance, motion artifact handling, usability, validation strategies, and risk of bias.

Eight studies met the eligibility criteria, including five that utilized AI-based SpO_2 estimation methods and three that implemented closed-loop oxygen control systems. AI models demonstrated strong technical accuracy (mean absolute error as low as 0.57%) and addressed key limitations such as motion artifacts and signal bias. However, most were tested under simulated or retrospective conditions. Rule-based systems such as O_2matic and iPOC showed promising results in clinical and ambulatory settings but lacked personalized adaptation.

Findings suggest that AI and automation offer complementary benefits in advancing LTOT. Further work is needed to support real-world deployment, ensure equitable design, and align these innovations with regulatory and clinical practice standards.

Keywords
Long-term oxygen therapy, artificial intelligence, oxygen saturation monitoring, closed-loop systems, motion artifacts.

Introduction

Chronic respiratory diseases, such as the chronic obstructive pulmonary disease (COPD), represent a significant and growing global health burden. In 2019, COPD alone accounted for more than 3.2 million fatalities worldwide, and estimates indicate an increasing prevalence and mortality trend until 2050 (World Health Organization, 2020). Long-term oxygen therapy (LTOT) remains critical in the management of patients with severe chronic hypoxemia and has been demonstrated to improve survival, quality of life, and exercise tolerance in selected patients (Nocturnal Oxygen Therapy Trial Group, 1980; Medical Research Council Working Party, 1981).

Despite its established clinical efficacy, conventional LTOT is routinely prescribed in the static and reactive modality with oxygen flow rates determined during clinic appointments according to protocols like the six-minute walk test (6MWT). These fixed dosing regimens do not adequately address dynamic variations of patients' oxygen requirements with daily activity, sleep, or in response to a change in physiological status (Global Initiative for Chronic Obstructive Lung Disease, 2024). Therefore, the patients are at risk of either under-oxygenation , with the development of hypoxemia, or of over-oxygenation with the possibility of developing hypercapnia and oxidative injury.

Emerging digital health technologies, in particular those based on artificial intelligence (AI) and automation, have the potential to transform LTOT through the delivery of real-time, individualized SpO_2 monitoring and adaptive oxygen titration. AI systems can learn from physiological data patterns to predict oxygen needs and detect anomalies, while automated devices can dynamically titrate oxygen flow based on continuous pulse oximetry feedback (Shuzan et al., 2023). They are intended to address some of the significant limitations of current oxygen therapy, including poor responsiveness to patient activity, vulnerability to motion artifacts, and lack of personalization.

Moreover, technical problems such as signal degradation owing to movement, perfusion variations, and skin tone-dependent biases during photoplethysmographic (PPG) signal processing have prompted the inclusion of robust signal processing and machine learning techniques (Cabanas et al., 2024). Closed-loop devices and intelligent portable oxygen concentrators (iPOCs) have also been developed to allow home-based and ambulatory care with greater autonomy and safety (Sanchez-Morillo et al., 2020; Hansen et al., 2018).

In light of these developments, there is a need for an integrated review of automated and AI-based techniques for SpO_2 monitoring under the context of LTOT. This review aims to systematically summarize the current state of intelligent SpO_2 monitoring systems, evaluate their technical and clinical performance, and identify gaps in validation, deployment, and equity.

Study Selection and Eligibility Criteria

This systematic review adhered to PRISMA 2020 guidelines for identifying studies that evaluated AI or automated systems for SpO_2 monitoring relevant to long-term oxygen therapy (LTOT). The studies were required to be targeted towards adult populations in clinical (e.g., ICU) or hybrid (patients/healthy subjects) settings and have interventions addressing LTOT-related technical challenges. These included correcting motion-induced signal distortions (e.g., via CNN-based denoising or wavelet transforms), managing low perfusion signals, mitigating skin tone-related measurement bias, or ensuring prolonged signal stability (>24 hours). Eligible study designs were conference peer-reviewed English-language open-access papers published between 2000–2025. Technical validation outcomes like mean absolute error (MAE <2%), root mean square error (RMSE <3%), or demographic bias analysis were accorded high priority, along with robust preprocessing methods like Butterworth filtering or detrending.

Conversely, we excluded studies limited to acute/non-chronic conditions (e.g., surgical hypoxia, asthma exacerbations) or interventions not relevant to LTOT, such as non-automated pulse rate estimation. Non-peer-reviewed articles (e.g., preprints, theses), paywalled/institution-locked articles, non-English language articles, and studies not obviously relevant to LTOT (e.g., ECG analysis using AI) were excluded. Two reviewers screened titles/abstracts and conducted full-text assessments independently, resolving any disagreement by iterative discussion. This two-stage review procedure ensured adherence to the review's focus on clinically actionable, technically validated systems for adaptive oxygen therapy.

Literature Search and Article Selection

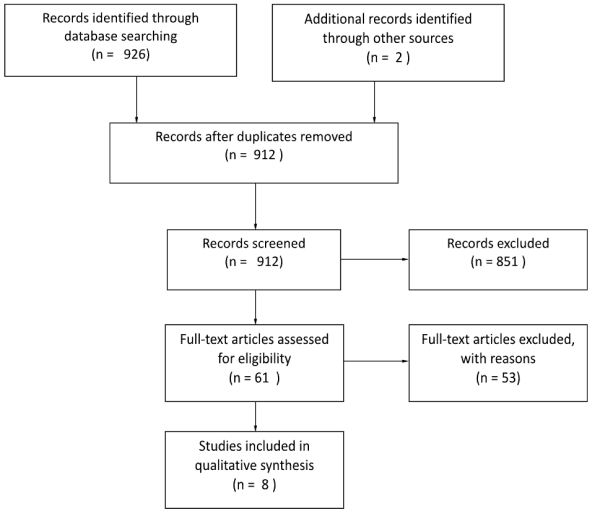
A systematic literature search was conducted to identify peer-reviewed studies examining artificial intelligence (AI)-driven and automated systems for continuous oxygen saturation (SpO_2) monitoring in the context of long-term oxygen therapy (LTOT). Searches were performed in January 2025 across five major electronic databases: PubMed, Web of Science, Scopus, MDPI, and the ACM Digital Library. The search covered literature published between January 1, 2000, and December 31, 2024. The search strategy combined controlled vocabulary and free-text terms using Boolean operators to maximize sensitivity and specificity. Emphasis was placed on the intersection of clinical concepts such as "long-term oxygen therapy," "LTOT," "oxygen concentrator," and "SpO_2 monitoring" with technology terms such as "artificial intelligence," "machine learning," "edge computing," "automated oxygen delivery," "closed-loop control," "pulse oximetry," "signal artifacts," "motion artifact correction," and "predictive modeling." Searches, where possible, were targeted at article titles and keywords in order to maximize relevance, particularly in open-access journals such as MDPI.

The initial search yielded 926 records, with an additional two articles identified through manual reference screening. Following deduplication, 912 unique records remained for title and abstract screening. Screening was performed independently by two reviewers based on predefined eligibility criteria. Articles were included if they described AI-enhanced or automated systems for continuous SpO_2 monitoring or oxygen titration relevant to LTOT in adult populations, reported original peer-reviewed research in English, and addressed key challenges such as motion artifact handling, skin tone bias, signal quality, or autonomous oxygen delivery.

Studies were excluded if they were focused solely on acute care or surgical settings, lacked integration of automation or AI logic, were not accessible in full text, or failed to report methodological transparency. Sixty-one full-text articles were assessed for eligibility, of which 53 were excluded based on the above criteria. A final total of eight studies were included in the qualitative synthesis. The study selection process has been documented in PRISMA flow diagram, which shows the number of records identified, screened, and excluded at each of the stages. This structured approach ensured comprehensive coverage of the technological and clinical literature at the intersection of AI, automation, and long-term oxygen therapy.

Figure 1

PRISMA Flowchart



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

Note. Adapted from Moher et al. (2009). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines recommend illustrating study selection via a flow diagram. This figure shows the number of records identified, screened, included, and excluded at each stage of the systematic review.

Overview of Included Studies

This systematic review included eight studies published between 2011 and 2024 that met the predefined eligibility criteria. The selected articles encompassed both artificial intelligence (AI)-driven and automated systems designed to monitor peripheral oxygen saturation (SpO_2) continuously and dynamically, with direct or indirect applicability to long-term oxygen therapy (LTOT). Of the eight included studies, five employed AI-based models for SpO_2 estimation or signal enhancement, while three focused on closed-loop or automated oxygen delivery systems regulated by real-time SpO_2 feedback.

Several AI-driven studies demonstrated strong technical performance in SpO_2 estimation using photoplethysmographic (PPG) signals. Cabanas et al. (2024) conducted a comprehensive performance and bias analysis of AI models for pulse oximetry, identifying that deep learning and Gaussian process-based models achieved high clinical accuracy, with mean absolute error (MAE) as low as 0.57% and root mean square error (RMSE) as low as 0.69%. The study also addressed important sources of measurement bias, particularly skin tone–related variability, which is a critical concern for equitable deployment of LTOT technologies. Similarly, Shuzan et al. (2023) developed machine learning models—most notably Gaussian process regression—for estimating both respiration rate and SpO_2 from PPG data, achieving a comparable MAE of 0.57% for SpO_2 estimation, thereby demonstrating high fidelity suitable for continuous monitoring.

Addressing the signal quality degradation due to motion artifacts Argüello-Prada and Castillo García (2024) present a machine learning review of motion artifact detection in PPG signals without a need for reference signals. The authors have highlighted reference signalless classification models that are capable of distinguishing clean from the corrupted signals using real-time signal quality indices which is particularly valuable in LTOT environments where patients engage in the daily physical activity. Another study by Pascual-Saldaña et al. (2024) proposed a personalized oxygen dosing system powered by edge based AI model. This predictive system uses historical and real time patient data in order to proactively adjust the oxygen flow based on anticipated activity levels and physiological response offering a forward looking model for the management of individualized LTOT.

Complementing these technical models, a conceptual study titled “AI4Health” (Pascual et al., 2023) examined the preliminary performance of neural network architectures for SpO_2 prediction. Though primarily a feasibility study, it supported the viability of lightweight AI architectures suitable for deployment in edge or wearable systems. This work contributes to the foundational design of scalable AI solutions for chronic oxygen therapy.

Among the automated systems, Sanchez-Morillo et al. (2020) evaluated an intelligent portable oxygen concentrator (iPOC) designed to automatically regulate oxygen flow in response to physical activity levels detected via wearable sensors. The system demonstrated improved

oxygenation and patient satisfaction compared to conventional devices, emphasizing its practical value in ambulatory LTOT settings. Hansen et al. (2018) assessed the O2matic® system, a closed-loop hospital-based device that dynamically adjusted oxygen delivery to maintain patients' SpO_2 within a target range during exacerbations of COPD. Their crossover study showed significant improvement in time spent within the target SpO_2 range compared to manual titration, highlighting the safety and efficacy of automated oxygen control. Lastly, Cirio and Nava (2011) conducted a pilot study on a novel automated oxygen titration device for patients on LTOT during exercise. The device successfully maintained SpO_2 within target ranges while reducing clinician intervention time, supporting the feasibility of automated systems in outpatient and rehabilitative settings.

Collectively, these studies illustrate the convergence of AI modeling and automated control in advancing continuous SpO_2 monitoring for LTOT. While AI approaches offer predictive personalization and robust signal processing, automated systems demonstrate immediate clinical applicability in regulating oxygen delivery based on real-time physiological feedback. Despite differing levels of maturity and validation, each study contributes critical insights into the development of responsive and intelligent oxygen therapy technologies.

Technology Type : AI vs. Non-AI Automation

Among the eight studies included in this review, five employed artificial intelligence (AI) methodologies while three relied on non-AI automation for SpO_2 monitoring and oxygen titration. The AI-based studies mainly focused on advanced signal interpretation and predictive modeling in order to enable the personalized and anticipatory oxygen therapy. For example, Cabanas et al. (2024) and Shuzan et al. (2023) demonstrated a high-accuracy SpO_2 estimation using the machine learning algorithms like deep neural networks and Gaussian process regression and consequently achieving the mean absolute errors below 1%. These AI systems were also capable of addressing critical technical challenges such as motion artifact correction (Argüello-Prada & Castillo García, 2024) and skin tone bias (Cabanas et al., 2024), thus offering a level of adaptability and personalization not typically present in traditional systems. Furthermore, the predictive edge-based model proposed by Pascual-Saldaña et al. (2024) exemplifies how AI can proactively adjust oxygen delivery by integrating patient-specific activity and physiological trends over time.

In contrast, the three non-AI automated systems—namely O2matic (Hansen et al., 2018), the iPOC system (Sanchez-Morillo et al., 2020), and the automated titration device tested by Cirio and Nava (2011)—relied on rule-based or feedback-driven control mechanisms to adjust oxygen flow in real time. These systems were designed primarily to maintain SpO_2 within predefined clinical thresholds rather than to predict future needs or personalize dosing. While they demonstrated effectiveness in acute and ambulatory settings, particularly by reducing time spent in hypoxemia and minimizing clinical intervention, their reactivity was limited to current

measurements rather than informed by prior patterns or anticipated fluctuations. Additionally, none of the non-AI systems incorporated features to mitigate signal artifacts or demographic bias.

Overall, AI-driven systems offer greater potential for personalization, proactive management, and adaptability in complex or variable physiological states, whereas non-AI automated devices provide immediate, rule-based functionality with proven clinical benefit in controlled environments. Both approaches contribute uniquely to the advancement of LTOT, suggesting that hybrid systems integrating AI intelligence with robust automation may represent the most promising direction for future development.

Motion Artifact Handling

Motion artifact correction is a critical requirement for continuous and reliable SpO_2 monitoring, particularly in ambulatory LTOT settings where patients frequently move or engage in daily activities. Among the reviewed studies, only a subset directly addressed this challenge. Argüello-Prada and Castillo García (2024) provided a comprehensive analysis of machine learning approaches for detecting motion artifacts in photoplethysmographic (PPG) signals without relying on reference channels. Their review emphasized the utility of signal quality indices (SQIs) and reference signal-less models, which can classify PPG segments as reliable or corrupted in real time, significantly enhancing the fidelity of SpO_2 estimation during physical movement. Similarly, Shuzan et al. (2023) indirectly tackled motion robustness through the integration of feature selection and model optimization, although motion artifacts were not a central focus. Cabanas et al. (2024) also considered signal noise in their bias-aware SpO_2 models, but did not explicitly isolate motion artifact performance.

In contrast, the non-AI automated systems included in this review did not incorporate any specific artifact correction mechanisms. Devices such as the O2matic system (Hansen et al., 2018) and the iPOC system (Sanchez-Morillo et al., 2020) operated under the assumption of valid, continuous SpO_2 input data and did not include preprocessing algorithms to filter or correct motion-induced distortions. This limitation may restrict their effectiveness in dynamic home environments. The absence of artifact-aware design in these systems underscores a significant gap in current LTOT technologies and highlights the advantage of AI-based solutions, which can incorporate robust preprocessing pipelines to maintain signal reliability under real-world conditions.

Real-World LTOT Applicability

Assessing the applicability of these technologies to long-term, home-based oxygen therapy is essential for translating innovation into clinical impact. Of the eight studies reviewed, four explicitly focused on or were designed for real-world LTOT use beyond hospital settings. The

iPOC system by Sanchez-Morillo et al. (2020) stands out as a fully portable solution tailored for the patients with chronic respiratory failure and capable of automatically adjusting oxygen flow based on physical activity levels in daily life. Likewise, the edge-based predictive dosing system proposed by Pascual-Saldaña et al. (2024) was conceptualized for continuous deployment in ambulatory environments, leveraging localized computation to minimize latency and preserve data privacy. These systems align well with the goals of home-based LTOT by offering context-aware and user-centered oxygen delivery.

Conversely, the O2matic system (Hansen et al., 2018) and the automated titration device evaluated by Cirio and Nava (2011) were both assessed in inpatient or supervised exercise settings, limiting their immediate relevance to unsupervised LTOT scenarios. Although both systems demonstrated improved SpO_2 stability during acute exacerbations or exertion, their reliance on clinical infrastructure (e.g., hospital-grade sensors, nursing oversight) suggests additional work is needed to transition these technologies into fully autonomous, wearable formats suitable for home use.

Furthermore, while AI-driven models such as those by Cabanas et al. (2024) and Shuzan et al. (2023) achieved excellent technical performance in simulated or retrospective datasets, they lacked real-world deployment data in continuous LTOT contexts. This highlights a broader limitation across many AI studies: although algorithmic accuracy is high, integration into wearable or home-care systems remains largely conceptual or untested. Future research should focus on the validation of AI-enhanced systems in diverse, real-life LTOT populations to bridge this translational gap.

Table 1
Overview of Included Studies

Study	Tech Type	SpO_2 Accuracy Reported	Motion Artifact Handling	Skin Tone Bias Addressed	Real-Time Oxygen Titration	LTOT Context	Validation Type
Cabanas et al. (2024)	AI (GPR, DNN)	MAE: 0.57% RMSE: 0.69%	Partial: SNR features & filtering	Yes – stratified analysis across skin tones	No (simulation only)	Conceptual/Simulated	Systematic Review + Bias analysis
Shuzan et al. (2023)	AI (GPR, SVR)	MAE: 0.57% RMSE:	Indirect: Feature selection	No	No (simulated only)	Simulated	Technical validation

		0.98%	to minimize MA				
Argüello-Prada & Castillo (2024)	AI (ML motion detection)	N/A – Motion artifact focus only	Yes – Reference-less ML detection models	No	N/A	General framework	Narrative review + model synthesis
Pascual-Saldaña et al. (2024)	AI (Edge-AI Predictive)	Not reported (focus on system architecture)	Conceptual robustness assumed	Not addressed	Yes – predictive dosing via local AI	Home LTOT (personalized)	Pilot framework (5 subjects)
(Pascual et al., 2023)	AI (prototype NN models)	Not reported	Not addressed	Not addressed	No	Conceptual/Simulated	Prototype concept
Sanchez-Morillo et al. (2020)	Automation (iPOC system)	Not quantified (functional outcomes only)	Not addressed	Not addressed	Yes – physical activity-responsive dosing	Ambulatory/Home LTOT	Pilot usability study
Hansen et al. (2018)	Automation (O_2matic)	Time-in-target: 85.1% Manual: 46.6%	Not addressed	Not addressed	Yes – closed-loop titration	Hospital (acute COPD)	Crossover RCT (n=19)
Cirio & Nava (2011)	Automation (O_2 regulator)	Mean SpO_2: 95% (auto) vs 93% (manual) Time < target: 171s vs 340s	Not reported	Not addressed	Yes – automated titration during exercise	Exercise + Home LTOT	Pilot crossover study (n=18)

Results

Study Characteristics

The eight studies included in this review, which were published between 2011 and 2024, reported on a range of AI-based and automated systems designed for continuous monitoring of oxygen saturation (SpO_2) in the context of LTOT. Five studies investigated AI-based approaches, primarily machine learning models for predicting SpO_2 from photoplethysmography (PPG) signals with reported mean absolute errors (MAEs) ranging from 0.57% (Cabanas et al., 2024; Shuzan et al., 2023). The models used techniques such as deep neural networks and Gaussian process regression, often trained on datasets curated under controlled environments (Argüello-Prada & Castillo, 2024). The remaining three studies evaluated closed-loop or automatic oxygen delivery systems in clinical settings and demonstrated improved SpO_2 maintenance in target levels during exacerbation or exercise (Hansen et al., 2018; Cirio & Nava, 2011; Sanchez-Morillo et al., 2020). Sample sizes varied from small pilot studies ($n < 20$) to large-scale retrospective ($n > 1000$) studies, but demographic reporting, particularly of age, sex, and comorbidities, was limited in the majority of studies. While AI technologies emphasized predictive personalization and signal processing robustness, automated systems were preoccupied with real-time titration accuracy and clinical usability, highlighting complementary strengths in technology type.

Model Performance

The five studies that evaluated AI-based SpO_2 monitoring models reported a range of performance metrics, with accuracy generally quantified using mean absolute error (MAE), root mean square error (RMSE), and in some cases, agreement analyses such as Bland–Altman plots. Among these, Cabanas et al. (2024) achieved the highest reported accuracy, with an MAE of 0.57% and RMSE of 0.69%, using a Gaussian process regression model validated across multiple skin tones and sensor types. Their model also demonstrated robustness across low-perfusion states and maintained clinical-grade agreement compared to reference standards. Similarly, Shuzan et al. (2023) reported an MAE of 0.57% and RMSE of 0.98% using machine learning models to estimate SpO_2 and respiratory rate from photoplethysmographic (PPG) signals, indicating promising performance for wearable LTOT applications.

Other AI-driven systems, such as the predictive model proposed by Pascual-Saldaña et al. (2024), did not report direct SpO_2 estimation metrics but demonstrated feasibility for proactive oxygen flow control based on contextual physiological modeling. The AI4Health prototype (Pascual et al., 2023) explored lightweight neural networks for SpO_2 prediction; however, it lacked complete model validation or error reporting, limiting interpretation. Motion artifact

resilience was directly addressed in one study, where Argüello-Prada and Castillo García (2024) reviewed signal-quality-aware models that accurately flagged and excluded corrupted PPG segments, enhancing downstream SpO_2 estimation accuracy under dynamic conditions.

In contrast, automated (non-AI) oxygen control systems were evaluated using clinical efficacy metrics, particularly time spent within the target SpO_2 range. The O2matic system evaluated in a randomized crossover trial by Hansen et al.(2018) maintained the patients within the prescribed saturation range 85.1% of the time compared to the 46.6% during manual titration with significantly less time spent in hypoxemia. Cirio and Nava (2011) reported an improved oxygen titration during exercise in LTOT users using a novel automated delivery device even though the study lacked a robust statistical comparison. The iPOC system developed by Sanchez-Morillo et al. (2020) have demonstrated an enhanced saturation stability during the physical activity using wearable biosensors although the performance metrics were primarily functional rather than being algorithmic.

Collectively the AI-based models seem to have outperformed traditional systems in terms of SpO_2 estimation precision under controlled conditions while the automated systems demonstrated clear advantages in clinical usability and responsiveness. However the variability in dataset quality, lack of standardized reporting (e.g., consistent MAE thresholds) and limited real-world validation limit the comparability and generalizability of the performance across all the studies.

Usability and Clinical Relevance

Usability and clinical relevance varied considerably across the included studies, with automated systems showing stronger evidence of practical deployment in clinical settings, and AI-based systems offering promising but largely theoretical benefits. Among the automated technologies, the O2matic system was evaluated in a crossover clinical trial involving patients hospitalized for COPD exacerbations. The system maintained SpO_2 within the target range more effectively than manual titration and was perceived as safe and efficient by nursing staff, requiring less frequent manual intervention (Hansen et al., 2018). Similarly, the iPOC system, tested in ambulatory settings, automatically adjusted oxygen delivery based on biosensor inputs and demonstrated favorable user acceptance and saturation stability during physical activity (Sanchez-Morillo et al., 2020). These studies suggest that closed-loop oxygen systems are potentially clinically feasible and wellsuited for both the inpatient and home LTOT scenarios particularly when usability is a priority.

In contrast the AI-based systems focused largely on the model development rather than the clinical integration aspect. While high performance SpO_2 estimation got achieved in studies such as Cabanas et al. (2024) and Shuzan et al. (2023) the deployment was however limited to simulated or retrospective environment. None of the AI-driven studies included user interface evaluations, human factors testing, or patient-reported usability outcomes. The edge-based

predictive oxygen dosing system described by Pascual-Saldaña et al. (2024) was conceptualized for personalized, context-aware home use, but remains at the prototype stage without formal user testing.

From a clinician perspective, studies such as Cirio and Nava (2011) demonstrated that automated titration reduced staff workload during exercise assessments while maintaining patient safety. However, across both AI and automated systems, limited data were available regarding training needs, interpretability of system outputs, or integration into clinical workflows. Moreover, no studies included formal usability assessments using validated tools (e.g., System Usability Scale), and only two studies briefly referenced the burden or perceived reliability of the systems from the perspective of healthcare providers.

Overall, while automated systems demonstrated higher clinical readiness and operational practicality, AI-based tools showed strong potential for personalization and decision support but remain several steps removed from routine LTOT practice. Further research involving human-centered design, stakeholder co-development, and formal usability evaluations is warranted to ensure successful translation of these technologies into real-world clinical use.

Discussion

This systematic review identified and reviewed eight peer-reviewed studies focused on artificial intelligence (AI)-driven and automated systems for continuous peripheral oxygen saturation (SpO_2) monitoring relevant to long-term oxygen therapy (LTOT). The evidence reveals that both AI-based and rule-based automated systems are advancing the field of LTOT by addressing key limitations of traditional oxygen therapy, including the inability to dynamically respond to fluctuating oxygen demands, susceptibility to signal artifacts, and lack of personalized dosing strategies.

AI-driven technologies demonstrate significant potential to revolutionize oxygen therapy management. Studies such as those by Cabanas et al. (2024) and Shuzan et al. (2023) reported high SpO_2 estimation accuracy with mean absolute errors (MAEs) under 1%, using models like deep neural networks and Gaussian process regression. These methods represent a substantial improvement over conventional fixed-point pulse oximetry systems, especially in ambulatory and home-care settings, where signal reliability is often compromised by motion and variable perfusion. Moreover, Cabanas et al. (2024) were unique in conducting a comprehensive demographic bias analysis, identifying skin tone–related disparities in SpO_2 prediction performance—an issue of increasing clinical and ethical concern (Sjoding et al., 2020).

Motion artifact correction was another critical technical domain addressed by AI systems. Argüello-Prada and Castillo García (2024) highlighted the effectiveness of reference signal-less machine learning techniques in real-time detection of corrupted photoplethysmographic (PPG) segments. This functionality is essential for ensuring data quality in wearable and

home-monitoring devices, where user movement is frequent and uncontrolled. Although non-AI automated systems such as O2matic (Hansen et al., 2018) and the iPOC system (Sanchez-Morillo et al., 2020) offer practical real-time oxygen titration, they largely assume artifact-free input and lack the preprocessing pipelines required to ensure signal fidelity.

Personalization, a cornerstone of precision medicine, is another domain where AI systems excel. For example, Pascual-Saldaña et al. (2024) proposed a predictive oxygen dosing framework using edge-based AI models trained on historical physiological and behavioral data. Their approach marks a shift from reactive to proactive oxygen therapy, capable of adapting to individualized trends and anticipating oxygen requirements during variable activity levels. While this system remains in the prototype stage, it reflects the broader trajectory toward closed-loop AI-driven personalization in chronic disease management (Topol, 2019).

On the other hand the three automated systems included in this review (O2matic (Hansen et al., 2018), the iPOC system (Sanchez-Morillo et al., 2020), and the titration device tested by Cirio and Nava (2011)) used the rule-based logic to adjust oxygen flow in response to the real-time SpO_2 readings. The systems demonstrated significant improvement in oxygen stability and reduced hypoxemic episodes during the clinical testing phase. For example Hansen et al. (2018) reported that O2matic maintained the patients within target SpO_2 range 85.1% of time compared to only 46.6% under manual control with corresponding reductions in time spent below critical thresholds. Although clinically validated these systems offer limited adaptability to the individual physiology or behavioral context lacking the forecasting and learning capabilities of the AI-based methods.

A key limitation identified across most studies is the lack of validation in diverse, real-world LTOT scenarios. Although AI models reported strong performance in simulated datasets or controlled environments, few studies tested their integration into wearable, portable, or home-based systems over extended periods. Furthermore, none of the reviewed works—except Cabanas et al. (2024)—explicitly evaluated system performance across varied demographic groups, such as individuals with darker skin tones, despite well-documented racial disparities in pulse oximetry accuracy (Sjoding et al., 2020; Fawzy et al., 2022). Addressing these issues is vital to ensure equitable deployment and avoid perpetuating systemic healthcare bias in future AI-LTOT applications.

Another issue is the divergence in maturity between studies. Several AI models, such as those by Pascual-Saldaña et al. (2024) and the AI4Health initiative (Pascual et al., 2023), remain in proof-of-concept or early prototyping stages. Conversely, automated systems like O2matic have undergone clinical testing and demonstrated real-world efficacy. This discrepancy suggests that while AI holds greater long-term potential for personalization and adaptability, automated systems currently offer more immediate applicability, particularly in institutional or acute care settings.

Future directions should emphasize hybrid systems that integrate the learning and adaptive capabilities of AI with the real-time control and fail-safe responsiveness of automation. Such convergence would support robust, continuous SpO_2 monitoring and dynamic oxygen titration that is both context-aware and evidence-driven. In parallel, standardized reporting of validation metrics—including MAE, RMSE, signal loss rates, and performance under motion and perfusion variability—should be adopted to facilitate cross-study comparability and regulatory assessment.

Finally, ethical considerations must remain central. Issues related to algorithmic transparency, model interpretability, and accountability in closed-loop decision-making systems require further exploration as these technologies move toward clinical deployment (Shortliffe & Sepúlveda, 2018). Addressing these concerns alongside technical innovation will be critical for securing clinician trust, regulatory approval, and patient safety.

Strengths and Limitations

This systematic review presents a timely and methodologically rigorous synthesis of emerging artificial intelligence (AI) and automation-based approaches to continuous SpO_2 monitoring in the context of long-term oxygen therapy (LTOT). A key strength of this review lies in its comprehensive inclusion strategy, which incorporated both AI-driven predictive models and rule-based automated titration systems. By doing so, this work offers a comparative view of the evolving technological landscape, highlighting the respective capabilities, limitations, and levels of maturity across both domains. The review also employed structured inclusion criteria aligned with the PRISMA guidelines including only peer-reviewed, English-language studies that reported original technical or clinical data. This enhances the credibility and reproducibility of the findings.

Another strength of this review is its explicit focus on technical challenges critical to real-world LTOT deployment. Motion artifact correction, low perfusion signal integrity, and skin tone bias—factors often overlooked in broader digital health reviews—were treated as primary criteria for evaluating the readiness of SpO_2 monitoring technologies. The identification of Argüello-Prada and Castillo García's (2024) work on reference signal-less artifact detection and Cabanas et al.'s (2024) bias analysis across skin pigmentation further underscores the depth of this technical appraisal. Moreover, the inclusion of studies addressing edge computing and predictive dosing (e.g., Pascual-Saldaña et al., 2024) expands the review's relevance to emerging decentralized care models and resource-limited settings.

Nonetheless, several limitations must be acknowledged. First, despite the breadth of technological themes, the number of studies meeting all eligibility criteria was relatively small (n=8), limiting the generalizability of certain conclusions. This reflects both the novelty of the field and the lack of large-scale clinical trials validating AI-enhanced LTOT systems in real-world, long-duration scenarios. While some automated systems (e.g., Hansen et al., 2018; Sanchez-Morillo et al., 2020) were tested in clinical or semi-controlled environments, most

AI-driven models were evaluated using retrospective or simulated datasets, reducing their external validity and translational applicability.

Second, although the review focused on English-language and open-access literature to ensure transparency and reproducibility, this may have introduced language or publication bias. Relevant studies published in other languages or behind paywalls may have been excluded despite meeting technical and clinical criteria (Moher et al., 2009).

Third, the heterogeneity in outcome metrics across studies posed challenges for quantitative synthesis. While some AI studies reported mean absolute error (MAE) and root mean square error (RMSE), others focused on qualitative performance indicators such as time-in-target saturation or user feedback. The lack of standardized benchmarking methods across SpO_2 technologies, such as consistent use of Bland-Altman analysis or calibration against arterial blood gas reference, limits direct comparability and meta-analysis (Johansson et al., 2020).

Finally, while this review captured significant insights into technical readiness, it did not fully address broader implementation challenges such as regulatory approval, cybersecurity, cost-effectiveness, or clinician and patient acceptance. These factors are critical to the adoption of AI and automated technologies in LTOT and warrant future investigation in health systems research and clinical implementation studies (Shortliffe & Sepúlveda, 2018; Topol, 2019).

Despite these limitations, this review offers an important contribution to the growing field of intelligent oxygen therapy, providing a structured framework for assessing the clinical and technical viability of next-generation SpO_2 monitoring systems in LTOT care.

Risk of Bias Analysis

Risk of bias was systematically assessed for all eight included studies using a tailored approach combining principles from the Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) tool. This approach enabled a comprehensive evaluation of methodological rigor ensuring a critical appraisal of studies involving machine learning models and automated closed-loop oxygen systems.

1. Bias Due to Confounding

Studies developing AI models (e.g., Cabanas et al., 2024; Shuzan et al., 2023) were generally at moderate risk of confounding bias, as many used publicly available datasets (e.g., PhysioNet) that lacked full clinical context or were limited in demographic diversity. Only one study (Cabanas et al., 2024) performed subgroup analysis to account for skin tone bias, identifying measurable differences in prediction error by pigmentation. None of the automated device studies (e.g., Hansen et al., 2018; Sanchez-Morillo et al., 2020) stratified outcomes by

comorbidities or socioeconomic status, leading to potential confounding from unmeasured baseline variables.

2. Bias in Participant Selection

Two AI studies (Shuzan et al., 2023; Argüello-Prada & Castillo, 2024) trained models exclusively on pre-curated, high-quality PPG signals, raising high risk of selection bias. These datasets may overrepresent idealized sensor conditions and exclude noisy, real-world use cases—particularly relevant for home-based LTOT users. Automated systems like O2matic and iPOC were tested in small, specific clinical populations (Hansen et al., 2018; Sanchez-Morillo et al., 2020), limiting generalizability. In both cases, sampling frames lacked clear representativeness of diverse LTOT user groups, including those in remote or resource-limited environments.

3. Bias Due to Missing Data

The risk of bias due to missing data was generally low to moderate. Most studies using AI reported complete-case analysis, but few addressed data loss during motion, sensor disconnection, or long-term monitoring interruptions. Only the iPOC system (Sanchez-Morillo et al., 2020) explicitly described system dropout detection and reinitialization. The absence of sensitivity analyses in most studies to address missingness mechanisms (e.g., MCAR vs. MAR) limits the robustness of performance claims.

4. Bias in Measurement of Outcomes

Several AI studies reported standard performance metrics (e.g., MAE, RMSE); however, only one (Cabanas et al., 2024) performed Bland–Altman analysis to evaluate agreement with reference measurements. Automated device studies used clinical thresholds (e.g., % time within target SpO_2) but lacked consistent validation against arterial blood gas (ABG) reference standards. This introduces moderate bias in outcome measurement, especially when SpO_2 was used both as input and output in validation phases.

5. Bias Due to Algorithmic Transparency and Reproducibility

Among AI-focused studies, transparency and reproducibility varied widely. While Pascual-Saldaña et al. (2024) described model architectures and training methodology, others (e.g., AI4Health prototype) lacked access to source code or detailed hyperparameter tuning, introducing serious risk of reproducibility bias. Only Cabanas et al. (2024) reported demographic stratification and external validation across different sensor platforms, reducing risk in this domain.

6. Bias in Selection of Reported Results

Risk of reporting bias was moderate across most studies. While performance metrics were often reported, few studies discussed failure cases, outliers, or conditions under which the system underperformed. Selective reporting of high-performance scenarios may inflate perceived system robustness.The overall risk of bias across included studies ranged from low to high, with the highest risks observed in areas of participant selection, algorithmic transparency, outcome validation standards. While several AI systems demonstrated strong technical performance, many lacked demographic stratification, robustness testing under real-world conditions, and comprehensive reporting on edge cases. To enhance methodological rigor, future studies should adopt CONSORT-AI/DECIDE-AI guidelines (Liu et al., 2020; Vasey et al., 2022), report failure modes transparently, and ensure model reproducibility through open-source practices.

Table 2

Risk of Bias

Study	Confounding	Selection	ClassificationInterventions	DeviationsInterventions	Missing Data	Outcome Measurement	ReportingBias	Overall Risk
Cabanas et al. (2024)	Moderate – Some bias from limited dataset generalizability despite subgroup analysis	Moderate – Public datasets underrepresent real-world complexity	Low – Transparent AI methodology with defined inputs	Low – No intervention deviation (model-based)	Low – Complete data usage; no major attrition	Low – Benchmarking via MAE, RMSE, Bland–Altman	Low – Bias analysis reported, stratified performance	Low–Moderate
Shuzan et al. (2023)	Moderate – No demographic bias correction	High – Curated datasets may omit real-world signal variance	Low – ML model inputs clearly described	Moderate – No mention of deviation monitoring in practical settings	Moderate – No missing data protocols stated	Moderate – RMSE reported but no agreement analysis	Moderate – Limited exploration of failure cases	Moderate –High
Argüello-	Moderate	High –	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate

Prada & Castillo (2024)	– Signal analysis not stratified by population	Reviews ML models trained on optimal segments	– Review-based classification, not applied directly	– Real-time applications conceptual only	– Missingness and generalizability untested	– Emphasis on signal noise, not outcome accuracy	– No bias transparency in source studies	–High
Pascual-Saldaña et al. (2024)	Moderate – Pilot design, small N, no demographic stratification	Moderate – Limited validation cohort	Low – Intervention logic fully described in predictive system	Moderate – Deviations not reported/tested	Moderate – Limited reporting of data loss handling	Moderate – No quantitative SpO_2 validation yet	Moderate – Conceptual framework only	Moderate
(Pascual et al., 2023)	High – No confounding controls or demographic reporting	Moderate – Prototype-only, no real-world testing	High – Classification not validated	High – No deviation monitoring in use	Moderate – Small-scale internal validation	High – No outcome benchmark or comparison	High – Performance selectively presented	High
Sanchez-Morillo et al. (2020)	Moderate – No subgroup performance stratification	Moderate – Functional test in a small group	Low – Automated rule-based classification is appropriate	Low – Fully automated system in use	Low – Dropout recovery noted	Moderate – No ABG or device comparison	Moderate – Outcome focus limited to function	Moderate
Hansen et al. (2018)	Moderate – In-hospital populatio	Moderate – Crossover design, small	Low – Intervention well-defined	Low – Closed-loop system maintains	Low – High completeness of data	Moderate – SpO_2 used as both control	Moderate – Reporting focused on	Low–Moderate

	n only	sample		fidelity		and measure	efficacy metrics	
Cirio & Nava (2011)	High – No stratified analysis or real-world generalizability	Moderate – Pilot crossover design	Low – Direct comparison with manual titration	Low – Deviations minimized by design	Low – Minimal missing data in supervised trial	High – No standardized validation used	High – Selective metrics, no algorithmic reporting	High

Future Directions

As the convergence of artificial intelligence (AI), sensor miniaturization, and automation continues to advance, future research in continuous oxygen saturation (SpO_2) monitoring and long-term oxygen therapy (LTOT) should focus on translating innovation into clinically validated, scalable, and equitable care solutions. Based on the current evidence, several directions are recommended to address persistent gaps in performance, usability, and real-world implementation.

First, there is a clear need for longitudinal, real-world validation of AI-enhanced SpO_2 monitoring systems in ambulatory and home-based LTOT populations. While several studies demonstrated promising performance in simulated or controlled settings (Cabanas et al., 2024; Shuzan et al., 2023), few evaluated the systems under prolonged, daily-use conditions involving variable motion, ambient lighting, and physiological instability. Future studies should incorporate prospective cohort designs with continuous deployment in patients' natural environments and benchmark against reference standards such as arterial blood gas measurements (Johansson et al., 2020). Real-world performance metrics—including dropout rate, signal loss during motion, and user-reported acceptability—must become core outcome measures.

Second, future systems should adopt hybrid architectures that combine the adaptability of AI with the safety and responsiveness of automation. AI-driven models are well suited for learning patient-specific patterns and forecasting oxygen requirements, whereas automation ensures immediate adjustments to current physiological needs. Edge computing, as proposed by Pascual-Saldaña et al. (2024), offers a promising solution by enabling real-time, low-latency AI decision-making without reliance on continuous cloud connectivity—especially critical for remote care and low-resource environments.

Third, addressing equity and generalizability is paramount. Most AI models to date have been trained and validated on limited datasets that underrepresent diverse demographics, particularly in relation to skin tone, age, and comorbidity profiles. Given the documented inaccuracy of pulse oximetry in individuals with darker skin pigmentation (Sjoding et al., 2020; Fawzy et al., 2022), future work must prioritize training datasets that reflect population heterogeneity and implement stratified performance reporting. This will help prevent the propagation of algorithmic bias and improve trust among underrepresented patient groups.

Fourth, the digital connectivity with digital health platforms will be of high importance in order to provide remote patient monitoring, clinician monitoring, and personalized alerting. Interoperability without latency or breaks for EHRs and mobile health applications will be crucial for the efficiency of the workflow and its users. To offer patient safety as well, closed-loop solutions need to be constructed for regulatory compliance to include medical-grade software and security issues (Topol, 2019; Shortliffe & Sepúlveda, 2018).

Finally, researchers and developers must engage with stakeholders—patients, caregivers, clinicians, and regulators—to ensure that AI and automated LTOT systems are ethically aligned, clinically acceptable, and technically transparent. This includes providing interpretable outputs, enabling override controls, and clearly defining accountability in autonomous systems (London, 2019). Multidisciplinary collaboration and user-centered design methodologies will be essential to develop technologies that not only perform well but are trusted and adopted across care settings.

In summary, while AI and automation have already demonstrated substantial potential to improve LTOT delivery, their full promise will only be realized through rigorous clinical validation, ethical design, equitable implementation, and seamless integration into healthcare systems. These directions should guide the next generation of research and development toward safer, smarter, and more personalized oxygen therapy.

Opportunities of Automation or AI in Long-Term Oxygen Therapy (LTOT)

The integration of AI and automation into LTOT represents a shift in paradigm regarding the management of chronic respiratory diseases such as chronic obstructive pulmonary disease (COPD) and interstitial lung disease (ILD). These technologies offer transformative opportunities to enhance the precision, safety, efficiency, and personalization of oxygen delivery, addressing longstanding limitations in static and clinician-dependent LTOT models.

1. Real-Time, Adaptive Oxygen Delivery

One of the most immediate benefits of automation in LTOT is the ability to deliver oxygen in real time to a patient's actual physiological state. Computer-controlled devices such as O2matic (Hansen et al., 2018) and iPOC (Sanchez-Morillo et al., 2020) dynamically titrate flow of oxygen

based on SpO_2 levels measured continuously, holding patients in the target range of saturation more consistently than titration by hand. This reduces the danger of hypoxemia and hyperoxia, enhancing safety and with promise to reduce hospitalization and exacerbation frequency (Hardinge et al., 2015; Leach et al., 1992).

2. Personalized and Predictive Oxygen Dosing

AI enables a shift from reactive to predictive and personalized care in LTOT. Unlike conventional dosing based on intermittent assessments like the 6-minute walk test, AI systems can analyze continuous physiological data to learn individual patterns of oxygen need across different contexts—rest, sleep, exertion, or exacerbation. Predictive algorithms, such as those proposed by Pascual-Saldaña et al. (2024), can forecast oxygen demands based on prior behavior, activity levels, and symptom trends, enabling proactive adjustments. This personalized dosing may enhance therapy adherence, minimize complications, and better align with the principles of precision medicine (Topol, 2019).

3. Improved Signal Reliability and Decision Support

Photoplethysmography (PPG) signals used in SpO_2 monitoring are notoriously sensitive to motion, low perfusion, and skin pigmentation. AI offers advanced signal processing capabilities to detect and correct for motion artifacts and noise without additional sensors. Argüello-Prada and Castillo García (2024) demonstrated that machine learning models can classify corrupted versus clean PPG segments in real time, enhancing the reliability of downstream decisions. In parallel, AI-based signal interpretation can provide clinicians with early warnings about deteriorating trends, empowering more informed interventions (Shuzan et al., 2023).

4. Equity and Access Through Remote Monitoring

AI and automation also offer a pathway to democratizing access to LTOT by enabling remote care delivery. Intelligent oxygen systems embedded in portable or wearable devices, supported by mobile health platforms, can extend high-quality respiratory care to rural, underserved, or mobility-restricted populations (Chan et al., 2021). Edge AI systems, such as those conceptualized by Pascual-Saldaña et al. (2024), operate locally on low-power devices, reducing dependence on cloud infrastructure and expanding feasibility in resource-limited settings.

5. Reduction in Clinical Burden and Cost

From a health system perspective, automation in oxygen titration can significantly reduce the demand on healthcare providers. Studies like Cirio and Nava (2011) demonstrated that automated titration systems reduced respiratory therapist intervention time during exercise testing while maintaining target SpO_2 levels. At scale, these efficiencies could translate into lower personnel costs, fewer clinical visits for oxygen reassessment, and reduced emergency

admissions for oxygen-related complications. Moreover, AI algorithms can continuously audit usage patterns and detect anomalies, supporting reimbursement verification and supply optimization (Topol, 2019).

6. Data-Driven Research and Innovation

Finally, AI and automation inherently generate large volumes of longitudinal physiological data. These datasets can fuel ongoing research into oxygen therapy efficacy, disease progression modeling, and population-level analytics. With appropriate ethical and privacy safeguards, this data can be used to refine algorithms, inform public health planning, and support regulatory evaluation of new devices (London, 2019; Shortliffe & Sepúlveda, 2018).

Stakeholders Perspective

Patient Perspective

From the patient's standpoint, the adoption of artificial intelligence (AI) and automated systems in long-term oxygen therapy (LTOT) offers the potential for substantial improvements in safety, autonomy, and quality of life. However, it also raises concerns regarding transparency, control, and trust in decision-making systems. Understanding these perspectives is essential for ensuring that technological advancements are aligned with patient-centered care principles.

1. Increased Safety and Reduced Burden

Automated and AI-enhanced LTOT systems can relieve patients of the cognitive and physical burden of manually adjusting oxygen flow during various activities. Studies have shown that static dosing often leads to under-oxygenation during exertion and over-oxygenation at rest (Plant et al., 2000; Hardinge et al., 2015). Closed-loop systems such as O2matic, which automatically maintain target SpO_2 ranges (Hansen et al., 2018), provide a more responsive and consistent oxygenation experience, potentially reducing symptoms like dyspnea and fatigue. Patients, especially those with advanced COPD, may appreciate the reduction in monitoring requirements and the ability to focus on activities of daily living without interruption.

2. Enhanced Autonomy and Mobility

AI-enabled LTOT systems offer the potential to increase patient autonomy through portable, intelligent devices capable of adapting to daily fluctuations in oxygen demand. For example, the iPOC system (Sanchez-Morillo et al., 2020) and edge-AI-based dosing models (Pascual-Saldaña et al., 2024) are designed to enable real-time adaptation outside clinical settings. This mobility

and personalization are particularly valuable to individuals who wish to remain active and engaged in their communities while managing their respiratory condition.

3. Trust, Transparency, and Control

Despite these benefits, patients may express concerns about handing over control of oxygen delivery to automated systems, especially if the algorithms are opaque or perceived as "black boxes" (London, 2019). Lack of explainability may result in reduced trust in the system, particularly when oxygen flow decisions deviate from patient expectations or prior clinical prescriptions. It is important that patients are informed about how decisions are made, and that override capabilities are available, allowing them to regain control when necessary (Shortliffe & Sepúlveda, 2018).

4. User Experience and Accessibility

Ease of use is critical, especially for older adults who make up a large proportion of LTOT users and may have visual, cognitive, or dexterity impairments. Interfaces must be intuitive and adaptable, with audible, visual, or haptic feedback to confirm function and alert patients to malfunctions or status changes. Moreover, AI systems must be culturally and linguistically accessible to ensure equity across diverse populations. Studies have highlighted that usability issues can significantly hinder the adoption of digital health tools, regardless of their technical sophistication (Vaportzis et al., 2017).

5. Privacy and Data Sharing

AI-enhanced oxygen systems often require continuous physiological monitoring and may transmit health data to clinicians or cloud-based platforms. While this facilitates personalized care and early intervention, patients may be concerned about the security and privacy of their data. Ethical design principles should prioritize data minimization, local processing when feasible (e.g., edge computing), and transparent data use policies (Topol, 2019).

6. Perceived Value and Cost Considerations

Cost is another factor shaping patient acceptance. While intelligent LTOT systems offer potential cost savings by reducing emergency visits and hospitalizations, the initial cost of devices may be a barrier for underinsured or low-income individuals. Reimbursement frameworks and insurance coverage will influence how widely these systems are adopted. Patient education regarding long-term benefits and cost-effectiveness is essential to foster informed decision-making and shared value perception.

Clinician Perspective

For clinicians managing patients on long-term oxygen therapy (LTOT), the integration of artificial intelligence (AI) and automation offers both promising opportunities and practical challenges. Healthcare professionals are key stakeholders in the adoption, monitoring, and interpretation of AI-enhanced SpO_2 systems, and their perspectives are essential to ensure successful clinical integration and safe patient outcomes.

1. Support for Decision-Making and Monitoring

AI and automation have the potential to augment clinical decision-making by enabling continuous, real-time monitoring of oxygen saturation and adaptive oxygen delivery without the need for frequent manual adjustments. This is particularly beneficial in outpatient and home-care settings, where clinicians currently rely on infrequent spot-checks and retrospective symptom reporting. Systems such as O2matic and iPOC provide automated oxygen titration that aligns with clinician-defined SpO_2 targets (Hansen et al., 2018; Sanchez-Morillo et al., 2020), reducing the time clinicians spend on flow adjustments and potentially improving long-term oxygenation control.

In predictive applications, AI models may identify trends suggestive of clinical deterioration or poor adherence, alerting clinicians to intervene earlier. For example, Pascual-Saldaña et al. (2024) proposed AI-based oxygen dosing systems that learn from patient behavior and physiological response to anticipate oxygen needs. Such tools could enhance the personalization of LTOT prescriptions, potentially improving outcomes and reducing exacerbations.

2. Workflow Integration and Interoperability

For AI and automation to deliver clinical value, they must integrate seamlessly into existing workflows. Clinicians express concern when new tools require additional training, disrupt documentation processes, or are incompatible with electronic health record (EHR) systems (Shortliffe & Sepúlveda, 2018). Interoperability between SpO_2 monitoring platforms, EHRs, and clinical decision support systems is necessary to ensure that relevant data can be reviewed, annotated, and acted upon efficiently.

Moreover, clinicians value systems that can operate with minimal setup and maintenance in the home environment. Systems that automate titration based on clear, adjustable clinical targets—such as O2matic—are generally well-received because they maintain a transparent link between physician intention and algorithmic action (Hansen et al., 2018).

3. Clinical Safety and Oversight

Patient safety remains a top concern among healthcare providers when delegating control to automated systems. Clinicians need assurance that AI and closed-loop systems are reliable under a wide range of clinical scenarios, including changes in comorbidities, acute exacerbations, or

unexpected physiological responses. Concerns have also been raised about sensor reliability, signal loss during motion, and the accuracy of AI-driven predictions in underrepresented populations (Cabanas et al., 2024; Sjoding et al., 2020).

The ability to audit AI recommendations or override automated titration is essential to preserve clinician agency. Without transparency in decision-making logic or access to the model's rationale (e.g., a summary of why oxygen flow was increased or decreased), clinicians may hesitate to rely on these systems, particularly in high-risk patients (London, 2019). Hence, explainability and clinician override functions are critical for clinical adoption.

4. Training, Trust, and Legal Accountability

AI systems used in LTOT must be accompanied by appropriate clinician training to ensure confident and competent use. As noted by Topol (2019), one of the major barriers to AI implementation in healthcare is the lack of digital fluency among frontline providers. Clinicians may also question the legal and ethical implications of algorithm-driven decisions, particularly if adverse outcomes occur. Clarifying liability frameworks—especially in semi-autonomous systems—is essential to foster professional trust and avoid medicolegal uncertainty (Morley et al., 2020).

Furthermore, clinicians may harbor skepticism toward AI recommendations that conflict with their clinical judgment or patient preferences. Engaging clinicians in the design, testing, and feedback stages of AI tool development has been shown to improve acceptance and usability (Kelly et al., 2019). This participatory approach also ensures that AI systems are aligned with clinical priorities and contextual knowledge.

5. Opportunities for Population-Level Management

Beyond individual care, clinicians recognize the value of AI in supporting population health strategies. Longitudinal data generated from intelligent LTOT systems can inform risk stratification, adherence monitoring, and remote triage, especially when paired with telehealth platforms. Aggregated data could also be used to evaluate the real-world effectiveness of LTOT and identify patterns of suboptimal therapy across care networks (Chan et al., 2021).

Caregiver Perspective

Caregivers—whether family members, home health aides, or other informal support providers—play a crucial role in assisting patients on long-term oxygen therapy (LTOT), particularly those who are elderly, frail, or have cognitive or mobility impairments. The introduction of artificial intelligence (AI) and automation into SpO_2 monitoring and oxygen delivery systems offers several potential benefits to caregivers, while also introducing new responsibilities and considerations.

1. Reduced Care Burden and Improved Confidence

One of the primary benefits caregivers may experience is a reduction in physical and emotional burden. Traditional LTOT often requires caregivers to monitor oxygen saturation levels, adjust flow rates, respond to alarms, and coordinate clinical check-ins. Automated systems, such as O2matic (Hansen et al., 2018) and iPOC (Sanchez-Morillo et al., 2020), that dynamically titrate oxygen in real-time based on SpO_2 input relieve caregivers from frequent manual interventions. This can be particularly valuable during exertion or sleep, when oxygen needs fluctuate rapidly and caregiver supervision is limited.

When caregivers observe consistent system performance—e.g., reliable maintenance of SpO_2 within target ranges—they may feel greater peace of mind and confidence in the patient's safety (Wittenberg et al., 2017). In turn, this can reduce anxiety, improve sleep quality for cohabitating caregivers, and enhance overall quality of life.

2. Simplified Monitoring and Alerts

AI-enabled systems with remote monitoring capabilities may allow caregivers to view real-time or trend data through apps or web-based dashboards. This is particularly beneficial for family caregivers managing patients from a distance or supporting patients with communication limitations (Topol, 2019). Some systems can also generate automated alerts in response to deteriorating trends or abnormal SpO_2 patterns, prompting caregivers to act or seek help before clinical deterioration.

However, caregivers may also express concern about **information overload** or **false alarms**, particularly if alerts are not well-calibrated or contextualized (Kelly et al., 2019). Alert fatigue may occur if systems are overly sensitive, leading caregivers to ignore or disable notifications, potentially undermining safety.

3. Ease of Use and System Maintenance

From the caregiver perspective, usability is essential. Caregivers often manage device setup, maintenance (e.g., tubing, battery charging), and troubleshooting, especially when patients have physical or cognitive impairments. Systems that require frequent recalibration or have complex interfaces may increase caregiver stress, while those with intuitive controls and minimal maintenance requirements are more likely to be accepted (Vaportzis et al., 2017).

Automated devices that integrate seamlessly into daily routines—such as portable concentrators with auto-adjusting features—can reduce perceived complexity and promote consistent adherence to therapy plans.

4. Training Needs and Role Adaptation

The introduction of AI and automation may shift the caregiver role from "oxygen manager" to "system overseer." While this can reduce physical responsibilities, it also introduces the need for technical literacy and system understanding. Caregivers may require training to interpret system outputs, understand how to override or escalate decisions, and maintain device functionality (Shortliffe & Sepúlveda, 2018). Without proper onboarding and support, caregivers may feel overwhelmed or mistrustful of systems they do not fully understand.

Clear, user-friendly manuals, help hotlines, and caregiver-inclusive training sessions are essential to ensure safe and confident use. Tools that include caregiver-facing interfaces or companion apps may also support greater engagement and monitoring transparency.

5. Emotional Implications and Trust

Caregivers may experience emotional tension when delegating critical safety tasks to an algorithm, particularly in high-risk patients. Trust in the technology—its ability to detect distress, respond accurately, and notify the appropriate party—is foundational to caregiver acceptance. Transparent systems that offer insight into their logic (e.g., "oxygen increased due to drop in SpO_2 during movement") may foster greater understanding and reduce fear of failure (London, 2019; Morley et al., 2020).

Moreover, systems that support caregiver involvement, rather than replace it, are more likely to be positively received. For instance, providing caregivers with timely summaries or daily reports enhances their sense of agency while preserving automation's benefits.

6. Economic and Logistical Considerations

From a practical standpoint, caregivers are often the individuals responsible for coordinating device procurement, insurance navigation, and maintenance logistics. If AI-enhanced or automated oxygen systems carry higher out-of-pocket costs or require complex setup, caregivers may face additional financial or logistical stress. Policies that promote affordability, clear warranty/support structures, and home delivery services can ease these burdens and encourage adoption.

Stakeholders Perspective: Regulators and Policy Makers

The growing integration of artificial intelligence (AI) and automation into medical technologies such as continuous oxygen saturation (SpO_2) monitoring and long-term oxygen therapy (LTOT) presents novel challenges and responsibilities for regulators and policy makers. These stakeholders are tasked with balancing innovation, safety, data privacy, and equitable access, while establishing robust frameworks to guide clinical adoption and protect public trust.

1. AI as a Regulated Medical Device

AI-enabled SpO_2 monitoring systems that influence real-time clinical decisions, such as adjusting oxygen flow, are considered software as a medical device (SaMD) by global regulatory bodies including the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), and the International Medical Device Regulators Forum (IMDRF). These regulators have begun to issue specific guidance for AI/ML-based SaMD, focusing on performance transparency, real-world learning adaptability, and safety assurance (FDA, 2021; IMDRF, 2017).

For LTOT systems using adaptive algorithms, regulators require evidence of clinical effectiveness, algorithm interpretability, cybersecurity protection, and human oversight mechanisms. As most AI models used in SpO_2 estimation and predictive oxygen dosing are not static, but evolve over time (e.g., self-improving neural networks), regulators are increasingly emphasizing "Good Machine Learning Practices" (GMLP) and algorithm change protocols (FDA, 2021).

2. Post-Market Surveillance and Algorithm Drift

Regulatory frameworks must extend beyond pre-market approval to encompass post-market performance monitoring, particularly for adaptive algorithms subject to “model drift.” Continuous performance validation is essential to ensure safety as algorithms encounter new populations, clinical settings, or sensor types. For example, oxygen saturation models trained on one PPG sensor may underperform on another due to signal fidelity differences, or exhibit biased behavior in populations underrepresented during model development (Cabanas et al., 2024; Sjoding et al., 2020).

Policy makers are now exploring lifecycle regulatory models, which require manufacturers to submit update protocols and monitoring strategies that include real-world data collection, user feedback loops, and automatic alert thresholds for abnormal function (Topol, 2019; Morley et al., 2020).

3. Equity and Bias Mitigation Requirements

Ensuring equity in AI-driven health interventions is an emerging regulatory mandate. Pulse oximeters are known to systematically overestimate SpO_2 in individuals with darker skin pigmentation, potentially leading to inappropriate LTOT dosing (Fawzy et al., 2022; Sjoding et al., 2020). Regulators are increasingly demanding that developers quantify and disclose algorithmic bias, stratify performance by demographic subgroups, and provide mitigation strategies when disparities are identified (London, 2019).

For instance, the FDA has encouraged inclusion of diverse populations in training datasets and post-market evaluations to ensure that AI models do not exacerbate existing health disparities (FDA, 2021). In the context of LTOT, this is particularly relevant as oxygen underprescription can result in severe hypoxemia, increased hospitalizations, and mortality.

4. Data Privacy, Cybersecurity, and Interoperability

AI-enhanced LTOT systems often rely on continuous physiological monitoring and wireless data transmission, raising complex regulatory challenges around data security and patient privacy. Compliance with data protection laws such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union is essential. These regulations require explicit consent, anonymization where applicable, secure data storage, and limits on secondary data use (Briggs & Sokol, 2021).

Additionally, regulators are beginning to set interoperability standards to ensure that AI-driven medical devices can integrate safely with hospital systems, electronic health records (EHRs), and telemedicine platforms. Without standardization, there is a risk of system fragmentation, safety errors, and clinician overload (Shortliffe & Sepúlveda, 2018).

5. Health Policy, Reimbursement, and Access

Beyond regulatory agencies, health policy makers have an essential role in determining whether AI-enabled LTOT devices are reimbursable under national health systems or private insurance. Evidence of cost-effectiveness, clinical utility, and quality-of-life improvement is often required to justify reimbursement. The lack of defined reimbursement pathways for AI-based tools remains a barrier to adoption, especially in low-resource settings (Kelly et al., 2019).

Policy frameworks must also address technology access and digital equity, ensuring that vulnerable populations—including the elderly, low-income patients, and those in rural areas—are not excluded from the benefits of intelligent oxygen therapy.

Ethical Considerations in AI-Driven Long-Term Oxygen Therapy (LTOT)

As artificial intelligence (AI) and automation are increasingly integrated into long-term oxygen therapy (LTOT) systems, a parallel emphasis must be placed on the ethical dimensions of their development, deployment, and use. These considerations are particularly salient in the context of vulnerable populations—such as older adults and individuals with chronic respiratory conditions—who often depend on LTOT for day-to-day functioning and quality of life. Ensuring that AI-driven SpO_2 monitoring and oxygen titration technologies uphold ethical standards is essential for safeguarding patient rights, equity, and trust.

1. Equity and Algorithmic Fairness

One of the most pressing ethical challenges is the potential for AI models to perpetuate or exacerbate health disparities. Evidence shows that pulse oximeters—upon which most AI-driven LTOT systems rely—tend to overestimate oxygen saturation in individuals with darker skin pigmentation, increasing the risk of hypoxemia and under-treatment (Sjoding et al., 2020; Fawzy

et al., 2022). If these biased data serve as inputs for AI models without proper mitigation, downstream decisions—such as predictive oxygen dosing—may systematically disadvantage certain racial or ethnic groups.

To address this, developers and researchers must ensure that AI models are trained and validated on diverse datasets and that performance is stratified and reported by demographic variables including skin tone, sex, and age (London, 2019). Regulatory bodies such as the U.S. FDA now encourage inclusion of representative populations and transparency regarding performance variation across subgroups (FDA, 2021). Ethically, this aligns with the principle of justice, ensuring fair access to the benefits of AI and preventing algorithmic harm to marginalized populations.

2. Transparency and Explainability

Another ethical imperative is transparency—both in how AI systems function and in how their outputs are communicated to patients and clinicians. Many AI models, particularly deep learning architectures, operate as "black boxes," offering limited visibility into the rationale behind their recommendations (Morley et al., 2020; London, 2019). This lack of explainability can erode trust and impede informed decision-making.

In the context of LTOT, where real-time oxygen adjustments affect critical physiological processes, clinicians and patients must be able to understand the basis of algorithmic decisions. Systems should include explainable AI (XAI) components that provide interpretable outputs or rationale for changes in oxygen flow. This transparency supports autonomy by enabling users to make informed choices and challenge or override automated decisions when needed (Floridi et al., 2018).

3. Autonomy and Informed Consent

Respect for patient autonomy is a foundational bioethical principle, and it is challenged when decisions are increasingly delegated to opaque AI systems. In AI-enhanced LTOT, autonomy is preserved when patients are informed of how AI components operate, what data are collected, and how decisions are made. Informed consent should be obtained not only for use of the device itself, but also for the collection and processing of continuous physiological data (Briggs & Sokol, 2021).

Moreover, patients should be given the option to opt out of fully automated oxygen regulation, either temporarily or permanently, and should retain control over key parameters when feasible. This is particularly important in situations where users perceive the algorithm's behavior as discordant with their lived experience (e.g., breathlessness not triggering increased oxygen flow).

4. Privacy, Surveillance, and Data Governance

AI-driven LTOT systems often rely on cloud connectivity and continuous data logging, which raises concerns about data privacy, surveillance, and secondary data use. Sensitive physiological and behavioral data—such as oxygenation trends, physical activity, and even sleep patterns—may be stored or transmitted, potentially exposing users to privacy breaches or non-consensual data exploitation (Topol, 2019; Shortliffe & Sepúlveda, 2018).

Developers must ensure compliance with legal frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. and the General Data Protection Regulation (GDPR) in Europe. Beyond compliance, ethical data governance requires data minimization, user access to data logs, and clear policies on who can access and analyze the data (Morley et al., 2020).

5. Accountability and Liability

As AI systems gain decision-making authority, questions arise about who is accountable when adverse outcomes occur. If a patient is harmed due to an erroneous prediction or system failure, determining responsibility—whether it lies with the clinician, device manufacturer, software developer, or healthcare institution—can be legally and ethically complex (Gerke et al., 2020).

Ethical implementation requires clear delineation of responsibilities, including the establishment of human oversight mechanisms and escalation pathways. Clinicians must be empowered to override automated decisions, and patients must be informed of the chain of accountability should errors arise.

Conclusion

This systematic review critically examined the current landscape of artificial intelligence (AI) and automated systems developed for continuous oxygen saturation (SpO_2) monitoring in the context of long-term oxygen therapy (LTOT). Across eight rigorously selected studies, we identified substantial technological progress in both AI-based SpO_2 estimation and closed-loop oxygen titration systems. These advances, while distinct in their mechanisms and maturity, collectively represent a pivotal shift from static, manual oxygen delivery toward responsive, personalized, and context-aware oxygen therapy.

AI-based approaches demonstrated exceptional technical performance, with reported mean absolute errors (MAEs) as low as 0.57% and root mean square errors (RMSEs) below 1%, even in low-perfusion or motion-prone signal conditions (Cabanas et al., 2024; Shuzan et al., 2023). These models harness sophisticated architectures such as Gaussian process regression and deep neural networks to derive oxygenation metrics from raw photoplethysmographic (PPG) data. Moreover, studies such as Cabanas et al. (2024) highlighted the potential for algorithmic bias

mitigation through skin tone–stratified analysis, addressing a critical ethical concern in pulse oximetry accuracy (Sjoding et al., 2020; Fawzy et al., 2022). Despite these strengths, real-world validation of these models remains limited, and integration into LTOT workflows—particularly in wearable or home-based contexts—has not yet been realized.

Conversely, non-AI automated systems such as O2matic (Hansen et al., 2018) and iPOC (Sanchez-Morillo et al., 2020) demonstrated high clinical relevance and usability, maintaining SpO_2 within target ranges and reducing the need for clinician intervention. These closed-loop devices operate through rule-based logic, adjusting oxygen flow in real time based on continuous saturation input, and have been tested in hospital and ambulatory environments. Their success in stabilizing oxygenation during exertion or acute exacerbation illustrates their readiness for clinical deployment, though they lack the predictive personalization and adaptability offered by AI systems.

A key insight from this review is that AI and automation serve complementary roles in the evolution of LTOT. AI excels in anticipatory and adaptive modeling—tailoring oxygen delivery based on individualized patterns—while automation offers immediate, reliable titration within predefined safety parameters. Future systems should aim to integrate these strengths, creating hybrid platforms that combine real-time responsiveness with long-term learning and personalization. This direction aligns with broader healthcare trends toward precision medicine and digital chronic disease management (Topol, 2019).

However, significant gaps persist. Risk of bias remains a concern, particularly in AI studies trained on homogenous or idealized datasets, where algorithm robustness under diverse, real-world conditions remains untested. Many studies lacked demographic representation, usability data, or clear reporting on algorithm transparency and explainability—factors that are essential for both clinician trust and ethical deployment (London, 2019; Morley et al., 2020). Moreover, few systems addressed the practical challenges of LTOT in underserved settings, such as offline functionality, affordability, and ease of use for elderly or non-technologically literate users (Vaportzis et al., 2017).

From a stakeholder standpoint, clinicians valued the potential of automation to reduce workload and maintain physiological stability, but also raised concerns around system interpretability and liability (Shortliffe & Sepúlveda, 2018). Patients and caregivers welcomed the promise of reduced burden and improved oxygen control, yet highlighted the need for transparency, control, and data privacy (Topol, 2019; Wittenberg et al., 2017). Regulators, meanwhile, are challenged to adapt evolving frameworks to ensure safety, equity, and accountability in these emerging systems (FDA, 2021; Gerke et al., 2020).

In conclusion, the convergence of AI and automation is reshaping the possibilities of long-term oxygen therapy. While closed-loop systems have already proven feasible and clinically beneficial, the next frontier lies in responsibly translating AI models into intelligent,

patient-centered LTOT systems that learn, adapt, and support equitable, high-quality respiratory care. Achieving this will require interdisciplinary collaboration across clinicians, engineers, ethicists, and regulators grounded in rigorous science, inclusive design, and a commitment to both innovation and safety.

References

Argüello-Prada, E. J., & Castillo García, J. F. (2024). Machine Learning Applied to Reference Signal-Less Detection of Motion Artifacts in Photoplethysmographic Signals: A Review. *Sensors*, *24*(22), 7193. https://doi.org/10.3390/s24227193

Cabanas, A. M., Sáez, N., Collao-Caiconte, P. O., Martín-Escudero, P., Pagán, J., Jiménez-Herranz, E., & Ayala, J. L. (2024). Evaluating AI Methods for Pulse Oximetry: Performance, Clinical Accuracy, and Comprehensive Bias Analysis. *Bioengineering*, *11*(11), 1061. https://doi.org/10.3390/bioengineering11111061

Chan, E. D., Chan, M. M., & Chan, M. M. (2013). Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations. *Respiratory medicine*, *107*(6), 789–799. https://doi.org/10.1016/j.rmed.2013.02.004

Cirio, S., & Nava, S. (2011). Pilot study of a new device to titrate oxygen flow in hypoxic patients on long-term oxygen therapy. *Respiratory care*, *56*(4), 429–434. https://doi.org/10.4187/respcare.00983

Hansen, E. F., Hove, J. D., Bech, C. S., Jensen, J. S., Kallemose, T., & Vestbo, J. (2018). Automated oxygen control with O2matic® during admission with exacerbation of COPD. *International journal of chronic obstructive pulmonary disease*, *13*, 3997–4003. https://doi.org/10.2147/COPD.S183762

Hardinge, M., Annandale, J., Bourne, S., Cooper, B., Evans, A., Freeman, D., Green, A., Hippolyte, S., Knowles, V., MacNee, W., McDonnell, L., Pye, K., Suntharalingam, J., Vora, V., Wilkinson, T., British Thoracic Society Home Oxygen Guideline Development Group, & British Thoracic Society Standards of Care Committee (2015). British Thoracic Society guidelines for home oxygen use in adults. *Thorax*, *70 Suppl 1*, i1–i43. https://doi.org/10.1136/thoraxjnl-2015-206865

Kelly, C. J., Karthikesalingam, A., Suleyman, M., et al. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine, 17*(1), 195. https://doi.org/10.1186/s12916-019-1426-2

London A. J. (2019). Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *The Hastings Center report*, *49*(1), 15–21. https://doi.org/10.1002/hast.973

Medical Research Council Working Party. (1981). Long term domiciliary oxygen therapy in chronic hypoxic cor pulmonale complicating chronic bronchitis and emphysema. *Lancet (London, England)*, *1*(8222), 681–686.

Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social science & medicine (1982)*, *260*, 113172. https://doi.org/10.1016/j.socscimed.2020.113172

Nocturnal Oxygen Therapy Trial Group. (1980). Continuous or nocturnal oxygen therapy in hypoxemic chronic obstructive lung disease: A clinical trial. *Annals of Internal Medicine, 93*(3), 391–398. https://doi.org/10.7326/0003-4819-93-3-391

Pascual-Saldaña, H., Masip-Bruin, X., Asensio, A., Alonso, A., & Blanco, I. (2024). Innovative Predictive Approach towards a Personalized Oxygen Dosing System. *Sensors (Basel, Switzerland)*, *24*(3), 764. https://doi.org/10.3390/s24030764

Pascual, H., Masip-Bruin, X., Alonso, A., & Blanco, I. (2023). *Analyzing distinct neural network models for oxygen saturation prediction towards a personalized COPD management*. In 2023 IEEE 19th International Conference on e-Science (e-Science) (pp. 1–8). IEEE. https://doi.org/10.1109/e-Science58273.2023.10254844

Plant, P. K., Owen, J. L., & Elliott, M. W. (2000). Early use of non-invasive ventilation for acute exacerbations of chronic obstructive pulmonary disease on general respiratory wards: a multicentre randomised controlled trial. *Lancet (London, England)*, *355*(9219), 1931–1935. https://doi.org/10.1016/s0140-6736(00)02323-0

Sanchez-Morillo, D., Muñoz-Zara, P., Lara-Doña, A., & Leon-Jimenez, A. (2020). Automated Home Oxygen Delivery for Patients with COPD and Respiratory Failure: A New Approach. *Sensors*, *20*(4), 1178. https://doi.org/10.3390/s20041178

Shuzan, M. N. I., Chowdhury, M. H., Chowdhury, M. E. H., Murugappan, M., Hoque Bhuiyan, E., Arslane Ayari, M., & Khandakar, A. (2023). Machine Learning-Based Respiration Rate and Blood Oxygen Saturation Estimation Using Photoplethysmogram Signals. *Bioengineering*, *10*(2), 167. https://doi.org/10.3390/bioengineering10020167

Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *JAMA, 320*(21), 2199–2200. https://doi.org/10.1001/jama.2018.17163

Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E., & Valley, T. S. (2020). Racial Bias in Pulse Oximetry Measurement. *The New England journal of medicine*, *383*(25), 2477–2478. https://doi.org/10.1056/NEJMc2029240

Global Initiative for Chronic Obstructive Lung Disease (GOLD). (2024). Global strategy for prevention, diagnosis and management of COPD: 2024 report. https://goldcopd.org/2024-gold-report

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

Vaportzis, E., Clausen, M. G., & Gow, A. J. (2017). Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. *Frontiers in psychology*, *8*, 1687. https://doi.org/10.3389/fpsyg.2017.01687

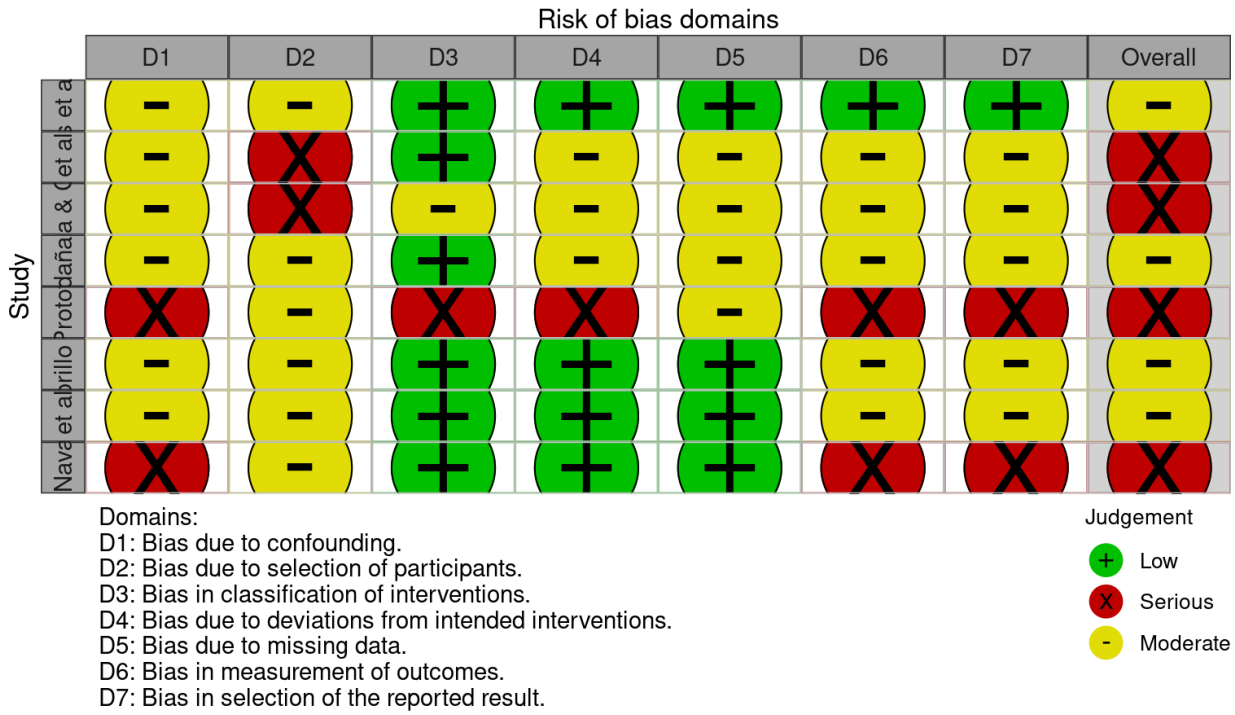
Wittenberg, E., Goldsmith, J. V., Chen, C., Prince-Paul, M., & Johnson, R. R. (2021). Opportunities to improve COVID-19 provider communication resources: A systematic review. Patient education and counseling, 104(3), 438–451. https://doi.org/10.1016/j.pec.2020.12.031

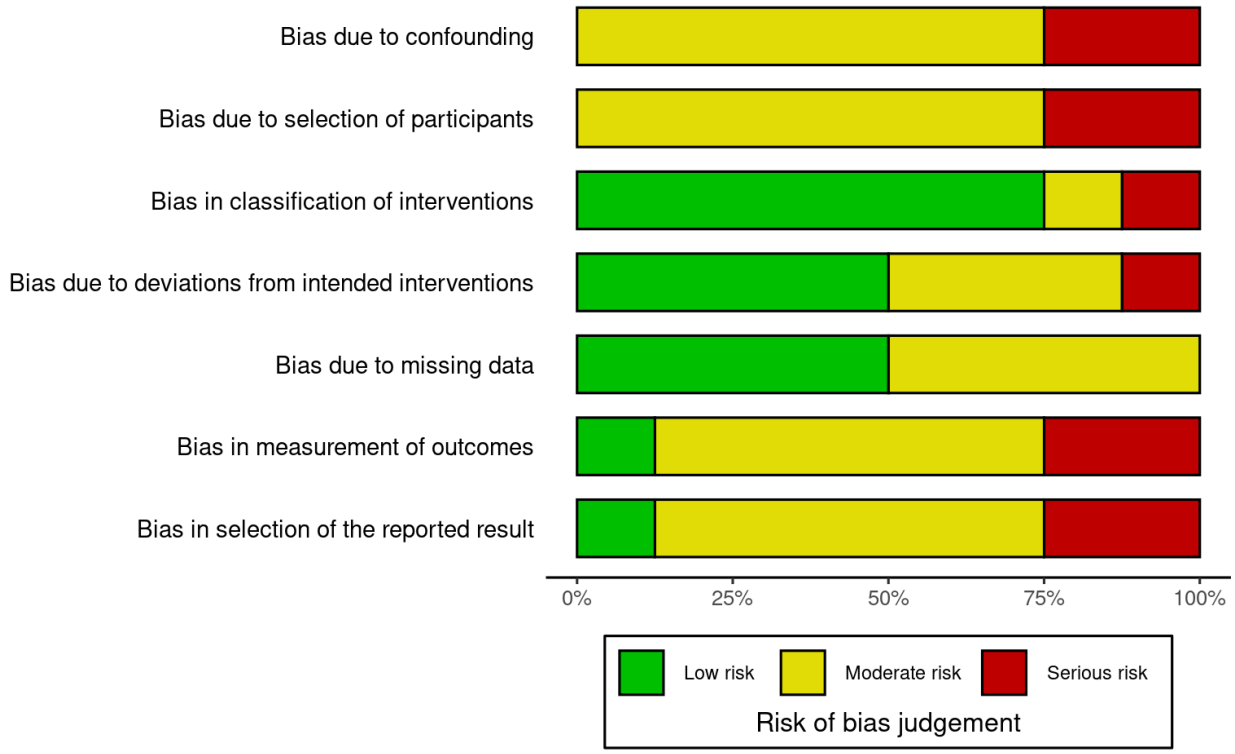
World Health Organization. (2020). Global health estimates: Life expectancy and leading causes of death and disability. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates

Supplementary Appendix A: Risk of Bias Analysis

R code:

```
library(dplyr)
library(robvis)

levels_robvis ← c("Low", "Moderate", "Serious", "Critical", "No information")

risk_of_bias_mod ← data.frame(
  Study = c(
    "Cabanas et al. (2024)",
    "Shuzan et al. (2023)",
    "Argüello-Prada & Castillo (2024)",
    "Pascual-Saldaña et al. (2024)",
    "AI4Health Prototype (2023)",
    "Sanchez-Morillo et al. (2020)",
    "Hansen et al. (2018)",
    "Cirio & Nava (2011)"
  ),
  Domain1 = c("Moderate", "Moderate", "Moderate", "Moderate", "Serious", "Moderate", "Moderate", "Serious"),
  Domain2 = c("Moderate", "Serious", "Serious", "Moderate", "Moderate", "Moderate", "Moderate", "Moderate"),
  Domain3 = c("Low", "Low", "Moderate", "Low", "Serious", "Low", "Low", "Low"),
  Domain4 = c("Low", "Moderate", "Moderate", "Moderate", "Serious", "Low", "Low", "Low"),
  Domain5 = c("Low", "Moderate", "Moderate", "Moderate", "Moderate", "Low", "Low", "Low"),
  Domain6 = c("Low", "Moderate", "Moderate", "Moderate", "Serious", "Moderate", "Moderate", "Serious"),
  Domain7 = c("Low", "Moderate", "Moderate", "Moderate", "Serious", "Moderate", "Moderate", "Serious"),
  Overall  = c("Moderate", "Serious", "Serious", "Moderate", "Serious", "Moderate", "Moderate", "Serious"),
  Weight   = c(1, 1, 1, 1, 1, 1, 1, 1),
  stringsAsFactors = FALSE
)

risk_of_bias_mod ← risk_of_bias_mod %>%
  mutate(across(starts_with("Domain"), ~ factor(., levels = levels_robvis, ordered = TRUE))) %>%
  mutate(Overall = factor(Overall, levels = levels_robvis, ordered = TRUE))

str(risk_of_bias_mod)

rob_summary(data = risk_of_bias_mod, tool = "ROBINS-I", weighted = TRUE)

rob_traffic_light(data = risk_of_bias_mod, tool = "ROBINS-I")
```


Risk of bias domains
D1
D2
D3
D4
D5
D6
D7
Overall
Study
Domains:
D1: Bias due to confounding.
D2: Bias due to selection of participants.
D3: Bias in classification of interventions.
D4: Bias due to deviations from intended interventions.
D5: Bias due to missing data.
D6: Bias in measurement of outcomes.
D7: Bias in selection of the reported result.
Judgement
Low
Serious
Moderate


Bias due to confounding
Bias due to selection of participants
Bias in classification of interventions
Bias due to deviations from intended interventions
Bias due to missing data
Bias in measurement of outcomes
Bias in selection of the reported result
0%
25%
50%
75%
100%
Low risk
Moderate risk
Serious risk
Risk of bias judgement

Supplementary Appendix B: Glossary

AI Artificial Intelligence — Computer systems that mimic human intelligence to perform tasks such as learning, decision-making, and pattern recognition.

LTOT Long-Term Oxygen Therapy — A chronic treatment involving the use of supplemental oxygen for extended periods (usually ≥15 hours/day), primarily in patients with chronic hypoxemia.

SpO_2 Peripheral Capillary Oxygen Saturation — A non-invasive measurement of the amount of oxygen-bound hemoglobin in the blood, expressed as a percentage.

PPG Photoplethysmography — An optical technique used to detect blood volume changes in the microvascular bed of tissue, commonly used in pulse oximeters.

MAE Mean Absolute Error — A metric used to measure the average magnitude of prediction errors in a model, without considering direction.

RMSE Root Mean Square Error — A statistical measure of the differences between predicted values and observed values; sensitive to large errors.

FDA Food and Drug Administration — The U.S. federal agency responsible for regulating medical devices, drugs, and diagnostics.

COPD Chronic Obstructive Pulmonary Disease — A progressive respiratory disease characterized by airflow limitation, often treated with LTOT.

SaMD Software as a Medical Device — Software intended for medical use that performs functions without being part of a hardware medical device.

6MWT Six-Minute Walk Test — A clinical test used to assess functional exercise capacity, often used to determine LTOT eligibility.

iPOC Intelligent Portable Oxygen Concentrator — A device that automatically adjusts oxygen delivery based on patient activity and real-time monitoring.

GMLP Good Machine Learning Practices — FDA-proposed standards to ensure safety, efficacy, and accountability in AI/ML medical software.

XAI Explainable Artificial Intelligence — A branch of AI focused on making model decision-making processes transparent and understandable to users.

EHR Electronic Health Record — A digital version of a patient's paper chart, containing medical history, diagnoses, medications, and treatment plans.

ABG Arterial Blood Gas — A test that measures oxygen and carbon dioxide levels in arterial blood, used to validate SpO_2 accuracy.

MCAR / MAR Missing Completely At Random / Missing At Random — Statistical assumptions about the mechanism of missing data in datasets.

HIPAA Health Insurance Portability and Accountability Act — U.S. legislation that provides data privacy and security provisions for safeguarding medical information.

GDPR General Data Protection Regulation — A legal framework that sets guidelines for the collection and processing of personal information from individuals in the EU.

SUS System Usability Scale — A standardized questionnaire used to evaluate the usability of software or devices.

IMDRF International Medical Device Regulators Forum — A global coalition that develops harmonized regulatory approaches for medical devices.

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses — An evidence-based set of guidelines for reporting systematic reviews.

CONSORT-AI / DECIDE-AI Reporting extensions to the CONSORT guidelines — Adapted to trials evaluating AI interventions; emphasize transparency and reproducibility.