AI-Based Non-Invasive VO_2max Estimation and Cardiovascular Risk Stratification: A Systematic Review

P. D. Fonseka and C. Wijesiriwardana

Abstract Cardiovascular diseases (CVDs) account for over 17 million deaths per year, thus representing the most significant single cause of death worldwide. One of the main factors that determines the risk of cardiovascular diseases and overall death in a person is cardiorespiratory fitness (CRF), which shows the level of an individual's physical fitness. CPET or cardiopulmonary exercise testing is the most accurate method to evaluate CRF, however, it is a highly invasive, costly procedure that is not appropriate for the screening of vast populations. Through this review, the articles released between the years of 2010 and 2025 that made use of artificial intelligence (AI) and machine learning (ML) approaches for non-invasive VO_2max estimation and cardiovascular risk evaluation were analysed. In order to discuss the prevailing scientific techniques of data collection, feature extraction, model calibration, and validation, the results of twenty-four articles were taken into consideration and subjected to peer review and synthesis. However, existing studies rely on small, homogeneous samples and lack standardized, calibrated risk mapping. The objective of this review is to conceptualize an AI-based model using the combination of demographic, clinical, and wearable sensor data to non-invasively estimate VO_2max and map these estimates onto easily interpretable cardiovascular risk categories.

***Index Terms*— Artificial Intelligence, Cardiorespiratory Fitness, Cardiovascular Risk Stratification, Machine Learning, VO_2max**

I. Introduction

THE World Health Organization (WHO) claims that cardiovascular diseases are the leading causes of death in the world, with about 45 million deaths recorded every year [1]. Cardiovascular diseases refer to any diseases that affect the heart and blood vessels. It includes coronary artery disease, stroke, and heart failure, depending upon the cardiovascular risk factors such as hypertension, diabetes, obesity, and immobility [3]. Given the increasing rate of cardiovascular diseases, there is heightened emphasis on cardiovascular health, prevention, and personalized therapy research. Cardiorespiratory fitness (CRF) has been proven to be one of the strongest indicators of cardiovascular and overall health. It is the capacity of the body to take in oxygen and utilize it during continuous physical activity [3]. VO_2max is the most common measure of cardiorespiratory capacity and fitness. It stands for maximum oxygen consumption/performance level during prolonged intense exercise [6], [3]. VO_2max is frequently viewed as the leading clinical sign of fitness and longevity because of its high inverse correlation with heart diseases and mortality [5], [6]. Cardiopulmonary exercise testing (CPET) is a controlled maximum or symptom limited graded exercise test that measures gas exchange (oxygen uptake and carbon dioxide production) breath by breath in order to obtain VO_2max and assess the cardiorespiratory system directly [6], [3].

However, despite being the "gold standard" for determining VO_2max, CPET is still not an effective method for screening the entire population, especially the high risk ones, because it is an invasive and expensive procedure [6], [3]. Hence, non-invasive methods have already been developed to get rid of these disadvantages and to estimate VO_2max by considering demographic, physiological, and lifestyle factors which do not require maximal exertion or gas exchange testing. Methods that facilitate the use of submaximal protocols and predictive algorithms not only adhere to ethical safety standards by reducing risks but also make research more practical by needing less advanced laboratory facilities. Modern practice for non-invasive assessment involves using demographic and physical attributes as well as data from deregulated sensors like heart rate monitors and accelerometers, including those that can generate photoplethysmogram (PPG) data. AI and ML advances allow estimation of VO_2max through the continuous and non-invasive methods of analysing physiological data patterns in a safe way, hence assisting in the large-scale monitoring of population health. Evidence from the work of Beltrame et al. [8] and Liu et al. [9] shows that the Random Forest (RF), Support Vector Machines (SVM), and Deep Neural Networks (DNN) models can analyze the wearable and lifestyle data to attain prediction accuracies that are on par with laboratory based CPET. Consequently, the use of multi-sensor physiological signals,

P.D. Fonseka is with the School of Computing, Asia Pacific Institute of Information Technology, Colombo, Sri Lanka.
(e-mail: parofonseka@gmail.com)

C. Wijesiriwardana is a Senior Lecturer at Department of Information Technology, Faculty of Information Technology, University of Moratuwa, Sri Lanka. (e-mail: chaman@uom.lk)

ISSN: 2961-5992

including PPG, ECG, and accelerometry, has made it possible to have fitness estimation be continuous and real time [10], [11].

The technological advancements have not yet occurred that could bridge the gap between the computational VO_2max estimation and the clinical cardiovascular risk evaluation. Most of the AI and ML models have only been constructed around their precision (e.g. R^2 and RMSE) and have not tried to relate their findings to the medical frameworks that are already in place. The work of Kaminsky et al. [12], Ross et al. [3], Mandsager et al. [6], and others has led to clinical investigations that have played an important role in the development of standardized reference systems. These reference systems incorporate the FRIEND Registry, which displays VO_2max percentiles according to age and sex, and the Mandsager mortality bands that correlate graded fitness categories with mortality differences. However, these frameworks remain unintegrated within AI-based predictive models, thus creating a gap between data based estimates and clinical decision making. The review fills the gap by suggesting an AI-based conceptual framework that merges non-invasive VO_2max evaluation and verified cardiovascular risk thresholds. The suggested framework essentially connects the AI-generated VO_2max estimates with the recognized cardiovascular risk levels, thus changing the model's output directly into the risk categories. Consequently, it depicts a route that makes it possible for the computational predictions to be converted into clinical interpretation, ethically used in large scale cardiovascular screening as a tool. Therefore, the main objective of this review is to examine how AI and ML approaches can be used to estimate VO_2max non-invasively and how these estimated values can be translated into clinically meaningful cardiovascular risk categories. This review makes three main contributions. First, it synthesizes existing AI and ML approaches used for VO_2max estimation from demographic, physiological, lifestyle, and wearable sensor data. Second, it identifies key methodological requirements, including preprocessing, feature engineering, validation, calibration, and explainability, that are necessary for reliable and generalizable VO_2max prediction. Third, it proposes a conceptual AI-based framework that links predicted VO_2max values with standardized cardiovascular risk interpretation using reference systems such as the FRIEND Registry and Mandsager's mortality-based fitness categories. This contribution distinguishes the present review from previous studies that mainly focus either on VO_2max prediction accuracy or cardiovascular risk assessment separately. Hence, the review is guided by three research questions (RQs):

- RQ1: How can AI and ML approaches utilize non-invasive demographic, physiological, and wearable sensor data to estimate VO_2max?
- RQ2: What data preprocessing, feature engineering, and model calibration strategies optimize performance and generalizability of AI-based VO_2max prediction models?
- RQ3: What gaps and limitations exist in aligning VO_2max estimates with standardized cardiovascular disease (CVD) risk bands?

II. Literature Review

This literature review is an intersectional study involving CRF, VO_2max, AI, and cardiovascular risk assessment. It commences with observations that CRF and VO_2max are the parameters of cardiovascular health and mortality clinically accepted and established through epidemiological studies and benchmarking frameworks such as the FRIEND Registry. The VO_2max estimation approaches are then segregated into maximal, submaximal, and non-exercise, with a strong focus on the safety, scalability, and AI compatibility of the submaximal and non-exercise methods. Further, it discusses how AI/ML algorithms are developed toward estimating VO_2max from demographic, physiological, and wearable sensor data, highlighting a detailed explanation of some comparative advantages between Random Forests and deep neural network methods. Then, this review explains VO_2max predictions being related to main cardiovascular risk stratification models such as Mandsager's mortality bands and FRIEND percentiles, adding a significant justification for clinical application of AI in these predictions. The review then closes with literature gaps identified, including data availability, underutilization of accidental applications of probabilistic calibration, and absence of clinical integration, which demand a sound AI-based conceptual framework linking prediction performance to pragmatic clinical risk stratification.

A. Cardiorespiratory Fitness (CRF) and VO_2max as Clinical Vital Indicators

Cardiorespiratory fitness (CRF), which is defined by the maximum oxygen consumption (VO_2max), is an essential predictor of heart disease and all causes of death [3]. As the AHA (American Heart Association) considers CRF a "clinical vital sign," its predictive power for health outcomes is on par with blood pressure, cholesterol, and glucose [3]. A thorough meta-analysis reported that the increase of 1-MET in CRF leads to a reduction of about 13% in total mortality and 15% in heart disease events [4]. Some clinical records like the FRIEND Registry [12] have provided value for VO_2max in each sex and age group, grading people into the categories of very poor, poor, fair, good, or excellent fitness. The CRF provides a clinical benchmark at these standard points for interpretation and comparison of population health; however, these are not integrated into scalable, non-invasive AI estimation methods that could make such benefits accessible to the general population.

B. Approach Classification: Maximal, Submaximal, and Non-Exercise Methods

CRF can be estimated through three main methodological approaches:

- Maximal protocols: Measuring VO_2max using a CPET remains the preferred methodology whereby the gas exchange at the lungs is measured during exercise until exhaustion. However, it is time-consuming, invasive, and impractical for older or high-risk populations [1], [3].

ISSN: 2961-5992

• Submaximal protocols: The tests estimate VO_2max based on heart rate response during moderate exercise levels; these tests include YMCA cycle ergometer, six-minute walk, or 20-m shuttle run test [6], [7]. These tests are safer but less precise.

• Non-exercise models: These predict VO_2max from accessible variables like age, sex, BMI, physical activity level, and resting heart rate [8], [9].

The aim of this review is to discuss various safety methods of estimating VO_2max using submaximal and non-exercise approaches that are cost effective and compatible with an AI/ML system for input. These methods not only comply with the ethical research standards but also reduce the risk to participants and the inconvenience brought about by the physiological strain of the maximal exertion test such as CPET. The latter may be unmanageable for older adults, sedentary individuals, or patients with cardiovascular risk factors [7], [3]. These protocols, which do not require maximal effort or exercise, are easier to organize and more affordable in outdoor or community settings since they utilize minimal equipment and are suitable for large populations [9], [13]. Additionally, these techniques are naturally aligned with AI-based modeling as they offer lifestyle data that are well structured, low risk, and repeatable, such as those for large population groups and wearable devices that are scalable [8], [11]. However, mostly due to the submaximal, non-exercise inputs being favored for deployment, we yet maintain validity by referring to the CPET gold standard. We do this by calibrating predicted VO_2max to CPET derived norms (FRIEND percentiles) and stratifying risk via Mandsager bands; thus, maximal tests are labelled as reference and clinical benchmarks, not as prerequisites for screening. Therefore, this does not only uphold non maleficence but also makes possible equitable, population-level screening through digital health platforms.

C. Progress in Artificial Intelligence and Machine Learning Techniques for VO_2max Estimation

The prediction of VO_2max with the use of AI and ML has gone through a huge developing process and finally the classical regression has been left behind by employing the above-mentioned technologies as they are able to model nonlinear relationships and thus are able to remove generalization errors across different and mixed population groups [7], [14], [9]. Non-exercise models that have been trained with population scale data and that make use of demographic and lifestyle features (e.g., NHANES) now come to a point where they can provide a comparable accuracy to laboratory estimates which in turn increases the possibility of low burden screening that can be carried out on a large scale, i.e., Liu et al., (2023) [9]. In fact, similar recommendations from the reviews of methodological studies also emphasized systematic improvements resulting from the use of ML methods as compared to the traditional ones and at the same time pointed out the different datasets and methods used for testing [14]. The use of wearable and sensor-based systems has further increased the performance by being able to record and continuously monitor the physiological processes at rich time series under free living conditions. Supervised models that are trained on data streams of heart rate, accelerometry, and related sources have been determining the dynamics of oxygen uptake with a high degree of similarity to lab measures [8], [10]. Using minimal sensor setups combining heart rate and step count resulted in mean errors close to 4 mL/kg/min which suggests that such a system might be very useful and practically implemented in field settings [11]. The recent signal processing innovations that have successfully reduced motion artifacts in PPG and the reconstruction of reliable heart-rate traces have helped VO_2max estimation to a great deal [15], [16]. Also, specially designed devices and optical systems have presented oxygen use estimation based on ML significantly indicating maturing of technology at the device level [17].

Tree ensemble methods such as random forests and gradient boosting, along with neural networks like CNNs and LSTMs for 1-D physiological data, reliably surpass linear models regardless of the dataset selected [8], [9], [10]. Clinical and surgical settings have also been beneficiaries: machine learning based on patch ECG achieved the same accuracy as the reference standards in surgical cohorts by reporting zero mean bias [18], while the combination of regression strategies based on statistical data plus extrapolation led to VO_2max estimation that was more precise in the laboratory testing [19]. External validation studies conducted on consumer grade devices have shown both the positive aspects and the limitations of the technology. For instance, independent studies performed on wrist devices have found that the overall accuracy is moderate and that the potential applications are unevenly distributed among various age and fitness groups. Thus, it becomes essential that the model is device sensitive and subgroup analyses are performed [20], [21], [22]. The non-exercise ML model for older adults is developed and transferred with reasonable accuracy but still heavily based on the demographics, protocol consistency, and outcome definitions [22]. In clinical risk populations, it has been noted in studies that the CRF estimates derived from wearables are valid in comparison with CPET, thereby making it possible to carry out targeted screening [23].

Preprocessing methods that are robust (artifact handling, scaling, and windowing), feature design (e.g., HRV, cadence variability), and leakage safe evaluation (subject-level splits, stratified cross-validation, and external test sets) should be considered as a basis for credible performance claims [7], [14], [10]. Reviews are still pointing out inconsistencies in dataset curation and benchmarking which are making it difficult to carry out direct comparisons thereby making it necessary for standardized pipelines and reporting [13], [14]. In the field of applied cardiology, the role of AI is becoming more important and it is being placed in the context of precision medicine goals, linking models to actionable endpoints and clinical workflows [24], [6]. One important aspect is that calibration, uncertainty communication, and clinical translation are still less reported compared to error metrics such as R^2 or RMSE. A recent study emphasizes that it is necessary to evaluate and, if relevant, to calibrate predictions across subgroups and devices in order to avoid systematic bias when deploying [14], [22]. To link the

ISSN: 2961-5992

algorithm to the clinic, it is suggested that the VO_2max estimates should be adjusted to age/sex percentiles (e.g., FRIEND) and then correlated with the validated fitness mortality bands (e.g., Mandsager) so that the outputs will correspond with the cardiovascular risk categories that are generally used in practice [5], [3]. By connecting VO_2max, percentile, and risk band, the process becomes easier to interpret and allows evaluation at the band level, such as agreement and reclassification, in addition to the usual CPET agreement analysis [9], [10]. In the end, complete systems should present the intended application, efficiency in various subcategories, and the limitations, particularly when the consumer platforms or vendor algorithms are part of the process. The product's market acceptance is a sign of its practicality, however, proprietary tiers have to be validated against laboratory standards independently before recommending them for preventive or clinical decision-making [21], [22].

TABLE I
VO2 MAX FITNESS CATEGORIES BASED ON PERCENTILE RANGES AND CVD RISK

Percentile Range	Fitness Category	Cardiovascular Risk Band	Relative Mortality Risk
≥ 97.7	Elite	Minimal	Lowest
75–97.6	High	Very low	Low
50–74	Above average	Low	Lower
25–49	Below average	Moderate	Higher
≤ 24	Low	High	Highest

D.Linking VO_2max to Cardiovascular Risk Stratification

VO_2max is not only an indicator of physical fitness but also a clinically accepted biomarker for cardiovascular risk and mortality [3], [4]. The American Heart Association (AHA) identifies cardiorespiratory fitness (CRF) as a "clinical vital sign" because of its strong predictive value for early disease detection and preventive health assessment [3]. Kodama et al. [4] reported that, among more than 100,000 individuals, each 1-MET increase in CRF was associated with a 13% reduction in all-cause mortality and a 15% reduction in cardiovascular events. This supports the use of VO_2max as a quantifiable long-term survival indicator. Similarly, Mandsager et al. [5] classified CRF into mortality-based fitness categories using exercise treadmill testing (ETT) data from more than 122,000 adults. The expected MET values were calculated as follows:

For Men:

$$Estimated\ METs = 18.7 - (0.15 \times Age) \quad (1)$$

For Women:

$$Estimated\ METs = 14.7 - (0.13 \times Age) \quad (2)$$

Achieved METs were then normalized against expected values by calculating the Performance Index:

$$Performance\ (\%) = (Achieved\ METs \div Estimated\ METs) \times 100 \quad (3)$$

Subjects were then classified into fitness percentile bands as shown in Table 1, which associates CRF categories with cardiovascular risk levels and relative mortality risk.

This stratification framework provides a measurable way to interpret VO_2max in relation to mortality and disease risk. Mandsager et al. [5] showed that individuals in the lowest fitness category had substantially higher mortality risk compared with those in the highest fitness group. These findings are further supported by the FRIEND Registry [12], which provides age- and sex-specific VO_2max percentiles for population-level comparison. Therefore, combining FRIEND Registry percentiles with Mandsager's mortality-based fitness categories provides a dual clinical framework for linking VO_2max with cardiovascular risk stratification.

However, this type of integration is rarely applied in computational VO_2max prediction research. Recent advancements in artificial intelligence have created new opportunities to bridge this gap. By aligning AI-predicted VO_2max values with FRIEND percentiles and Mandsager's mortality-based fitness categories, researchers can transform algorithmic outputs into medically interpretable cardiovascular risk categories. This can support preventive health, exercise safety, and large-scale digital screening. Nevertheless, consumer-grade devices and commercial fitness platforms, such as Apple Watch, Garmin, and DexaFit, often rely on proprietary algorithms that require further independent clinical validation before being used for clinical decision-making [20], [21], [25]. The INTERLIVE consortium also recommends that wearable-based fitness and physiological estimates should be validated against laboratory reference standards before clinical interpretation [28].

ISSN: 2961-5992

TABLE II
VO2 MAX RANGES AND CORRESPONDING FITNESS LABELS FOR MEN AND WOMEN

Risk Band	VO_2max (Men) (mL/kg/min)	Label (Men)	VO_2max (Women) (mL/kg/ min)	Label (Women)
High risk	13–16	HEART RISK	11–14	HEART RISK
Low risk	26–36	LOW RISK	22–33	LOW RISK
Active	36–45	ACTIVE	33–42	ACTIVE
Athlete	45–56	ATHLETE	42–50	ATHLETE
Pro	56–66	PRO	50–60	PRO
Elite	≥ 66	ELITE	≥ 60	ELITE

The commercially available VO_2max categories listed in Table 2 are based on DexaFit's public fitness assessment framework [25]. DexaFit offers VO_2max testing and fitness classification directly to consumers, supporting personalized exercise planning. The table illustrates consumer-level segmentation of fitness levels, including Active, Athlete, Pro, and Elite, based on estimated VO_2max ranges for men and women. Although these classifications show the increasing commercialization and accessibility of CRF assessment, they are not equivalent to clinically standardized references such as the FRIEND Registry [12] or Mandsager's mortality-based fitness categories [5]. This difference highlights the gap between consumer-level fitness analytics and clinical cardiovascular risk assessment. Therefore, this review recommends that AI-predicted VO_2max values should be systematically mapped to FRIEND percentiles and Mandsager mortality bands to support a clinically interpretable and standardized risk stratification pipeline for scalable, non-invasive cardiovascular health monitoring.

E. Gaps and Limitations in Existing Literature

While technology is advancing swiftly, some restrictions remain, such as:

- Lack of integration: Various models based on ML have confirmed their results by laboratory benchmarks but only a few have mapped the results to the category related to clinical risk [9], [8].
- Limited dataset availability: Few publicly accessible datasets link VO_2max with long-term cardiovascular disease outcomes, restricting the ability to externally validate the models [9], [22].
- Underuse of probabilistic calibration: Only a few models have incorporated the uncertainty quantification through the method of Platt scaling or isotonic regression, which are very necessary for clinical trust [16], [17].
- Bias and generalizability: The performance of the model is different for different age groups, sexes, and ethnicities because of the imbalances in the sample [21], [22].
- Research fragmentation: The ongoing research is very limited when it comes to cross-domain synthesis as the studies are either predicting VO_2max (AI/ML focused) or assessing CVD risk (clinical focus) [3], [9].

By proposing an AI-integrated risk stratification framework that adheres to the predictive modeling requirements of FRIEND and Mandsager, the review aims to bridge the gaps [12], [5]. The proposed model would therefore allow for non-invasive cardiovascular risk assessment that is calibrated probabilistically, thus enhancing the process's interpretability, clinical importance, and global scalability [16], [17].

III. RESEARCH METHODS

A. Review Design and Framework

This work was sorted as a Systematic Review following PRISMA 2020 guidelines for transparency, reproducibility, and replicability [17]. The review protocol included important aspects such as research questions, eligibility criteria (PICOS), information sources, search strategy, screening workflow, and the data extraction schema before commencing with actual searching processes. According to the protocol, the review considered three research questions: RQ1: How do various AI and ML methods use non-invasive demographic, physiological, and wearable-sensor inputs to estimate VO_2max? ; RQ2: What data preprocessing, feature engineering, and model calibration approaches best prepare AI-based VO_2max prediction models in terms of performance and generalizability? ; RQ3: What are the gaps and limitations within the current literature, and how can VO_2max estimations be standardized into cardiovascular disease (CVD) risk categories? To enhance reproducibility, the review followed a predefined workflow consisting of database searching, duplicate removal, title and abstract screening, full-text eligibility assessment, data extraction, and thematic synthesis. Each selected study was reviewed using the same extraction fields, including population characteristics, input features, AI/ML method, validation approach, performance metrics, and risk-mapping strategy. This ensured that all studies were assessed consistently and that the synthesis could be repeated using the same methodological process.

B. Eligibility Criteria (PICOS)

Inclusion/exclusion criteria for review followed the PICOS framework [17]. The population (P) comprised human participants who were healthy, athletic, or clinical, with VO_2max determination as an outcome. The study considered different indexes (I) that included AI, ML, DL based, wearable and sensor based approaches to non-invasive VO_2max estimation. Comparisons (C) were considered between those using direct VO_2max from CPET or submaximal protocols that are validated. Outcomes (O) reflected gains in model metrics (e.g., RMSE, MAE, SEE, R^2) and the ability of VO_2max to correlate with cardiovascular risk thresholds such as percentiles

ISSN: 2961-5992

or mortality bands. Study design (S) included experimental, observational, or validation, while reviews were excluded from synthesis but referenced for context. Among other criteria considered for inclusion were: (1) a publication window of 2010-2025 to ensure relevance with the AI and wearables; (2) selection of English language studies only to keep terminological consistency; and (3) selections of full text articles for appropriate assessment of methodology and results. These criteria, taken together, ensured that recent, methodologically sound, and rigorously synthesized studies would be considered for review [17].

B. Information Sources

To ensure diversity in the studies handled within this review and to promote multi topic coverage, a number of databases were employed, such as PubMed, IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, Google Scholar, and Semantic Scholar. These databases were considered to cover a larger array of topics ranging from biomedical sciences, computational modeling, wearable technologies, and cardiovascular health analytics. All of these are important factors for estimating VO_2max in prediction models based on AI. The pooling of more databases also prevented selection bias and provides related peer reviewed literature to ensure that the results proposed are as credible and as strong as possible.

C. Search Strategy

Table 3 encapsulates the search strategy, centering on the four major concept clusters of VO2max / CRF, AI and machine learning, data acquisition modality, and cardiovascular risk, to maximize the retrieval of relevant studies across different multidisciplinary databases. Each concept was widened by Boolean operators and its synonyms so that the query could capture the different terminologies used in the clinical and computational research contexts. By including terms related to physiology and AI (e.g., maximal oxygen uptake, deep learning, non-exercise, risk stratification), this approach ensured the maximum sensitivity for the retrieval of documents, while staying true to the core concept of AI-based cardiorespiratory fitness estimation and cardiovascular risk assessment.

TABLE III
SEARCH CONCEPTS AND EXAMPLE TERMS USED IN DATABASE QUERIES

Concept	Example Terms
VO_2max / CRF	"VO_2max" OR "maximal oxygen uptake" OR "oxygen consumption" OR "cardiorespiratory fitness"
AI / ML	"Machine learning" OR "artificial intelligence" OR "deep learning" OR SVM OR "random forest" OR "neural network*"
Modality	submaximal OR "non-exercise" OR wearable* OR ECG OR PPG OR SCG OR accelerometer* OR smartphone
Risk	"Risk stratification" OR percentile* OR threshold* OR "mortality risk" OR "cardio

The search was conducted using Boolean combinations of the four concept clusters shown in Table 3. An example search string used across databases was: ("VO_2max" OR "cardiorespiratory fitness" OR "maximal oxygen uptake") AND ("machine learning" OR "artificial intelligence" OR "deep learning" OR "random forest" OR "support vector machine") AND ("wearable" OR "non-exercise" OR "submaximal" OR "ECG" OR "PPG" OR "accelerometer") AND ("risk stratification" OR "mortality risk" OR "percentile" OR "cardiovascular risk"). The search was limited to English-language studies published between 2010 and 2025. Where databases allowed filtering, publication year, article type, and full-text availability filters were applied. Google Scholar and Semantic Scholar were used mainly to identify additional relevant studies and citation links.

D. Study Selection and Screening Flow

Depicted is the PRISMA 2020 flow diagram (Figure 1) for the literature selection process of this systematic review [17]. Across seven electronic databases 200 records were retrieved from: PubMed, IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, Semantic Scholar, and Google Scholar. Thirty-five duplicate records were removed, while the remaining 165 records were subjected to screening based only on titles and abstracts. One hundred and five were excluded because they did not fall under the relevance criteria. Sixty full text articles were sought to be retrieved; for three articles, retrieval was simply not possible. Fifty-seven full text articles were then determined eligible based on set PICOS criteria. Continuing with the eligibility discussion, 94 studies were excluded for various reasons, chief among them were no model training, invasive CPET only studies, or inadequate clinical association, and Non-English language papers. 24 studies were finally identified as eligible for the final qualitative synthesis around which thematic discussions involving non-invasive VO_2max estimation and cardiovascular risk stratification were structured. All of these studies are referenced in the References section. These 24 studies were picked to ensure there was adequate representation of methods spanning regression based, neural networks and hybrid AI approaches. The Final set proved their balance by way of equal representation in clinical and non-clinical populations and this has given comparative insights into real world applicability. This selection thus forms the backbone when we captured in the results and discussions, making the subsequent synthesis both reliable and exhaustive.

ISSN: 2961-5992

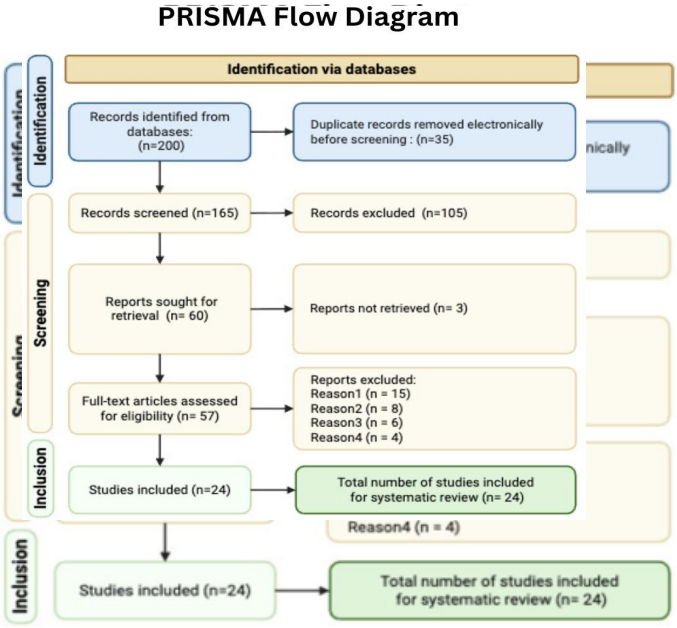

Fig. 1. PRISMA 2020 Flow Diagram showing study identification, screening, and inclusion process.

The overall inclusion and exclusion criteria are summarized in Table 4 and were formulated following the predefined PICOS framework, ensuring the methodological rigor and pertinence of the study toward the research objectives. Included studies were ones involving humans, published in English from 2010 to 2025, and used any kind of AI/ML based or statistical technique for the prediction of VO_2max from any of the three data sources: wearable, sensor, or non-exercise. Studies that linked these model predictions to clinical risk stratification frameworks were preferred over others. Exclusions were non-human studies, non-English articles, pure CPET analysis without any machine learning part, review articles, or those for which full texts were inaccessible. This systematic approach to the screening ensured that this review synthesized only experimentally grounded studies that could be compared by method, hence increasing the reliability, validity, and clinical interpretability of the final study.

TABLE IV
INCLUSION CRITERIA AND EXCLUSION CRITERIA VO_2MAX ESTIMATION AND CARDIOVASCULAR RISK STRATIFICATION

Inclusion Criteria	Exclusion Criteria
Human studies, English language, published between 2010–2025	Non-human or non-English studies
AI/ML-based or statistical VO_2max estimation approaches	Invasive CPET only studies with no model training
Wearable, sensor based, or non-exercise prediction models	Review papers, editorials, letters, or studies without full-text access
Models linked to clinical risk stratification or VO_2max-based thresholds	Pediatric-only or highly domain-specific studies without generalizability

E. Data Extraction

TABLE V
DATA EXTRACTION SCHEMA

Field	Description
Bibliography	Authors, year, venue, DOI/URL
Population	N, age, sex, health status
Features	Demographics, HR, HRV, sensors (ECG, PPG, accelerometer), submaximal test type
Method	Algorithm (Regression, RF, SVM, ANN, CNN, LSTM)
Validation	Train/test split, cross-validation, external validation
Metrics	RMSE, MAE, SEE, R^2, bias/limits of agreement
Risk Mapping	Percentiles (FRIEND, ACSM) or Mandsager thresholds
Limitations	Confounders, missing data, biases

As outlined in Table 5, the process of data extraction was based on a standardized building plan that guaranteed consistency and repeatability across different studies. Each entry consisted of bibliographic metadata entries, participant characteristics, input features, machine learning methods, validation approaches, and performance metrics. From a clinical viewpoint, there was a separate risk mapping field for cardiovascular risk, either through FRIEND percentiles [12] or Mandsager mortality timepoints [5]. This standardized extraction framework allowed for the systematic contents' comparison of algorithms, datasets, and outcome measures, hence providing a transparent synthesis of evidence on AI based VO_2max estimation and cardiovascular risk stratification.

Figure 2 describes the proposed modular AI-based workflow for non-invasive VO_2max estimation and cardiovascular risk stratification. The workflow begins with user input collection, including demographic data such as age, sex, height, and weight; lifestyle information such as physical activity level; and wearable sensor data such as heart rate, ECG, PPG, accelerometer, or step-count data where available. These inputs are then processed through data cleaning, missing value handling, normalization, outlier detection, and feature engineering. Engineered features may include BMI, resting heart rate, heart-rate variability, activity patterns, cadence variability, and sensor-derived physiological indicators.

After preprocessing, the processed feature set is passed into a machine learning model such as Random Forest, Support Vector Machine, Artificial Neural Network, CNN, or LSTM to estimate VO_2max. The predicted VO_2max value is then normalized using age- and sex-specific reference values such as the FRIEND Registry [12]. Finally, the normalized VO_2max output is mapped to cardiovascular risk categories using validated fitness-risk frameworks such as Mandsager et al. [5]. The final system output includes the estimated VO_2max, fitness category, cardiovascular risk band, and confidence or uncertainty score. This workflow improves reproducibility by

ISSN: 2961-5992

clearly defining each stage of the AI pipeline from input collection to clinical risk interpretation.

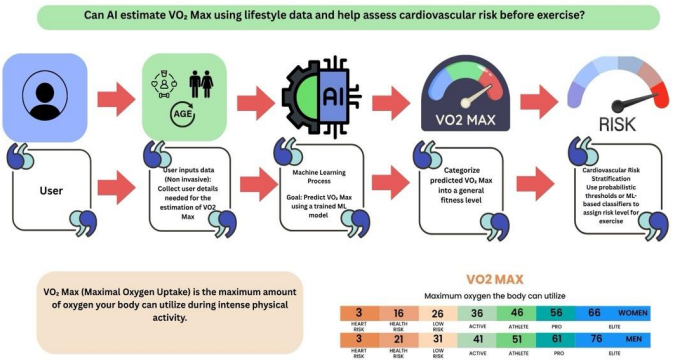

Fig. 2. The system flow suggesting how user demographic and lifestyle inputs are processed through a machine learning model to predict VO_2max

IV. Results and Discussion

The systematic review had selected 24 main works and publications between the years 2010 and 2025 after applying the PRISMA screening method. The findings are structured concerning the three primary research queries (RQ1 to RQ3), which delve into diverse issues of non-invasive VO2max estimation and cardiovascular risk stratification by means of AI and ML techniques. The systematic review was a comprehensive overview of research through AI and ML methods in VO2max estimation during 2010–2025 period which led to grouping the findings under three primary research questions (RQ1–RQ3). A brief description of the principal methods, datasets, findings, and limitations is provided in the subsections below.

A. RQ1. How can AI and machine learning approaches effectively utilize non-invasive demographic, physiological, and wearable sensor inputs to estimate VO_2max?

To answer the research question, a systematic review was conducted, and studies were grouped based on non-invasive VO_2max estimation methodologies, namely demographic-based, physiological signal based, and hybrid sensor-integrated models. Earlier statistical regression approaches mainly relied on demographic variables including age, sex, body mass index (BMI), and resting heart rate. Although these models provided scalability and interpretability, their predictive performance was constrained due to the exclusion of dynamic physiological features [7], [9]. Through the application of machine learning (ML) methods, non-invasive VO_2max estimation has reached significantly higher accuracy. Models such as Random Forest (RF), Support Vector Machines (SVM), Gradient Boosting Machines (GBM), and Artificial Neural Networks (ANN) capture nonlinear relationships between demographic and physiological factors. Therefore, hybrid sensor-based architectures ensure higher accuracy. Modern wearable technologies records ECG, PPG, SCG, and movement data to calculate VO_2max, concentrating on certain cardiac and biomechanical signals that relate to oxygen utilization. CNNs and LSTMs have gained much interest during recent years as deep learning architectures for the modeling of physiological data with spatiotemporal dependencies [10], [18]. In Beltrame et al. [8], the wearable-based deep neural network method used heart rate and movement data and almost reached the accuracy of laboratory equipment. Rissanen et al. [18] further confirmed ML VO_2max estimates in populations at cardiovascular risk in a clinical setting. An RMSE of around 3.95 mL/kg/min using inputs from step count and heart rate was recently reported by Neshitov et al. [11], thus confirming minimal sensor setup applicability for field measurements. Still, several issues go on to challenge the discipline. Most datasets available contained fewer than 500 participants, often from healthy populations, limiting external generalization. Protocol variation from study to study suppresses benchmarking. Moreover, despite accurate VO_2max from AI, very few studies tie the outputs to a validated cardiovascular risk category, which compromises clinical interpretation [5], [3].

The paragraph above describes the current situation and future prospects following the development of an AI-powered VO_2max prediction model. Concluding that AI methods might be used in studying nonlinear relationships in data, carrying out computations for probabilistic risk, or generally assisting clinicians in studies. Despite progress, considerable work remains in expanding dataset diversity, developing uniform assessment procedures, and fine-tuning clinical calibration before these applications can be widely implemented. Increasing interpretability through transparent feature attribution and explainability approaches would certainly build trust among clinicians [14], [26], [6]. Hybrid approaches that combine physiological sensing with contextual metadata on lifestyle, sleep, and activity patterns hold great promise for the next generation of personalized fitness and cardiovascular risk prediction [9], [8], [11], [10]. These multidimensional models could finally enable constructing adaptive real-time monitoring systems that evolve along with a person's health profile and hence actualize precision exercise medicine. Essentially, results for RQ1 indicate that AI-based VO_2max-prediction models can yield accurate, noninvasive VO_2max estimates from demographic, physiological, and wearable-sensor inputs, while hybrid sensor systems appear especially well suited to continuous monitoring of the real life situations. At the same time, considerations about clinical viability should follow from a trio of aspects: standardized preprocessing and validation guidelines, thorough external validation, and an explicit mapping of predicted VO_2max values to established cardiovascular risk categories. These topics are further elaborated in Sections B (RQ2) and C (RQ3).

B. RQ2. What are the most suitable data preprocessing, feature engineering, and model calibration strategies for optimizing the performance and generalizability of AI-based VO_2max prediction models?

The reviewed studies are sorted into three major methodological phases of an AI-based VO_2max prediction

ISSN: 2961-5992

pipeline: data preprocessing, feature engineering, and model calibration. These stages are inherent to the accuracy of prediction, such as reduction of overfitting or being generalized to various populations.

The primary purpose of data preprocessing is to ensure that VO_2max models perform reliably and effectively. Data cleaning, normalization, and transformation were highlighted in the literature to handle physiological datasets given that these datasets usually contain noise or missing values [7], [9]. Variables such as heart rate, age, and BMI are usually min/max scaled, or z-score normalized to stabilize the training. Outlier detection techniques such as statistical filters and the interquartile range (IQR) method further improve quality when heterogeneous sources are fused, such as when the sources consist of wearable sensor streams and clinical tests [13], [18]. The newest AI-based workflows are further applying advanced preprocessing to a more diverse kind of signals being received from wearables and IoT devices than mere cutting-edge basics including normalization. The applications of artifact suppression, signal quality indexing, and sliding-window segmentation of ECG and PPG streams are in wide use to ensure that models actually learn from the most physiologically plausible segments and not from noise, hence stabilizing the VO_2max estimation in free-living conditions [22], [25], [21]. In this scenario, advanced non-invasive VO_2max testing protocols meet the highest standards of medical AI, wherein the system is considered safe only with carefully curated high-quality input data.

Feature engineering deals with turning raw inputs into meaningful representations. Recent studies implement a plethora of temporal and frequency domain feature extraction methods such as Fast Fourier Transform (FFT) and wavelet decomposition of HRV and PPG signals to bring out better representations of cardiorespiratory dynamics [8], [10]. Further, recursive feature elimination, LASSO regularization, and principal component analysis are various strategies for feature selection and dimensionality reduction that help reduce redundancy, combat overfitting, and ease computational burden [7], [13]. Recently, there has been a significant influx of AI techniques, with more attention being granted toward representation learning. AI methodologies are mainly considered when models like convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are trained directly on raw or minimally processed physiological time-series data obtained from submaximal tests and wearable devices so that the models can autonomously detect temporal patterns linked to VO_2 kinetics [8], [10], [20]. This modeling surely necessitates less reliance on handcrafted features and bridges the gap between easy measurements of complicated heart and lung responses in the real world. Aligned with the main concept of this review, VO_2max estimation should be fine-tuned, ongoing, and non-invasive, making it hard to distribute for usual cardiovascular risk estimation. The main methods employed from a machine learning engineering perspective for equalizing the output with different input variations and ensuring lack of overfitting with small samples of VO_2max are ensemble methods like Random Forests, Gradient Boosting, or XGBoost as conducted by Abut et al. [7], Ashfaq et al. [4], and Schumacher et al. [19], with LASSO on a regularization basis. Symptomatic of this continuous underestimation of risk on the changing outcome variable among low-fitness or clinical subgroups is sample imbalance, found to be a serious hurdle. Usually, the methods that are pinpointed are resampling, class weighting, or threshold tuning to select more fit or healthier individuals. This is often addressed as a subitem in the general area of cardiometabolic forecasting research [17].

According to Mandsager et al. [5] and Schumacher et al. [19], model calibration involves aligning model forecasts with actual outcomes. This step is critical in the clinical field as the raw probabilities of an event should represent real world occurrences. While this calibration step is seldom done, deep learning architectures give remarkably good predictive results. Unlike other calibration algorithms, Platt Scaling and Isotonic Regression have been deemed ideal for calibration of VO_2max prediction outputs to make them reliable for use in clinical and fitness applications. Moreover, validation protocols, and especially stratified k-fold cross-validation, are often applied in order to ensure the sample representation across fitness categories and demographic subgroups, thus decreasing bias and leading to better generalization [9], [8]. When it comes to making a method applicable to different situations, the importance of external validation through different cohorts, devices, and acquisition protocols has been demonstrated by a number of studies, considering deployment as a domain shift problem rather than just a test-split exercise [19], [24], [23]. Recalibration and device-specific error profiling play a significant role in case of the consumer wearable devices receiving signals from research-grade sensors, and thus, bringing the argument back to the current state that $\dot{V}O_2$max estimation pipelines should be adapted according to the variability of real-world hardware. When using AI for healthcare, it is essential that calibrated VO_2max estimates are accompanied by measures of their uncertainty, like confidence intervals or reliable confidence scores, to support safe clinical decisions, triage processes, and exercise assessments [16], [12]. One of the ways to do this is through the use of Explainable AI (XAI) techniques, which may include feature importance analysis for ensembles or attribution maps for neural networks, to show whether the models depend on physiologically correct features (such as age, resting heart rate, and activity volume) and to assist the doctors in confirming and monitoring the recommendations based on VO_2max [4], [15]. These techniques form the technical foundation for the subsequent clarification of the research question regarding the reliable cardiovascular risk bands mentioned in RQ3. In summary, the combination of organized preprocessing techniques such as normalization, artifact elimination, and outlier management, along with knowledge driven feature creation, representation learning, and clear calibration and validation processes, greatly enhances the preparedness of AI-based VO_2max estimation models for clinical application and widespread deployment. The future research should focus on common, publicly available pipelines

ISSN: 2961-5992

and benchmark datasets that incorporate these AI principles, strong preprocessing, ensemble and deep learning architectures, uncertainty aware calibration, and explainability to facilitate reproducible comparison of algorithms and smooth inclusion of VO_2max prediction into approved cardiovascular risk stratification frameworks.

C. RQ3. What gaps and limitations exist, and how can VO_2max estimates be standardized into CVD risk bands?

Both RQ1 and RQ2 highlight the element that AI/ML models are capable of estimating VO_2max non-invasively with utmost precision if demographic, physiological, and wearable sensor inputs are used, and that there is a gradual but certain establishment of the robust preprocessing, feature engineering, and validation strategies. The studies that do push these predictions into clinically standardized risk categories such as age- and sex-specific VO_2max percentiles or mortality-based fitness bands are very limited [12], [6], [3]. Consequently, the vast majority of AI-based VO_2max estimators are still considered as technical tools rather than integral parts of an entire cardiovascular risk stratification pipeline. A major drawback of this research area is the limited availability of open-access datasets that simultaneously include CPET-derived VO_2max, non-invasive AI-predicted VO_2max, and longitudinal cardiovascular outcomes. Datasets of this nature are crucial for external validation and assessment of risk-based thresholds, such as the gradients of incident cardiovascular events or mortality across different fitness levels. Without them, these activities are likely to be constrained [9], [19]. The shortage of data directly hampers the creation of models that can easily be transferred from one situation to another and prevents the developers from making a strong claim as to whether or not AI-derived VO_2max can safely guide clinical or preventive decision-making at a large scale. On the methodological side, calibration and risk standardization are still not fully exploited. Though RQ2 points out the significance of techniques such as Platt scaling, isotonic regression, and stratified validation as the best practices, still very few VO_2max modeling studies provide calibrated probability estimates, prediction intervals, or band-level agreement with the current clinical frameworks [16], [4]. In addition, the cooperation with reference systems such as FRIEND, ACSM, or Mandsager's mortality-based fitness bands is often not implemented at all, which means that model outputs are not translated into low-, moderate-, or high-risk strata that doctors can easily understand [6], [3]. The absence of a uniform mapping of the outputs of the various models to the risk categories that the clinicians use is the main translational bottleneck separating computational performance from actionable cardiovascular risk assessment.

TABLE VI
SUMMARY OF GAPS IDENTIFIED

Theme	Gap Identified
Dataset availability	Lack of open VO_2max + cardiovascular outcome datasets
Calibration	Limited use of probabilistic calibration methods (e.g., Platt scaling, isotonic regression)
Validation	Lack of alignment with FRIEND and ACSM cardiovascular risk frameworks
Interpretability	Limited model explainability and clinical interpretability

D.Identified Limitations

The employing of AI and ML based techniques for the VO_2max prediction is indeed very promising, but at the same time, there are several important limitations that restrict its use in clinics and in real-world situations. Initially, many studies are carried out with small, homogeneous groups mainly consisting of healthy adults or athletes, which complicates the ability to make broad inferences about older, multi-diseased, and high-risk patients [8], [9], [19]. Secondly, there are no large-scale longitudinal datasets available that would include CPET-derived VO_2max, AI-predicted VO_2max, wearable-derived features, and subsequent cardiovascular outcomes, which would otherwise be a great limitation for robust external validation and outcome based calibration [16], [6]. Thirdly, when the feature sets, preprocessing pipelines, exercise protocols, and evaluation metrics are not unified, it leads to making cross study comparisons and meta-analysis very complicated and thus prevents the setting of standardized benchmarks [4], [5]. Fourth, the majority of the models emphasize point estimates and accuracy metrics (R^2, RMSE) while hardly ever reporting confidence intervals, probabilistic calibration, or explainability outputs, all of which are indispensable for safe clinical decision support [15], [16]. Lastly, and most importantly, there is no widely accepted methodology to transform AI-predicted VO_2max values into recognized cardiovascular risk categories like age- and gender-based percentiles or mortality based fitness bands which constitutes a major impediment in bridging the gap between technical performance and clinically interpretable risk stratification [3], [6].All these limitations together support the requirement of the proposed conceptual framework that focuses on the standardization of pipelines, calibration of outputs, and systematic alignment of VO_2max estimates with established cardiovascular risk categories.

V.CONCLUSION AND FUTURE DIRECTIONS

This systematic review assessed an overview of non-invasive AI/ML-based VO_2max estimation and its potential use in cardiovascular risk stratification schemes between 2010 and 2025. A total of 24 relevant studies were selected using a PRISMA guided method, encompassing clinical, athletic, and general populations. Evidence shows that machine learning and deep learning models particularly Random Forests, Support Vector Machines, Gradient Boosting, and neural networks like ANNs and LSTMs consistently outperform traditional regression methods when applied to data obtained through non-exercise, wearable, or submaximal protocols. These advancements have set the stage for scalable, non-invasive VO_2max estimation suitable for large-scale population health monitoring. However, the integration of these methods into clinical practice has been limited since few studies have worked

ISSN: 2961-5992

towards relating predicted VO_2max values to accepted risk classification systems like age- and sex-adjusted percentiles or mortality bands based on fitness. This represents a gap between the computational prediction and the actual clinical decision thresholds.

To bridge this gap, several future directions are emphasized. First, diverse and longitudinal datasets should be developed and openly shared datasets that combine CPET derived VO_2max, wearable sensor data, and cardiovascular outcomes. Second, standardized feature engineering practices, preprocessing pipelines, and evaluation protocols that will ensure reproducibility and fair benchmarking among models and cohorts are required. Third, practical calibration methods such as Platt scaling or isotonic regression and probabilistic risk bands with confidence intervals would help make the model more reliable and safer for clinical and preventive settings. Thus, this review favors a robust AI-based pipeline for VO_2max estimation, which includes preprocessing, explainability, uncertainty quantification, and external validation, to make VO_2max a clinically meaningful, scalable tool for cardiovascular risk stratification and preventive care, away from being a predominantly.

ACKNOWLEDGMENT

The author also wishes to thank the APIIT School of Computing for guidance, academic support, financial support and resources in conducting research successfully and Ms. Thenuri Hettiarachchi for her invaluable guidance throughout the study.

REFERENCES

[1] Lang, J. J., et al. (2024). Cardiorespiratory fitness is a strong and consistent predictor of risk across many mortality outcomes in the adult general population: Overview of systematic reviews. British Journal of Sports Medicine, 58(10), 556–564. https://doi.org/10.1136/bjsports-2023-106915

[2] Caplan, E. O., Hayden, J., Pimple, P., Luthra, R., Prewitt, T., Chiguluri, V., Kattan, M. W., Goss, A., & Harvey, R. (2021). Cardiovascular risk prediction model and stratification in patients with type 2 diabetes enrolled in a Medicare Advantage plan. Journal of Managed Care & Specialty Pharmacy, 27(11), 1579–1591. https://doi.org/10.18553/jmcp.2021.27.11.1579

[3] Ross, R., Blair, S. N., Arena, R., Church, T. S., Després, J.–P., Franklin, B. A., et al. (2016). Importance of assessing cardiorespiratory fitness in clinical practice: A case for fitness as a clinical vital sign. Circulation, 134(24), e653–e699. https://doi.org/10.1161/CIR.0000000000000461

[4] Kodama, S., et al. (2009). Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: A meta-analysis. JAMA, 301(19), 2024–2035. https://doi.org/10.1001/jama.2009.681

[5] Mandsager, K., Harb, S., Cremer, P., Phelan, D., Nissen, S. E., & Jaber, W. (2018). Association of cardiorespiratory fitness with long-term mortality among adults undergoing exercise treadmill testing. JAMA Network Open, 1(6), e183605. https://doi.org/10.1001/jamanetworkopen.2018.3605

[6] Harber, M. P., Myers, J., Bonikowske, A. R., Muntaner-Mas, A., Molina-Garcia, P., Arena, R., & Ortega, F. B. (2024). Assessing cardiorespiratory fitness in clinical and community settings. Progress in Cardiovascular Diseases, 83, 36–42. https://doi.org/10.1016/j.pcad.2024.02.009

[7] Abut, F., Akay, M. F., & George, J. (2016). Developing new VO_2max prediction models from maximal, submaximal and questionnaire variables using SVMs with feature selection. Computers in Biology and Medicine, 79, 182–192. https://doi.org/10.1016/j.compbiomed.2016.10.018

[8] Beltrame, T., Amelard, R., Wong, A., & Hughson, R. L. (2017). Prediction of oxygen uptake dynamics by machine learning of wearable sensors during activities of daily living. Scientific Reports, 7(1), 45738. https://doi.org/10.1038/srep45738

[9] Liu, Y., Herrin, J., Huang, C., Khera, R., Dhingra, L. S., Dong, W., et al. (2023). Non-exercise ML models for VO_2max prediction in national population surveys. JAMIA, 30(6), 1062–1072. https://doi.org/10.1093/jamia/ocad035

[10] Spathis, D., Perez-Pozuelo, I., Gonzales, T. I., Wu, Y., Brage, S., Wareham, N., & Mascolo, C. (2022). Longitudinal cardio-respiratory fitness prediction through wearables in free-living environments. npj Digital Medicine, 5(1), 176. https://doi.org/10.1038/s41746-022-00719-1

[11] Neshitov, A., Tyapochkin, K., Kovaleva, M., Dreneva, A., Surkova, E., Smorodnikova, E., & Pravdin, P. (2023). Estimation of cardiorespiratory fitness using heart rate and step count data. Scientific Reports, 13(1), 15808. https://doi.org/10.1038/s41598-023-43024-x

[12] Kaminsky, L. A., Arena, R., Beckie, T. M., Brubaker, P. H., Church, T. S., Forman, D. E., et al. (2015). Reference standards for cardiorespiratory fitness measured with CPET: Data from the FRIEND Registry. Mayo Clinic Proceedings, 90(11), 1515–1523. https://doi.org/10.1016/j.mayocp.2015.07.026

[13] Alzamer, H., Abuhmed, T., & Hamad, K. (2021). A short review on ML-guided oxygen uptake prediction for sport science applications. Electronics, 10(16), 1956. https://doi.org/10.3390/electronics10161956

[14] Ashfaq, A., Cronin, N., & Müller, P. (2022). Recent advances in ML for VO_2max prediction: A review. Informatics in Medicine Unlocked, 28, 100863. https://doi.org/10.1016/j.imu.2022.100863

[15] Safari, S. S. B., Sivakumar, S., Lim, K. H., & Tan, T. P. L. (2022). A review on oxygen consumption and heart rate monitoring methods & applications. In Proc. GECOST 2022, 462–469. https://doi.org/10.1109/GECOST55694.2022.10010466

[16] Vicente-Samper, J. M., Tamantini, C., Ávila-Navarro, E., De La Casa-Lillo, M. Á., Zollo, L., Sabater-Navarro, J. M., & Cordella, F. (2023). ML-based reconstruction of heart rate from PPG with motion artifacts. Biosensors, 13(7), 718. https://doi.org/10.3390/bios13070718

[17] Hsiao, C.–T., Tong, C., & Coté, G. L. (2025). ML-based VO_2 estimation using a wearable multi-wavelength PPG device. Biosensors, 15(4), 208. https://doi.org/10.3390/bios15040208

[18] Lee, H. A., et al. (2023). Machine-learning VO_2max estimation using a patch-type single-lead ECG in lung resection candidates. Healthcare, 11(21), 2842. https://doi.org/10.3390/healthcare11212863

[19] Matsuo, T., So, R., & Murai, F. (2022). Improved VO_2max estimation by combining multiple regression and linear extrapolation. Journal of Cardiovascular Development and Disease, 10(1), 9. https://doi.org/10.3390/jcdd10010009

[20] Düking, P., Giessing, L., Frenkel, M. O., Koehler, K., Holmberg, H.–C., & Sperlich, B. (2020). Wrist-worn wearables for HR and energy expenditure across activities: Validation. JMIR mHealth and uHealth, 8(5), e16716. https://doi.org/10.2196/16716

[21] Lambe, R., O'Grady, B., Baldwin, M., & Doherty, C. (2025). Accuracy of Apple Watch VO_2max measurements: A validation study. PLOS ONE, 20(5), e12080799. https://doi.org/10.1371/journal.pone.0323741

[22] Schumacher, B. T., LaMonte, M. J., LaCroix, A. Z., Simonsick, E. M., Hooker, S. P., Parada, H., Bellettiere, J., & Kumar, A. (2024). Development, validation, and transportability of non-exercise VO_2max ML models for older adults. Frontiers in Physiology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11282333/

[23] Rissanen, A. P. E., et al. (2022). Validity of wearable-derived cardiorespiratory fitness estimates in clinical risk populations compared with CPET. [Journal/DOI to be updated].

[24] Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2020). Artificial intelligence in precision cardiovascular medicine. JACC, 74(3), 276–291.

[25] DexaFit. (2025). VO_2max testing services. Retrieved September 21, 2025, from https://www.dexafit.com/services/vo2-max-testing

[26] Acharya, U. R., Alizadehsani, R., Roshanzamir, M., Hussain, S., Khosravi, A., Koohestani, A., et al. (2021). Handling uncertainty in medical data using ML and probability: A 30-year review. Annals of Operations Research, 1–42. https://doi.org/10.1007/s10479-021-04006-2

ISSN: 2961-5992

[27] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., … Moher, D. (2021). The PRISMA 2020 statement. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

[28] Umiatin, U., Laut, T. N., Ifa, R. P. N., & Faisal, M. R. (2024). Smart-band for vital signs monitoring based on IoT. Journal of Physics: Conference Series, 2866(1), 012043. https://doi.org/10.1088/1742-6596/2866/1/012043

[29] Sadeghi, S. (2022). Diabetes mellitus risk prediction in the presence of class imbalance. BMC Medical Informatics and Decision Making, 22, 243. https://doi.org/10.1186/s12911-022-01775-z

[30] Pescatello, L. S., MacDonald, H. V., Lamberti, L., & Johnson, B. T. (2015). Exercise and hypertension: Advances in exercise prescription. ACSM's Health & Fitness Journal, 19(1), 20–27.

[31] Barrett, D. (2025, August 19). PREVENT™ Risk Calculator. Center for Health Technology & Innovation. https://ahahealthtech.org/prevent-risk-calculator/

[32] Framingham Risk Score for Hard Coronary Heart Disease. (n.d.). MDCalc. Retrieved November 4, 2025, from https://www.mdcalc.com/calc/38/framingham-risk-score-hard-coronary-heart-disease

[33] SCORE2 and SCORE2-OP calculators. (n.d.). European Society of Cardiology. Retrieved November 4, 2025, from https://www.escardio.org/Education/Practice-Tools/CVD-prevention-toolbox/SCORE-Risk-Charts

[34] Aug, L. R. (2025). Getting active to control high blood pressure. Heart.org. Retrieved November 4, 2025, from https://www.heart.org/en/health-topics/high-blood-pressure/changes-you-can-make-to-manage-high-blood-pressure/getting-active-to-control-high-blood-pressure

[35] High blood pressure dangers: Hypertension's effects on your body. (n.d.). Mayo Clinic. Retrieved November 4, 2025, from https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/high-blood-pressure/art-20045868

[36] Grafton, G., Cascino, T. M., Perry, D., Ashur, C., & Koelling, T. M. (2020). Resting oxygen consumption and heart failure: Importance for Fick principle cardiac output. Journal of Cardiac Failure, 26(8), 664–672. https://doi.org/10.1016/j.cardfail.2019.02.004

[37] Ho, C.–A., Yeh, H.–C., Lau, H.–T., Chang, E.–Y., Hsu, C.–W., Chang, C.–H., Huang, C.–C., Chien, W.–S. C., & Ho, C.–S. (2025). VO_2max prediction models from a step test. International Journal of Medical Sciences, 22(11), 2676–2685. https://doi.org/10.7150/ijms.109977

[38] Lin, C.–C., et al. (2022). Risk prediction model for chronic kidney disease among individuals with type 2 diabetes. Scientific Reports, 12, 4963. https://doi.org/10.1038/s41598-022-08284-z

[39] Myers, J., Tanaka, H., Yamamoto, D., Masuda, S., & Kaminsky, T. (2005). Physical activity, CRF, and the metabolic syndrome. Nutrition, Metabolism & Cardiovascular Diseases, 15(6), 441–449. https://doi.org/10.1016/j.numecd.2005.02.006

[40] Sloan, R. A., Scarzanella, M. V., Gando, Y., & Sawada, S. S. (2021). Accuracy of non-exercise estimated CRF in Japanese adults. International Journal of Environmental Research and Public Health, 18(23), 12288. https://doi.org/10.3390/ijerph182312288

[41] Theerthagiri, J., & Vidya, R. (2021). Heart disease prediction using ML algorithms. Materials Today: Proceedings, 45(2), 1154–1158.

[42] Assessment of maximal oxygen uptake (VO_2max) in athletes and non-athletes assessed in sports physiology laboratory. (n.d.). Cureus. Retrieved November 4, 2025, from https://www.cureus.com/articles/254748#!/

ISSN: 2961-5992