

Review

Not peer-reviewed version

---

# A Systematic Review of Machine Learning Approaches for ECG-Based Detection of Dysglycemia and Their Translational Readiness

---

[Zhadyra Alimbayeva](#), [Chingiz Alimbayev](#)<sup>\*</sup>, [Kassymbek Ozhikenov](#), [Aiman Ozhikenova](#)<sup>\*</sup>, [Ussen Shylmyrza](#),  
Kymbat Khaidarova

Posted Date: 27 April 2026

doi: 10.20944/preprints202604.1806.v1

Keywords: ECG; dysglycemia; machine learning; deep learning; HRV; non-invasive screening; diabetes detection; wearable ECG; AI diagnostics



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# A Systematic Review of Machine Learning Approaches for ECG-Based Detection of Dysglycemia and Their Translational Readiness

Zhadyra Alimbayeva <sup>1,2</sup>, Chingiz Alimbayev <sup>1,\*</sup>, Kassymbek Ozhikenov <sup>1</sup>, Aiman Ozhikenova <sup>1,\*</sup>, Ussen Shylmyrza <sup>1</sup> and Kymbat Khaidarova <sup>1</sup>

<sup>1</sup> Satbayev University, Department of Robotics and technical means of automation, Almaty 050013, Kazakhstan

<sup>2</sup> Kazakh National Women's Teacher Training University, Department of Information Technology and Library Science, Almaty 050000, Kazakhstan

\* Correspondence: c.alimbayev@satbayev.university (C.A.); a.ozhikenova@satbayev.university (A.O.);

## Abstract

This systematic review provides a comprehensive and quantitatively grounded synthesis of machine learning (ML) approaches for electrocardiography (ECG)-based detection of dysglycemia, with a specific focus on translational readiness for clinical screening. A structured literature search across PubMed, Scopus, Web of Science, and IEEE Xplore (February 2025) identified 183 records, of which 17 studies met predefined inclusion criteria following PRISMA-guided screening. The included studies demonstrate substantial heterogeneity in dataset size (ranging from <50 to >25,000 subjects), ECG acquisition modalities (single-lead, 12-lead, wearable), feature representations (raw signals, heart rate variability, engineered features), and ML strategies (classical algorithms, deep learning, and multimodal models). Reported model performance is generally high, with accuracy values frequently exceeding 0.85 and area under the curve (AUC) ranging from 0.78 to 0.99. Smaller experimental studies often report inflated performance (up to 96–99% accuracy), whereas large-scale population-based investigations demonstrate more moderate but clinically plausible results (AUC ≈ 0.80–0.85). External validation, a key requirement for clinical applicability, was performed in only a limited subset of studies (approximately 12%). From a physiological perspective, ML models exploit ECG alterations associated with dysglycemia, including reduced heart rate variability, QT interval prolongation, and changes in ventricular depolarization and repolarization dynamics. However, the relationship between metabolic dysfunction and ECG signals remains indirect. A key finding of this review is the mismatch between reported predictive performance and model maturity. The majority of studies (≈65–70%) are classified as early-stage (Level 1–2 or 2–3), relying on small, single-center datasets and internal validation. Only a minority of studies achieve near-translational maturity (Level 4), characterized by large-scale datasets and external validation. ECG-based dysglycemia detection represents a promising non-invasive and scalable screening paradigm. However, its clinical translation is constrained by the lack of standardized ECG acquisition protocols, limited dataset diversity, insufficient external validation, and fragmented methodological frameworks. Future research should prioritize large multi-center datasets, standardized feature extraction pipelines, hybrid interpretable models, and prospective validation to enable robust, generalizable, and clinically deployable screening systems.

**Keywords:** ECG; dysglycemia; machine learning; deep learning; HRV; non-invasive screening; diabetes detection; wearable ECG; AI diagnostics

---

## 1. Introduction

Type 2 Diabetes has emerged as one of the most pressing global health challenges of the 21st century and remains a leading cause of mortality, disability, and reduced quality of life worldwide. Over recent decades, the burden of type 2 diabetes has increased at an unprecedented rate. Epidemiological estimates indicate that the global number of individuals affected by type 2 diabetes rose from 148.4 million (135.5–162.6 million) in 1990 to 437.9 million (402.0–477.0 million) in 2019 [1], representing nearly a threefold increase within a single generation. This rapid growth is largely attributed to a combination of demographic and lifestyle transitions, including population aging, urbanization, reduced physical activity, and the global rise in obesity. Importantly, the burden of diabetes is no longer confined to older populations.

Despite advances in pharmacological treatment and disease management, early detection of type 2 diabetes remains a critical unresolved problem. The disease is characterized by a prolonged asymptomatic phase, during which metabolic dysregulation gradually progresses without overt clinical signs. As a result, a substantial proportion of individuals remain undiagnosed until the development of complications affecting multiple organ systems, including neuropathy, nephropathy, retinopathy, and cardiovascular disease [3–5]. From a clinical perspective, this delay in diagnosis significantly limits the effectiveness of preventive strategies and increases both morbidity and healthcare costs. From a public health standpoint, it underscores the need for scalable, accessible, and cost-effective screening approaches capable of identifying individuals at risk at earlier stages of disease progression.

Current diagnostic standards are based on laboratory measurements of blood glucose and glycated hemoglobin (HbA1c), which serve as the clinical reference for diabetes diagnosis [6]. Additional tools, such as capillary glucometers and continuous glucose monitoring systems, are widely used for disease monitoring and management [7,8]. While these methods provide accurate and clinically validated measurements, their implementation in large-scale screening remains constrained by practical considerations. Blood sampling, the need for laboratory infrastructure, device costs, and issues related to patient compliance limit their feasibility, particularly in low-resource or geographically remote settings. These limitations have driven increasing interest in the development of non-invasive, easily deployable technologies that could enable population-level screening without the need for biochemical testing.

Among such approaches, the use of Electrocardiography has attracted growing attention [9]. Electrocardiography is a widely available, low-cost, and non-invasive technique routinely used in clinical practice and increasingly integrated into wearable devices. The rationale for its application in diabetes screening is based on the well-established impact of metabolic disorders on cardiac electrophysiology. Diabetes is known to influence cardiac function through multiple interconnected mechanisms, including autonomic neuropathy [10], metabolic disturbances [11], and structural alterations of the myocardium [12]. These processes lead to measurable changes in electrocardiographic parameters, such as heart rate variability and ventricular repolarization [13].

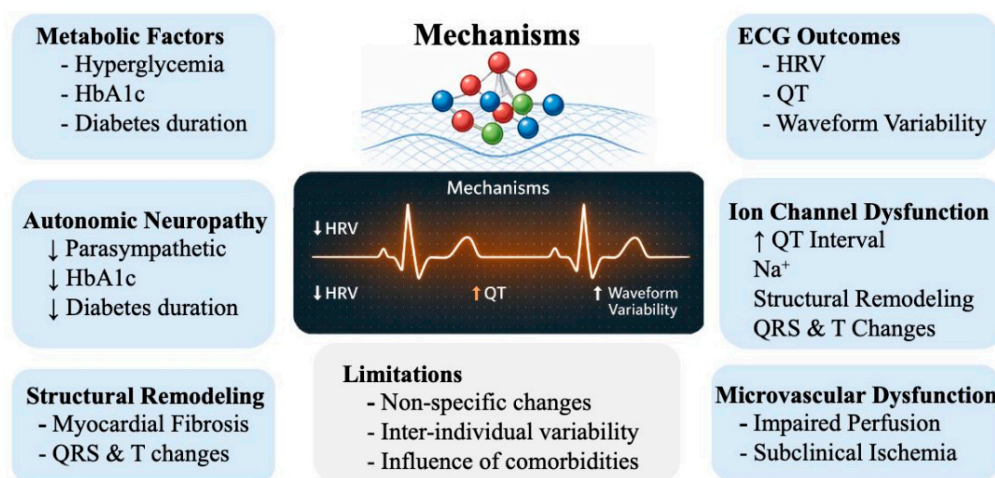
One of the earliest and most clinically relevant manifestations is cardiovascular autonomic neuropathy, which results from chronic hyperglycemia-induced damage to autonomic nerve fibers regulating cardiac function [14]. The extensive innervation of the heart involves a complex network of sympathetic and parasympathetic pathways, and disruption of this system leads to impaired regulation of cardiac output and electrophysiological stability [15]. A key measurable consequence of autonomic dysfunction is a reduction in heart rate variability (HRV), which has been consistently observed in individuals with diabetes and even in prediabetic states [16,17]. Importantly, HRV parameters show significant associations with glycemic markers, including fasting glucose and HbA1c, as well as with disease duration, suggesting a progressive deterioration of autonomic regulation over time. Structural and functional alterations in cardiac parasympathetic pathways, including postganglionic neurons within intracardiac ganglia, contribute to the withdrawal of parasympathetic tone and increased susceptibility to arrhythmias [18].

At the cellular level, metabolic disturbances associated with diabetes alter the function of cardiac ion channels. Chronic hyperglycemia, together with oxidative stress and inflammatory processes,

modifies the activity and expression of sodium, potassium, and calcium channels [19,20]. These alterations disrupt the normal dynamics of the cardiac action potential, affecting both depolarization and repolarization processes. In particular, impaired potassium channel function is associated with delayed repolarization and prolongation of the QT interval, a well-recognized electrocardiographic marker linked to increased risk of arrhythmias and sudden cardiac death [21]. Clinical studies have demonstrated that QT interval duration correlates with HbA1c levels and disease duration, indicating that poor glycemic control may directly influence cardiac electrophysiological stability [22]. In addition, changes in QRS duration and QT dispersion have been reported, particularly in patients with longer disease duration, further suggesting increased heterogeneity in electrical conduction [23].

Beyond cellular mechanisms, diabetes also induces structural and functional alterations in the myocardium, often described as diabetic cardiomyopathy. This condition is characterized by myocardial fibrosis, hypertrophy, and impaired calcium handling, resulting from dysfunction of key regulatory proteins such as ryanodine receptors, sarcoplasmic reticulum calcium ATPase, and sodium-calcium exchangers [24]. These changes lead to impaired contractility and altered propagation of electrical signals within cardiac tissue. At the same time, chronic inflammation and oxidative stress contribute to electrical remodeling, while microvascular dysfunction impairs myocardial perfusion and may lead to subclinical ischemia [25]. The combined effect of these processes is the emergence of subtle, heterogeneous, and often non-specific alterations in ECG signals. To provide a structured overview of these interconnected mechanisms and their implications for ECG-based screening, a schematic representation is presented in Figure 1.

The convergence of these mechanisms has motivated the exploration of machine learning techniques applied to ECG data for the detection of diabetes and related metabolic abnormalities [26]. In principle, ECG-based screening offers several attractive features, including non-invasive acquisition, low operational cost, and compatibility with large-scale deployment through wearable technologies. Despite growing interest in ECG-based detection of dysglycemia, the existing literature remains fragmented and methodologically heterogeneous. Previous reviews have summarized mechanistic links and ECG biomarkers in dysglycemia, including our prior work [27], which focused primarily on pathophysiological and electrophysiological aspects. However, these studies did not systematically evaluate machine learning methodologies, dataset characteristics, and translational readiness. Studies differ substantially in dataset scale, population characteristics, ECG acquisition protocols, feature representation, and validation strategies.



**Figure 1.** Pathophysiological mechanisms linking diabetes to ECG alterations and their limitations for clinical screening.

In this context, this review aims to provide a systematic and quantitative synthesis of machine learning approaches for ECG-based dysglycemia detection. Specifically, it categorizes existing studies by data sources, ECG modalities, feature representation, and modeling strategies, and evaluates reported performance in relation to dataset characteristics and validation design. The novelty of this work lies in its translational perspective. Beyond summarizing existing methods, this review analyzes how key methodological factors influence model robustness and introduces a structured maturity framework to assess the readiness of current approaches for clinical application. This integrated analysis provides a clearer understanding of the capabilities and limitations of ECG-based screening and outlines directions for future research toward scalable and clinically relevant solutions.

## 2. Literature Search Methodology

### 2.1. Search Strategy

To systematically synthesize evidence on machine learning approaches for ECG-based detection of dysglycemia, a structured literature search was conducted across major biomedical and engineering databases, including PubMed, Scopus, Web of Science, and IEEE Xplore. The search strategy targeted studies at the intersection of electrocardiography, glycemic disorders, and artificial intelligence, using keywords related to ECG (“ECG”, “electrocardiogram”, “heart rate variability”), dysglycemia (“diabetes”, “prediabetes”, “hyperglycemia”, “dysglycemia”), and machine learning (“machine learning”, “deep learning”, “artificial intelligence”). These terms were combined using Boolean operators as follows: (“ECG” OR “electrocardiogram” OR “heart rate variability”) AND (“diabetes” OR “prediabetes” OR “hyperglycemia” OR “dysglycemia”) AND (“machine learning” OR “deep learning” OR “artificial intelligence”). In addition, several targeted search phrases were used to capture variations in terminology across studies, including “ECG-based diabetes detection”, “electrocardiogram diabetes prediction”, “ECG-based diabetes screening”, “ECG machine learning diabetes”, “non-invasive diabetes detection ECG”, “ECG signals diabetes classification”, and “type 2 diabetes detection deep learning ECG”. The search was conducted in February 2025, and only studies published in English were considered. To ensure comprehensive coverage, reference lists of relevant articles were also screened. Given the exploratory and interdisciplinary nature of this review, no formal review protocol was registered, however, the review methodology was predefined and conducted according to PRISMA-based principles.

### 2.2. Eligibility Criteria

Study selection was guided by predefined inclusion and exclusion criteria. Studies were included if they utilized ECG signals or ECG-derived features, including heart rate variability, as input data and applied machine learning or deep learning methods for the detection, classification, or prediction of diabetes, prediabetes, or dysglycemia. Eligible studies were required to report quantitative performance metrics such as accuracy, area under the curve (AUC), sensitivity, or specificity. Studies were excluded if they did not involve ECG data, did not employ machine learning approaches, or focused exclusively on Type 1 diabetes, gestational diabetes, or disease management rather than detection. Additionally, studies lacking performance evaluation, as well as review articles, editorials, and non-peer-reviewed preprints, were excluded.

### 2.3. Study Selection and Data Extraction

All records identified through database searches were subjected to a structured multi-stage screening process. Initially, duplicate entries were identified and removed. The remaining records were screened based on titles and abstracts to exclude clearly irrelevant studies. Subsequently, full-text articles were assessed for eligibility according to the predefined inclusion and exclusion criteria. The study selection process followed a PRISMA-based workflow [28] and is illustrated in Fig. 2. A total of 183 records were initially identified, of which 52 duplicates were removed, resulting in 131 records for screening. Following title and abstract screening, 78 records were excluded. The

remaining 53 articles were assessed in full text, and 36 studies were further excluded for not meeting the eligibility criteria. Ultimately, 17 studies were included in the final analysis.

Data extraction was performed using a standardized framework to ensure consistency across studies. For each included study, the following information was collected: publication year, dataset characteristics, ECG acquisition type (single-lead, multi-lead, or wearable), feature representation (including raw signals, heart rate variability, or engineered features), sample size, population characteristics, glycemic markers (HbA1c, glucose levels, or diagnostic labels), study design, machine learning model, performance metrics, and reported limitations. The extracted data were used to construct the comparative summary presented in Table 1 and to support a structured analysis of methodological trends and model performance across the included studies. Screening and data extraction were performed by one reviewer and verified for consistency.

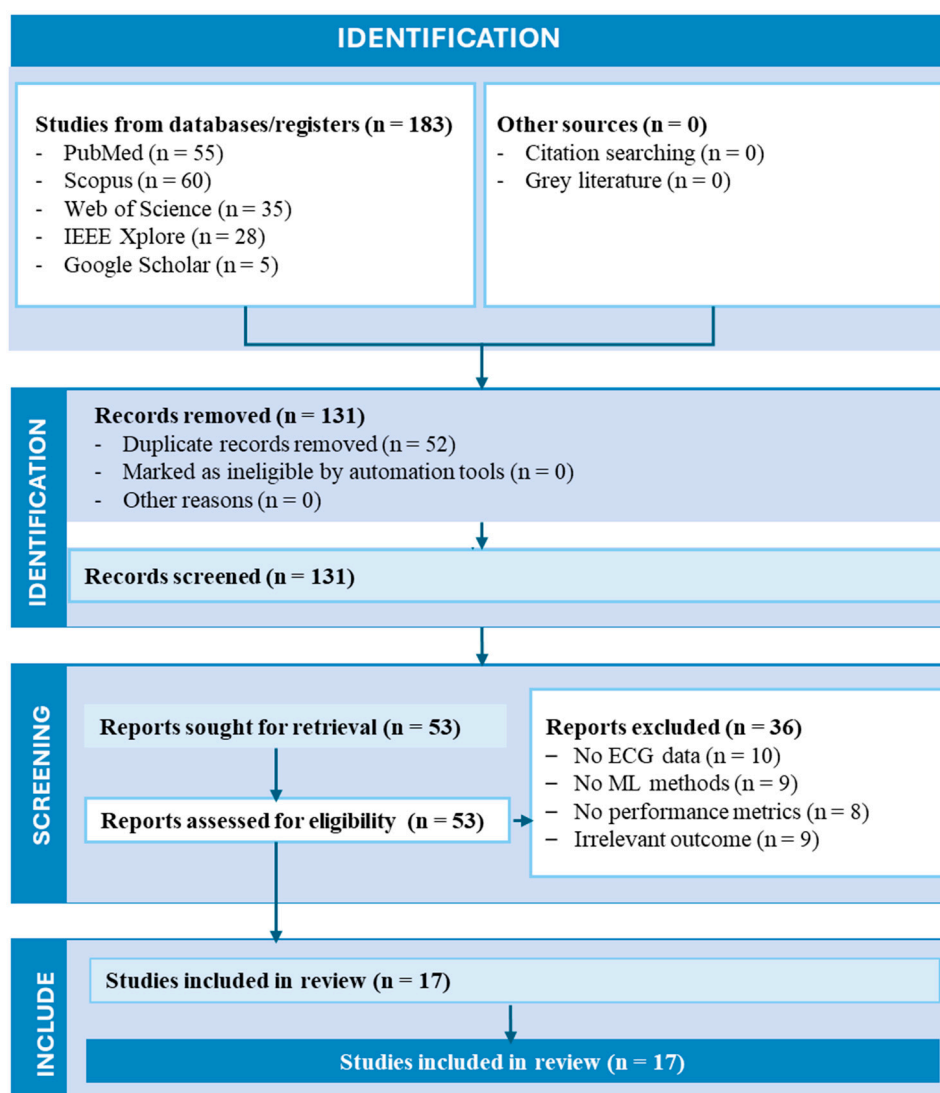


Figure 2. PRISMA flow diagram of study selection.

#### 2.4. Risk of Bias Assessment

The risk of bias in the included studies was evaluated to assess methodological rigor and the reliability of reported results, with particular attention to machine learning-based approaches for ECG-based dysglycemia detection. The assessment focused on the transparency and completeness of study reporting, including ECG data acquisition protocols, dataset sources, study populations, and clearly defined inclusion and exclusion criteria.

Additional evaluation considered the clarity of the prediction task, the description of preprocessing procedures, and the extent to which input features and model variables were explicitly reported. Particular attention was given to class distribution reporting, given the potential impact of class imbalance on model performance. The completeness and consistency of reported performance metrics, including accuracy, sensitivity, specificity, and AUC, were also examined. Studies were assessed based on the robustness of validation strategies, including the use of internal validation, cross-validation, or external validation on independent datasets. Studies relying on small sample sizes, lacking external validation, or providing insufficient methodological details were considered to have a higher risk of bias.

### 2.5. Data Synthesis

The synthesis of the included studies was performed using a structured approach to enable systematic comparison of methodologies and results. The synthesis was conducted in accordance with established systematic review principles to ensure transparency and consistency. The synthesis was organized around several key analytical dimensions relevant to ECG-based dysglycemia detection, including general study characteristics, dataset composition and population profiles, ECG acquisition protocols and signal configurations, feature representation approaches (such as raw signals, heart rate variability, and engineered features), machine learning strategies (including conventional algorithms, deep learning models, and multimodal approaches), and reported performance metrics together with validation schemes.

Each study was critically evaluated in terms of methodological clarity, consistency of reporting, and potential sources of bias, as described in the previous section. Performance indicators, including accuracy, sensitivity, specificity, F1-score, and area under the curve (AUC), were systematically extracted and compared where available. Due to substantial heterogeneity across studies in dataset characteristics, ECG configurations, preprocessing pipelines, model architectures, and outcome definitions, quantitative meta-analysis was not feasible. Instead, a narrative synthesis approach was adopted to identify recurring analytical patterns, compare modeling strategies, and highlight key limitations and research gaps in the current body of evidence.

The results were structured into thematic analytical categories, including data sources, ECG acquisition, feature representation, machine learning approaches, and model performance.

## 3. Results

### 3.1. General Characteristics and Aims

The studies included in this review ( $n = 17$ ) provide a comprehensive overview of current machine learning approaches for ECG-based detection of dysglycemia. As summarized in Table 1, the selected literature spans multiple years and reflects the progressive development of this research field, from early studies based on small cohorts and handcrafted features to more recent investigations utilizing large-scale datasets and deep learning architectures. The included studies demonstrate substantial variability in dataset size, population characteristics, ECG acquisition modalities, feature representation strategies, and analytical objectives. This heterogeneity is consistent with the interdisciplinary nature of the field, which integrates cardiology, metabolic research, signal processing, and artificial intelligence. The diversity of methodological approaches and study designs underscores the need for a structured synthesis, which is further developed in the subsequent subsections focusing on datasets, ECG configurations, feature extraction, machine learning models, and performance evaluation. The characteristics of the included studies are summarized in Table 1.

**Table 1.** Comparative summary of included studies.

Study	Year	Signal	Dataset	ECG	N	Population	Marker	Design	Model	Performance	Limitations	Maturity
-------	------	--------	---------	-----	---	------------	--------	--------	-------	-------------	-------------	----------

[29]	2021	ECG	Outpatient cohort (hospital-based)	12-lead, 500 Hz, 10 s; intervals (HR, PR, QRS, QT, QTc), axes (P, QRS, T)	4,832	Non-DM, prediabetes, T2DM; mean duration ~4.7 y	HbA1c	Retrospective cohort (validated)	CNN-based DL (ResNet + SE + attention)	AUC 0.826; Sens 71.9%; Spec 77.7%	Moderate accuracy; reduced performance in severe DM; single-center	Level 3 (retrospective clinical validation)
[30]	2023	ECG	Ethnic cohort (Sindhi, India; high-risk families)	12-lead, 10 s, 1000 Hz	1,262 (10,461 beats)	Mean age ~48 y; 61% female; high cardiometabolic burden	HbA1c, FPG, RBG	Observational; train/validation split	XGBoost (best); compared with RF, MLP, LSTM, CNN, Transformer	Acc 96.8%; Prec 97.1%; Rec 96.2%; F1 96.6%	Selection bias; no external validation; beat-level analysis; limited generalizability	Level 2 (model development, internal validation)
[31]	2022	ECG	Private dataset (non-public)	Single-lead; 256 Hz; resting	86 (24,630 segments)	35 T2DM / 51 healthy; age 20–70 y	Glucose (≥160 mg/dL)	Supervised classification	Decision Tree (DTC); compared with FT, MT, CT	Acc 86.9%; Sens 81.9%; Spec 90.6%; F1 82.8%	Small sample; no external validation; private dataset; limited generalizability	Level 2–3 (prototype; limited clinical validation)
[32]	2021	ECG	Private dataset (Taiwan; ECG + glucose)	Single-lead; 1000 Hz; 60 s	1,119	Age 38–80 y; mixed glycemic status	Blood glucose (≥100 mg/dL)	Retrospective; binary classification	Deep NN (10-layer); compared with LR, SVM	AUC 0.945; Sens 87.6%; Spec 85.0%	Private dataset; no external validation; sensitive to signal quality	Level 3 (advanced ML validation)
[33]	2020	ECG	Self-collected dataset	3-electrode setup (wrist + ankles)	24 (~1,500 samples)	10 diabetic / 14 healthy	Clinical status (no HbA1c/glycose)	Experimental; 5-fold CV	SVM (cubic); compared with DT, LDA, NB, KNN	Acc 96.8%	Very small sample; no objective biomarkers; no external validation; high overfitting risk	Level 1–2 (proof-of-concept)
[34]	2021	ECG	Hospital-based dataset (wearable ECG)	Single-lead; 60 s segments	370 (~317k segments)	T2DM only; mean age ~43.5 y	HbA1c	Retrospective; 5-fold CV	CNN-MFVW; compared with CNN, CNN-LSTM	Acc 90.2%; AUC 0.990; F1 0.901	No control group; small cohort; no external validation; sensitive to preprocessing	Level 2–3 (model development; limited clinical validation)
[35]	2021	ECG	Self-collected experimental dataset	Single-lead; 1000 Hz	21 (~22k segments)	Young adults; mixed glycemic status	Blood glucose (OGTT)	Prospective; 3-class classification	DBSCAN + CNN	Acc 81.7%; Sens 98.5%; Spec 76.8%	Very small sample; controlled setting; selection bias; no external validation	Level 2 (early-stage experimental study)
[36]	2025	ECG + clinical (multimodal)	Population-based cohort (Qatar Biobank)	12-lead (clinical)	2,043 + 395 (test)	Middle Eastern; mean age ~46 y	HbA1c, FPG	Cross-sectional + longitudinal (5-year follow-up)	DNN (ECG-DiaNet; ECG + CRFs)	AUC 0.845 (multimodal); 0.822 (CRF); 0.675 (ECG)	No external validation; single-region cohort; small longitudinal test set	Level 3 (advanced clinical ML; longitudinal validation)
[37]	2025	HRV (ECG-derived)	Retrospective cohort (AFT lab, India)	Lead II; 1000 Hz; 5-min segments	519 (261 T2DM / 258 controls)	Age 18–55 y; no major comorbidities	FBG, PPBG, HbA1c	Retrospective; binary classification	CatBoost (best); compared with LR, KNN, RF, GBM	Acc 91.3%; AUC 0.91; Sens 90.6%; Spec 91.9%	No external validation; controlled setting; HRV-only features; limited generalizability	Level 2–3 (validated ML model)
[38]	2025	ECG (engineered features)	Population-based cohort (Japan; external validation)	12-lead; 10 s; 500 Hz	16,766 + 2,456 (external)	General population; higher risk in older subjects	FPG, HbA1c	Retrospective; internal + external validation	LightGBM (best); compared with LR, RF, XGBoost, DNN	AUC 0.851 (internal); 0.785 (external)	Feature-based (no raw ECG DL); moderate specificity; class imbalance	Level 4 (advanced clinical ML with external validation)
[39]	2022	ECG + demographics (multimodal)	EHR cohort (NYU Langone)	12-lead; 10 s; 250–500 Hz	25,951 (test); large training cohort	Outpatients; new-onset diabetes subgroup	HbA1c ≥ 6.5%	Retrospective; prediction; external validation	DL (ResNet); ECG + demographics	AUC 0.80 (model); 0.68 (risk score)	Selection bias; multimodal dependence; no real-world validation; data not public	Level 4 (advanced clinical ML; near-translational)
[40]	2022	ECG (image-based)	Hospital cohort (China; 3 centers)	12-lead ECG images; 5 s	~2,914	Middle-aged/elderly; high-risk	FPG, OGTT	Retrospective; binary classification; CV + test set	CNN (JGRNet); compared with AlexNet, GoogleNet, SVM	Acc 0.781; AUC 0.777	Image-based ECG (information loss); no external validation;	Level 2–3 (early DL with internal validation)

										moderate performance		
[41]	2023	Multimodal (ECG + glucose + ACC + respiration)	DINAMO wearable dataset (free-living) (~4 days)	Wearable ECG; 250 Hz; continuous diabetic (~4 days)	29 (20 healthy / 9 diabetic)	Mixed cohort; continuous monitoring	Continuous glucose monitoring	Experimental; supervised classification	XGBoost (best); compared with LR, DT, RF, SVM	Acc 98.2% (multimodal); ~87.5% (ECG only)	Very small sample; uses glucose input; no external validation; high overfitting risk	Level 1–2 (exploratory multimodal study)
[42]	2024	ECG (high-density)	Private dataset (self-collected)	HD-ECG (up to 98 leads)	50	Healthy volunteers	Not specified	Experimental; supervised classification	CNN (HD-MVCNN)	Acc 99.0%; F1 94.5%	No glycemic ground truth; unclear labels; small sample; unrealistic setup (98 leads); no validation	Level 1 (concept study)
[43]	2023	ECG	MIMIC-III (ICU subset)	Single-lead; 125 Hz; 1 s windows	50	ICU patients; median age 64 y	Blood glucose	Retrospective; personalized classification	One-class SVM	AUC 0.92 (beat); 0.97 (10s)	ICU-only cohort; small sample; personalized model; no external validation	Level 3 (advanced ML validation)
[44]	2017	HRV (RR-interval)	Public dataset (PhysioNet)	RR intervals (QRS-based)	50 (33 normal / 17 diabetic)	Not specified	Not reported	Supervised classification	SVM	Acc ~95%	Very small sample; no glycemic markers; unclear labels; no external validation	Level 2 (early-stage study)
[45]	2024	HRV (ECG-derived)	Hospital cohort (Korea; prospective)	Wearable ECG; 250 Hz	83 → 21 (final)	T2DM only; elderly (mean ~69 y)	Continuous glucose	Observational; temporal prediction	1D CNN (ResNet-like; HRV input)	Acc 90.5%; Sens 87.5%; Spec 92.7%	Very small final cohort; no control group; HRV-only; no external validation	Level 3 (clinical ML validation)

As shown in Table 1, the reviewed studies can be grouped according to their primary objectives and analytical strategies. A subset of studies focuses on the estimation of glycemic biomarkers, particularly glycated hemoglobin (HbA1c), directly from ECG signals, demonstrating that ECG-derived representations can capture clinically relevant metabolic information and may serve as surrogate markers for disease progression [29,34,39]. Another major direction involves the classification of individuals into glycemic categories, including normoglycemia, prediabetes, and type 2 diabetes, where a wide range of machine learning techniques—such as gradient boosting, convolutional neural networks, and hybrid models—have been applied with consistently high reported performance [30], [32,35,40,42].

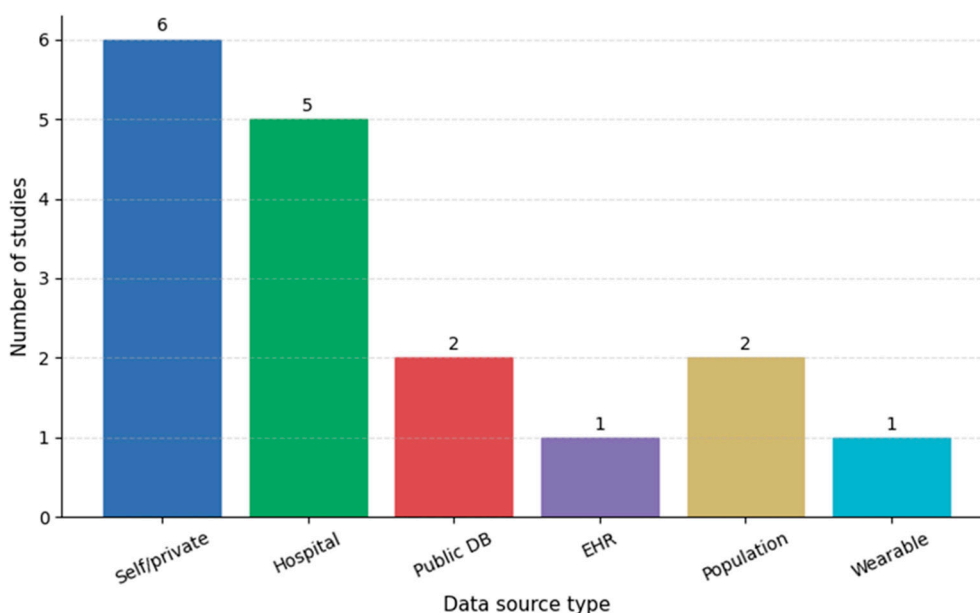
In parallel, several studies adopt a feature-engineering approach based on physiological signal analysis, employing techniques such as intrinsic time-scale decomposition, empirical mode decomposition, and entropy-based feature extraction, followed by classification using conventional machine learning algorithms [31,33,43]. Additional research directions include heart rate variability-based modeling, which emphasizes autonomic dysfunction as a key mechanism associated with dysglycemia [37,44], as well as multimodal frameworks integrating ECG signals with clinical risk factors to improve predictive performance [36,45].

Overall, the reviewed literature demonstrates that ECG-based dysglycemia detection has been explored across multiple methodological paradigms, including biomarker estimation, classification, physiological feature analysis, and multimodal prediction. This diversity of approaches highlights both the versatility of ECG as a data source and the absence of a unified modeling framework, which motivates the structured analysis presented in the following subsections.

### 3.2. Data Sources and Study Populations

The studies included in this review exhibit substantial heterogeneity in both data sources and study populations, reflecting the early-stage and interdisciplinary development of ECG-based dysglycemia detection. As summarized in Table 1 and illustrated in Figure 3, the reviewed datasets

span a wide spectrum, ranging from small self-collected experimental cohorts and private institutional databases to hospital-based clinical repositories, electronic health record (EHR) systems, public databases, and population-scale biobanks. A considerable proportion of studies relied on self-collected or non-public datasets with limited accessibility, including small experimental cohorts comprising 21–50 participants [33], [35,41,42,44]. In contrast, other studies utilized more structured clinical data sources, such as hospital outpatient cohorts [29,34], retrospective institutional datasets [32,37], and EHR-based cohorts [39]. Publicly available databases, including MIMIC-III and PhysioNet, were employed in only a limited number of studies [43,44], while large-scale population-based datasets—such as the Qatar Biobank and a Japanese health checkup cohort with external validation—were used in relatively few investigations [36,38]. As shown in Figure 3, this distribution highlights the predominance of small-scale and institution-specific datasets, with comparatively limited use of large, diverse, and externally validated data sources.



**Figure 3.** Distribution of data sources across included studies.

A pronounced imbalance is also evident in the distribution of sample sizes across studies. Many investigations were conducted on small cohorts, typically including fewer than 100 subjects, often under controlled or experimental conditions [31], [33,35,41,42,44,45]. In contrast, only a subset of studies leveraged larger datasets, including hospital-based cohorts with several thousand ECG records [29,34,40], a biobank-based dataset with more than 2,000 participants [36], a large population-based cohort comprising 16,766 records with an additional external validation cohort of 2,456 individuals [38], and an EHR-based study involving tens of thousands of patients [39]. This disparity indicates that much of the existing literature remains exploratory in nature, with limited statistical power and increased susceptibility to overfitting, whereas only a minority of studies approach the scale required for robust clinical validation and real-world deployment.

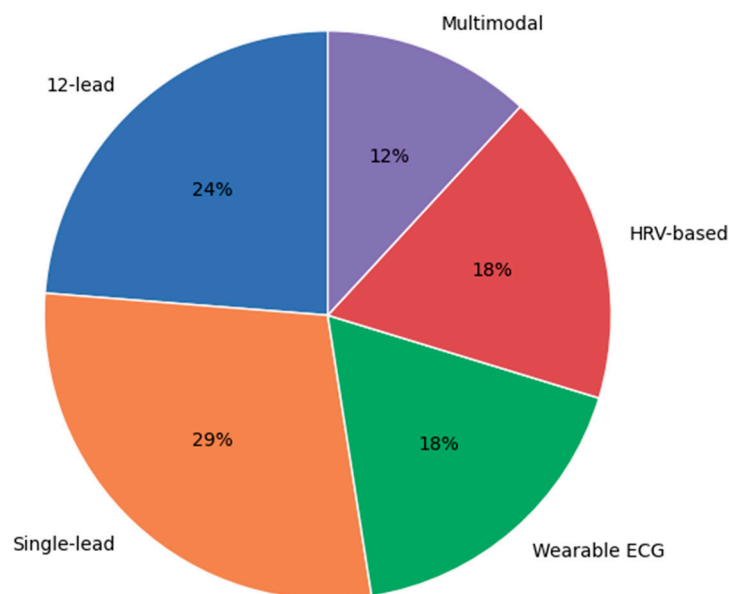
The composition of study populations further underscores the heterogeneity of the current evidence base. Several studies included mixed cohorts consisting of healthy individuals, prediabetic subjects, and patients with type 2 diabetes [29,30,37], whereas others focused on more specific or restricted populations, such as high-risk ethnic groups [30], outpatient cohorts [29], ICU patients [43], elderly individuals with established diabetes [45], or healthy volunteers [42]. In some cases, datasets were enriched with diabetic patients or lacked a control group entirely [34,45], thereby limiting their applicability to screening scenarios. Moreover, certain studies excluded participants with comorbidities [37] or were conducted under tightly controlled laboratory conditions [35,41], which may reduce ecological validity. The geographic distribution of datasets was also uneven,

encompassing studies from Asia, the Middle East, India, the United States, and public repositories, but with limited cross-regional validation [36,38,39].

From a translational perspective, a key limitation of the reviewed literature lies in the mismatch between study populations and intended clinical applications. Many studies were conducted in small, highly selective, or experimentally controlled cohorts rather than in representative community-based populations. Only a limited number of investigations utilized large outpatient, biobank, or population-based datasets that more closely reflect real-world screening conditions [36,38,39], and external validation was performed in only a small subset of studies [38,39]. Taken together, these findings suggest that while ECG-based dysglycemia detection is technically feasible, the representativeness, diversity, and scalability of available datasets remain critical barriers to clinical translation.

### 3.3. ECG Acquisition and Signal Configuration

The reviewed studies demonstrate substantial variability in ECG acquisition protocols and signal configurations, which represents a key methodological factor influencing both feature representation and model performance. As outlined in Table 1, ECG data were acquired using a wide range of configurations, including standard 12-lead clinical ECG systems, single-lead recordings, high-density multi-lead systems, and wearable devices operating under both controlled and free-living conditions (Figure 4). This diversity reflects the flexibility of ECG as a sensing modality, but also highlights the absence of standardized acquisition protocols for dysglycemia detection, which complicates cross-study comparison and limits reproducibility.



**Figure 4.** Distribution of ECG acquisition types.

A considerable proportion of studies relied on standard clinical 12-lead ECG systems, typically recorded over short durations of approximately 10 seconds under controlled conditions [29,30,38,39]. These datasets are generally associated with hospital-based or population-level cohorts and provide comprehensive spatial information on cardiac electrophysiology. In contrast, several studies employed single-lead ECG configurations, including both clinical and wearable setups, often with longer recording durations ranging from tens of seconds to several minutes [31,32,34,37]. While single-lead ECG enables simpler acquisition and improved scalability, it inherently reduces spatial

information and may limit the detection of subtle electrophysiological changes associated with dysglycemia.

Wearable ECG devices represent an important emerging direction, particularly for continuous monitoring and real-world data collection. Several studies utilized wearable sensors under free-living or semi-controlled conditions, enabling long-term signal acquisition and temporal analysis of glycemic states [34,41,45]. These approaches are particularly relevant for large-scale screening applications, as they offer non-invasive and scalable data collection. However, wearable ECG recordings are more susceptible to motion artifacts, signal noise, and variability in acquisition conditions, which introduces additional challenges for preprocessing and model robustness.

In addition to conventional and wearable ECG configurations, a small number of studies explored alternative signal representations, including high-density ECG systems with up to 98 channels [42], ECG images instead of raw signals [40], and heart rate variability-based representations derived from RR intervals [37,44]. While these approaches provide additional perspectives on cardiac electrophysiology, they also introduce methodological inconsistencies and may limit comparability across studies. In particular, high-density ECG systems and image-based representations are less practical for real-world deployment, whereas HRV-based approaches depend heavily on preprocessing quality and may omit relevant waveform information.

Another critical source of variability lies in signal duration and segmentation strategies. Some studies analyzed short, fixed-length ECG recordings (10-second segments), while others employed longer recordings segmented into windows or heartbeat-level samples for model training [30,34,43]. In certain cases, beat-level segmentation was used to increase dataset size, although this may introduce data leakage or artificially inflate performance metrics if not properly controlled. Conversely, longer recordings enable more robust estimation of temporal features, such as heart rate variability, but may increase computational complexity and sensitivity to noise.

The reviewed literature indicates that ECG acquisition and signal configuration remain highly heterogeneous, with no consensus on optimal recording protocols for dysglycemia detection. This variability directly influences downstream feature extraction strategies and model design, as further discussed in the following subsection. Importantly, the lack of standardized ECG acquisition protocols represents a key barrier to reproducibility, comparability, and clinical translation of ECG-based dysglycemia detection systems.

### 3.4. Feature Representation and ECG-Derived Biomarkers

The representation of ECG signals and the selection of informative features constitute a central component of machine learning frameworks for dysglycemia detection. As summarized in Table 1, the reviewed studies employ a wide spectrum of feature extraction strategies, ranging from raw ECG signals and deep learning-based representations to engineered features derived from waveform morphology and heart rate variability (HRV). This methodological diversity reflects both the complexity of the underlying physiological mechanisms and the absence of a standardized feature representation framework for ECG-based dysglycemia detection.

A substantial group of studies relies on raw ECG signals as direct input to deep learning models, enabling automatic feature extraction without explicit signal engineering. Convolutional neural networks and related architectures were used to learn hierarchical representations of ECG waveforms, capturing complex temporal and morphological patterns associated with glycemic status [29,32,34]. These approaches are particularly advantageous in large-scale datasets, where sufficient data allow models to learn subtle, non-linear relationships between ECG signals and metabolic abnormalities. For example, deep learning models have been shown to estimate HbA1c levels and detect hyperglycemia directly from ECG signals, suggesting that ECG may encode latent biomarkers of glycemic regulation. However, the interpretability of such models remains limited, and their performance is highly dependent on dataset size and quality.

In contrast, several studies employ engineered features derived from ECG waveform characteristics, including classical electrophysiological parameters such as heart rate, PR interval,

QRS duration, QT interval, QTc, and P- and T-wave morphology [29,30]. These features are physiologically interpretable and are directly linked to known mechanisms of diabetes-related cardiac dysfunction, including autonomic neuropathy, ion channel alterations, and myocardial remodeling. Feature-based approaches often incorporate preprocessing steps such as filtering, normalization, and fiducial point detection, followed by feature selection techniques to identify the most discriminative parameters [31,32]. While these methods provide greater transparency and clinical interpretability, they may fail to capture complex interactions present in raw signals.

Heart rate variability represents a distinct and widely used category of ECG-derived features, reflecting autonomic nervous system function. Multiple studies utilize time-domain, frequency-domain, and nonlinear HRV metrics, including SDNN, RMSSD, LF/HF ratio, and Poincaré plot parameters, to characterize alterations in cardiac autonomic regulation associated with dysglycemia [37,44]. These features are strongly grounded in physiological mechanisms, as reduced HRV is a well-established marker of autonomic dysfunction in diabetes. HRV-based models often achieve competitive performance and offer a simplified and computationally efficient representation of ECG data. However, they depend heavily on signal quality, accurate R-peak detection, and sufficiently long recording durations, and may omit important waveform-level information.

In addition to these primary approaches, several studies explore hybrid and alternative feature representations. Multimodal models combine ECG-derived features with clinical variables such as age, sex, and biochemical markers, demonstrating improved predictive performance compared to unimodal approaches [36]. Other studies employ feature-based machine learning pipelines using entropy measures, signal decomposition techniques (intrinsic time-scale decomposition or empirical mode decomposition), and statistical descriptors of ECG signals [31,33]. Furthermore, some approaches utilize ECG images or transformed representations rather than raw time-series data, although such methods may lead to information loss and reduced physiological interpretability [40].

Despite the diversity of feature extraction strategies, a common limitation across studies is the lack of standardization in feature definition, preprocessing pipelines, and evaluation protocols. Different studies use varying combinations of features, segmentation strategies, and normalization techniques, making direct comparison challenging. Moreover, the absence of consensus on which ECG-derived features are most relevant for dysglycemia detection reflects the complex and indirect relationship between metabolic disorders and cardiac electrophysiology. As discussed in Section 1, ECG alterations in diabetes are often subtle, non-specific, and influenced by multiple confounding factors, including comorbidities and inter-individual variability.

Overall, the current evidence suggests that both deep learning-based representations and engineered physiological features can capture relevant information for dysglycemia detection, but each approach has inherent limitations. Raw signal-based methods offer higher flexibility and potential performance, whereas feature-based approaches provide greater interpretability and physiological grounding. HRV-based representations offer a simplified and clinically meaningful alternative but may sacrifice signal richness. The lack of unified feature representation strategies remains a major barrier to reproducibility and clinical translation, highlighting the need for standardized feature extraction frameworks and multimodal approaches in future research.

### 3.5. Machine Learning Models

The methodological landscape of machine learning approaches applied to ECG-based dysglycemia detection is characterized by a coexistence of classical algorithms and deep learning architectures, with no clear consensus regarding the optimal modeling paradigm. The choice of model is closely intertwined with data characteristics and feature representation, as discussed in Sections 3.2–3.4, and therefore reflects both the scale of available datasets and the underlying assumptions about signal informativeness.

A substantial portion of the reviewed studies employs conventional machine learning algorithms operating on engineered feature sets. These include decision trees, random forests, support vector machines, k-nearest neighbors, and gradient boosting methods such as XGBoost and

LightGBM [30,31,37,38,43,44]. In many cases, these approaches are coupled with carefully selected physiological features derived from ECG waveforms or heart rate variability, allowing for relatively interpretable models that align with known mechanisms of diabetic cardiac dysfunction. Notably, gradient boosting methods frequently demonstrate strong performance in tabular settings, particularly when combining ECG-derived features with demographic or clinical variables [30,38]. However, their performance is inherently constrained by the quality and completeness of feature engineering, and their ability to capture complex temporal dependencies in raw ECG signals remains limited.

In parallel, deep learning models have been increasingly adopted, particularly in studies utilizing raw ECG signals or large-scale datasets. Convolutional neural networks (CNNs), including residual architectures, are the most commonly used models, enabling automated extraction of hierarchical features directly from waveform data [29], [32,34,39,45]. These models are well-suited to capturing subtle and distributed patterns in ECG signals that may not be accessible through manual feature engineering. In some cases, deep neural networks have demonstrated the ability to infer glycemic status or estimate HbA1c levels from ECG data alone, suggesting that latent representations of metabolic state may be encoded in cardiac electrical activity. Nevertheless, such models often require large training datasets, and their performance may degrade significantly when applied to smaller or heterogeneous cohorts.

A smaller subset of studies explores hybrid and alternative modeling strategies, including multimodal architectures that integrate ECG features with clinical or demographic data [36], as well as specialized approaches such as one-class classification for personalized modeling [43] or clustering-assisted pipelines for preprocessing and feature selection [35]. These approaches reflect attempts to address specific limitations of conventional supervised learning, such as class imbalance, limited data availability, or inter-individual variability. While promising, such methods remain relatively underexplored and are often evaluated in small or highly specific cohorts, limiting the generalizability of their findings.

Despite the diversity of modeling techniques, several recurring methodological issues can be identified across the literature. First, the majority of studies rely on retrospective datasets and internal validation strategies, with only a limited number incorporating external validation on independent cohorts [38,39]. Second, class imbalance and selection bias are frequently insufficiently addressed, particularly in studies involving enriched or highly selective populations. Third, the risk of overfitting remains substantial, especially in studies with small sample sizes and high-dimensional feature spaces. Finally, the lack of standardized evaluation protocols and reporting practices complicates the comparison of model performance across studies.

Taken together, the current evidence suggests that both classical machine learning methods and deep learning architectures are capable of achieving high predictive performance under specific conditions. However, these results are highly dependent on dataset characteristics, feature representation, and validation strategy. In particular, models trained on small, homogeneous, or institution-specific datasets may not generalize to broader populations. Therefore, the primary challenge is not the absence of effective algorithms, but rather the development of robust, generalizable models supported by large, diverse, and externally validated datasets. Addressing these limitations will be essential for translating ECG-based machine learning models from proof-of-concept studies to clinically applicable screening tools. The choice of machine learning model appears to be secondary to dataset characteristics and feature representation, as model performance is primarily driven by data quality, scale, and validation strategy rather than algorithmic complexity.

### *3.6. Model Performance and Validation*

The reported performance of machine learning models for ECG-based dysglycemia detection is generally high across the reviewed studies; however, these results should be interpreted with caution in light of substantial heterogeneity in evaluation protocols, dataset characteristics, and validation strategies. As summarized in Table 1, most studies report conventional metrics such as accuracy,

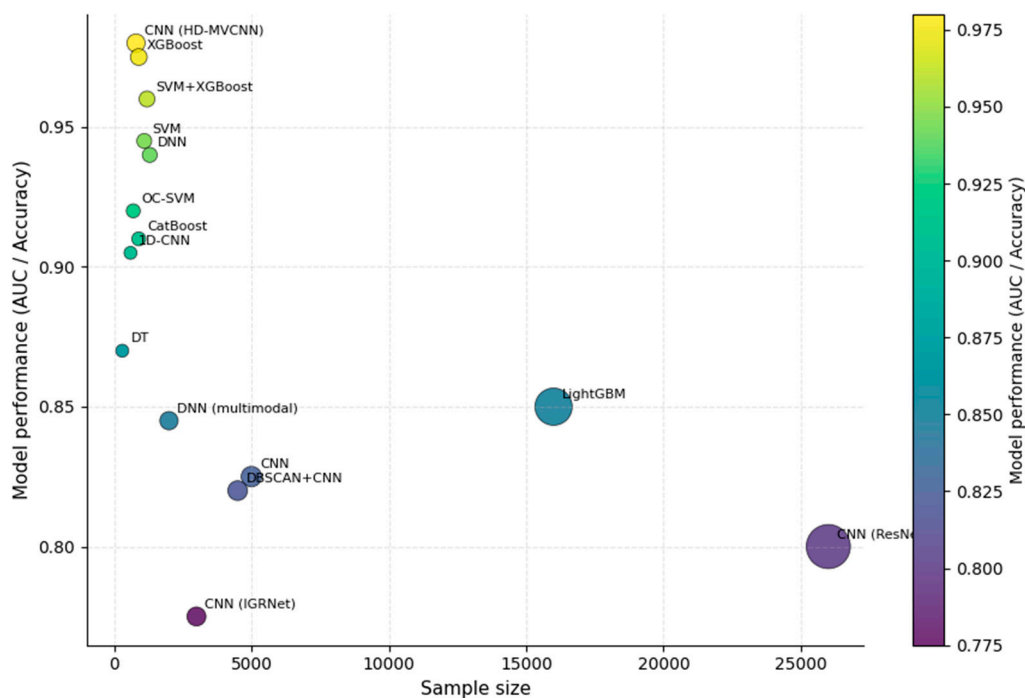
sensitivity, specificity, and area under the receiver operating characteristic curve (AUC), with many models achieving values in the range of 0.80–0.99. While these results may suggest strong predictive capability, they do not necessarily reflect real-world performance due to differences in study design, outcome definitions, and data composition.

A key observation emerging from the comparative analysis is the strong dependence of reported performance on dataset size and structure. As illustrated in Figure 5, studies based on small cohorts tend to report substantially higher performance metrics, often exceeding 0.90 or even 0.95, whereas studies utilizing larger and more heterogeneous datasets typically demonstrate more moderate results. This pattern suggests that model performance is not solely determined by algorithmic sophistication, but is strongly influenced by dataset scale, variability, and representativeness. In particular, small and homogeneous datasets may lead to overly optimistic estimates due to overfitting, limited variability, and reduced complexity of classification tasks.

Direct comparison of model performance across studies is further complicated by differences in problem formulation and outcome definitions. Some studies address binary classification tasks (diabetic vs. non-diabetic), while others consider multi-class classification or regression-based estimation of glycemic markers such as HbA1c. In addition, diagnostic thresholds vary across studies, including different cut-offs for fasting glucose or HbA1c levels, resulting in inconsistencies in labeling and evaluation. Consequently, similar numerical values of accuracy or AUC may correspond to fundamentally different prediction tasks and should not be interpreted as directly comparable measures of model effectiveness.

Another critical limitation lies in the design of validation strategies. The majority of studies rely on internal validation approaches, such as train–test splits or cross-validation applied to retrospective datasets. Although these methods are appropriate for initial model development, they are prone to optimistic bias, particularly when patient-level independence is not strictly enforced. In studies employing segmentation of ECG signals into multiple samples, the risk of data leakage increases if segments from the same individual appear in both training and testing sets. Notably, only a limited number of studies perform external validation using independent cohorts, which is essential for assessing model generalizability.

Dataset composition and population characteristics further affect reported performance. Several studies are conducted on enriched or highly selective populations, including high-risk groups, ICU patients, or cohorts without significant comorbidities. In such cases, classification tasks may be artificially simplified, leading to inflated performance metrics that do not reflect real-world screening conditions. Conversely, studies based on large, heterogeneous populations tend to report lower but more realistic performance values, reinforcing the importance of dataset diversity for robust model evaluation.

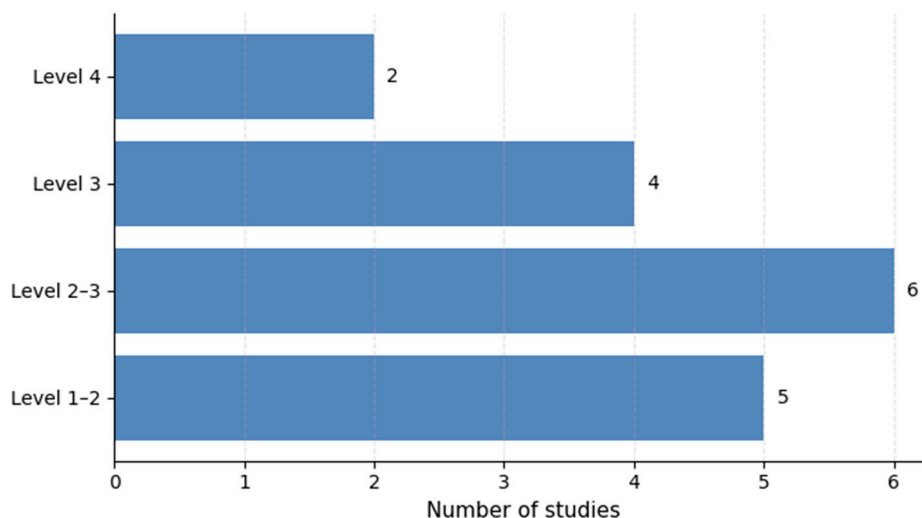


**Figure 5.** Relationship between dataset size and reported model performance.

Taken together, the evidence indicates that high model performance is achievable under controlled or small-scale conditions, but may not generalize to broader clinical populations. As highlighted in Figure 5, the apparent trade-off between dataset size and reported performance underscores the need for careful interpretation of published results. The primary limitation of the current literature is not the absence of predictive signal in ECG data, but rather the lack of standardized validation frameworks and the reliance on limited or biased datasets. Future studies should prioritize external validation, patient-level data separation, and evaluation in representative screening populations to ensure that reported performance metrics accurately reflect clinical applicability.

### 3.7. Model Maturity and Translational Readiness

The current body of evidence reveals a pronounced imbalance between reported model performance and the actual level of methodological maturity required for clinical translation. Despite consistently high predictive metrics across studies, the majority of approaches remain at early stages of development. As summarized in Table 1 and illustrated in Figure 6, most studies fall within Level 1–2 and Level 2–3 maturity, corresponding to proof-of-concept investigations and initial model development. These studies are typically based on small or highly controlled datasets and rely predominantly on internal validation strategies, which limits the reliability of their reported performance in real-world settings [33], [35,41,42,44].



**Figure 6.** Distribution of model maturity levels.

A subset of studies demonstrates a higher level of methodological rigor and can be classified within Level 3 maturity. These investigations generally utilize retrospective clinical datasets and apply more structured validation strategies, including cross-validation and independent test sets [29,32,34,37]. However, this group remains constrained by single-center data sources and limited population diversity. The absence of external validation in most of these studies raises concerns regarding their robustness and reproducibility across different clinical environments.

Only a small number of studies approach translational readiness (Level 4). These works are characterized by the use of large-scale datasets, such as population-based cohorts or electronic health record systems, and by the inclusion of external validation cohorts [36,38,39]. Such designs provide more realistic estimates of model performance and represent an important step toward clinical implementation. Nevertheless, even these studies remain predominantly retrospective and are often limited to specific geographic or institutional contexts, which restricts their broader applicability.

Taken together, the distribution of maturity levels presented in Figure 6 highlights a fundamental limitation of the field. High predictive performance is frequently achieved in low-maturity settings, whereas studies with more rigorous design and validation tend to report more moderate but clinically plausible results. This indicates that model maturity—defined by dataset scale, population representativeness, and validation strategy—provides a more meaningful indicator of translational potential than performance metrics alone. Advancing ECG-based dysglycemia detection will therefore require a shift toward externally validated, population-level, and prospectively evaluated models capable of integration into real-world clinical workflows.

#### 4. Discussion

The present review indicates that ECG-based dysglycemia detection is a technically promising yet methodologically immature field. While many studies report high predictive performance, these results are largely driven by dataset characteristics, including limited sample sizes, controlled conditions, and population bias, rather than intrinsic model capability. At the same time, substantial heterogeneity persists in ECG acquisition protocols, feature representation strategies, and validation approaches, preventing meaningful comparison across studies. Importantly, most models remain at early or intermediate stages of maturity, with minimal external validation and limited alignment with real-world screening scenarios. Collectively, these findings suggest that the primary challenge is not the absence of informative ECG biomarkers, but the lack of standardized data acquisition, robust validation frameworks, and scalable systems required for clinical translation.

**Table 2.** Synthesis of key findings, limitations, and required improvements for clinical translation.

Domain	Key finding	Strengths	Limitations	Required improvement
<b>Data characteristics</b>	ECG-based dysglycemia detection is feasible across multiple datasets	Large-scale studies demonstrate predictive potential	Most studies rely on small, single-center datasets; limited diversity	Large, multi-center, population-level datasets
<b>ECG acquisition</b>	Signal characteristics strongly influence model performance	Multilead ECG provides richer physiological information	Heterogeneous acquisition protocols; frequent use of single-lead ECG	Standardized, wearable multilead ECG systems
<b>Feature representation</b>	Both engineered and deep features capture relevant information	HRV and repolarization features show physiological relevance	Lack of standardization; inconsistent preprocessing	Hybrid feature frameworks with standardized pipelines
<b>Machine learning models</b>	ML and DL models achieve high performance under controlled conditions	CNN, boosting models show strong results	Performance depends on dataset rather than model; limited interpretability	Robust, interpretable models validated across datasets
<b>Model performance</b>	High reported accuracy in many studies	Strong results in experimental settings	Overfitting, optimistic bias, poor comparability	Standardized evaluation metrics and protocols
<b>Validation strategy</b>	Validation is a key bottleneck	Some studies include external validation	Most rely on internal validation; data leakage risk	External and prospective validation
<b>Model maturity</b>	Majority of studies at early/intermediate levels	Emerging Level 3–4 studies	Limited translational readiness	Maturity-driven development frameworks
<b>Clinical applicability</b>	ECG has potential for non-invasive screening	Scalable and low-cost modality	No real-world deployment; lack of screening studies	Integration into clinical workflows and screening programs
<b>System integration</b>	End-to-end systems are required	Advances in wearable ECG devices	Fragmented pipelines; lack of standardization	Integrated acquisition–ML–validation systems

#### 4.1. Requirements for Clinical Translation

The translation of ECG-based dysglycemia detection from experimental studies to clinically applicable screening systems requires the fulfillment of several methodological and technological criteria. Based on the analysis of the reviewed literature, these requirements can be structured into five key domains: data, acquisition, feature representation, modeling, and validation [46–48].

The availability of large-scale and representative datasets is a fundamental requirement. Models should be trained and evaluated on heterogeneous populations that reflect real-world screening conditions, including variability in age, sex, ethnicity, and comorbidities. The limitations of small and single-center datasets, which often lead to optimistic performance estimates, have been widely documented in machine learning–based medical studies [49]. In addition, standardized outcome definitions based on clinically validated biomarkers, such as HbA1c or oral glucose tolerance test (OGTT), are necessary to ensure consistency across datasets and enable meaningful comparison between studies [50].

ECG acquisition must be standardized, reproducible, and scalable. As discussed in Section 3.3, substantial variability exists in signal configurations, ranging from single-lead recordings to multilead clinical ECG systems. Previous studies have shown that simplified or non-standardized ECG acquisition may lead to information loss and reduced diagnostic reliability [51]. From a translational perspective, the use of multilead ECG systems with controlled signal quality, appropriate filtering, and consistent acquisition protocols is essential. At the same time, these systems must be designed for accessibility and deployment outside specialized clinical environments, which aligns with recent developments in wearable and portable ECG technologies [52,53].

Feature representation should balance physiological interpretability and robustness. While deep learning approaches enable automatic extraction of complex signal patterns, their reliance on large datasets and limited interpretability remain important challenges [54,55]. Conversely, engineered ECG features, including heart rate variability metrics and repolarization-related parameters, are grounded in well-established physiological mechanisms but may be sensitive to signal quality and preprocessing variability [56]. Hybrid approaches that combine data-driven and physiologically informed features have been increasingly proposed as a way to improve both performance and interpretability [57].

Rigorous validation strategies are required to ensure model generalizability. Internal validation methods, such as cross-validation, are insufficient to establish clinical applicability, particularly in the presence of dataset bias and potential data leakage [58,59]. External validation using independent datasets, preferably from different institutions or geographic regions, is widely recognized as a critical step in the development of reliable AI-based diagnostic systems [60]. Furthermore, prospective validation and real-world evaluation are necessary to assess performance under practical screening conditions and to account for variability in signal acquisition and patient behavior [61,62].

Clinical translation requires integration into scalable and user-friendly screening workflows. This includes not only model performance, but also system-level considerations such as ease of use, automation of signal processing, and compatibility with digital health infrastructures. The importance of end-to-end system design, encompassing data acquisition, processing, and interpretation, has been emphasized in recent studies on AI-based medical technologies [63]. In this context, the development of accessible ECG acquisition platforms combined with standardized analytical pipelines represents a key enabler for large-scale, non-invasive screening.

Taken together, these requirements define a structured pathway for advancing ECG-based dysglycemia detection toward clinical implementation. Addressing these conditions will be essential to bridge the gap between promising experimental results and reliable, scalable screening systems suitable for real-world use.

#### 4.2. Toward Practical ECG-Based Screening Systems

The analysis presented in this review indicates that the primary barrier to the clinical adoption of ECG-based dysglycemia detection is not the absence of predictive signal, but the lack of an integrated and scalable framework that connects signal acquisition, feature representation, model development, and validation within a unified pipeline. Addressing this gap requires a transition from isolated, performance-driven studies toward system-oriented approaches that explicitly account for data quality, reproducibility, and deployment constraints. Such a transition is consistent with recent trends in digital health, where end-to-end system design has been recognized as a critical factor for successful clinical translation [64].

A central element of this transition is the standardization of ECG acquisition. As demonstrated in Section 3.3, many existing studies rely on heterogeneous signal configurations, including single-lead recordings or retrospective datasets acquired under uncontrolled conditions. From a screening perspective, this variability limits both reproducibility and physiological interpretability. A practical solution is the adoption of accessible wearable multilead ECG systems capable of providing clinically relevant signal fidelity while remaining suitable for large-scale deployment. Recent developments in wearable multilead ECG platforms, including systems capable of synchronized 12-lead acquisition, demonstrate the feasibility of combining clinical-grade signal quality with scalable deployment [65]. In particular, wearable 12-lead configurations with integrated filtering, signal quality control, and synchronized acquisition offer a promising compromise between clinical-grade diagnostics and scalability [66]. Such systems enable consistent extraction of multilead features, including repolarization indices, which may be sensitive to metabolic disturbances.

Recent developments in wearable multichannel ECG platforms further support this direction by demonstrating the feasibility of combining compact hardware design with structured signal processing and data management frameworks. Importantly, these systems should be viewed not as standalone diagnostic tools, but as standardized data acquisition platforms that facilitate the generation of high-quality datasets for subsequent analysis. This perspective aligns with the requirements outlined in Section 4.1, where the availability of reliable and reproducible input data was identified as a prerequisite for meaningful model development and validation.

Beyond acquisition, the integration of feature extraction and machine learning into a coherent analytical pipeline is essential. Rather than treating these components independently, future systems should ensure that feature representation is aligned with both signal characteristics and clinical objectives. Hybrid approaches that combine physiologically interpretable features with data-driven representations may offer a viable pathway toward robust and explainable models [67]. At the same time, model development must be coupled with rigorous validation strategies, including external and prospective evaluation, to ensure generalizability across populations and settings [68].

The overall pathway from ECG signal acquisition to clinically applicable screening is illustrated in Figure 7. This framework highlights the interdependence of key components, including standardized acquisition, feature consistency, model robustness, and validation rigor, as well as the methodological bottlenecks identified in the current literature. Notably, the incorporation of

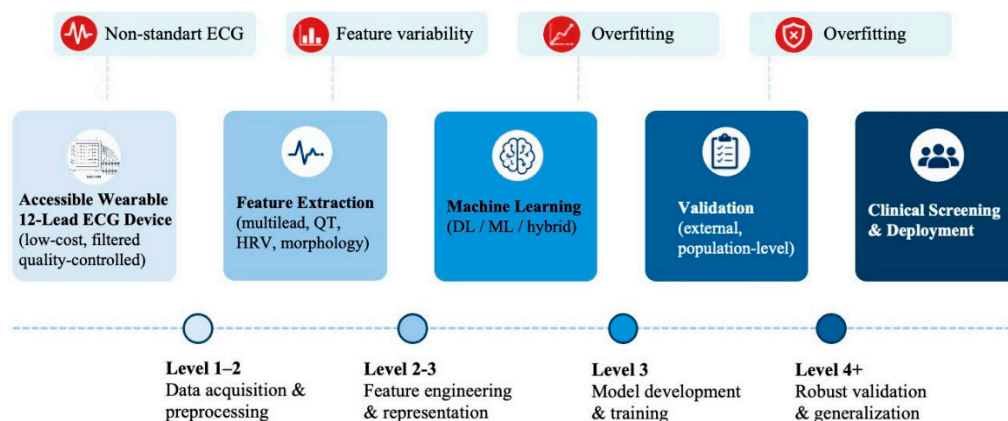
accessible wearable multilead ECG systems at the initial stage of the pipeline represents a critical enabler for overcoming existing limitations and supporting large-scale screening applications.

From a translational perspective, the development of practical ECG-based screening systems should prioritize accessibility, scalability, and integration into existing healthcare infrastructures. This includes the use of low-cost wearable devices, automated processing pipelines, and compatibility with digital health platforms. In addition, future research should focus on prospective validation studies and real-world deployment scenarios, particularly in resource-limited settings where non-invasive and scalable screening tools are most needed [69]. A proposed five-level maturity framework for translational evaluation of ECG-based dysglycemia models is presented in Table 3.

**Table 3.** Proposed maturity framework for ECG-based dysglycemia detection.

Maturity level	General characteristics	Validation status	Dataset requirements	Translational meaning
Level 1	Proof-of-concept / exploratory study	Internal only or absent	Small, highly selective cohorts	Technical feasibility only
Level 2	Initial model development	Cross-validation / train-test split	Single-center datasets	Early methodological evidence
Level 3	Clinical ML validation	Independent test set, structured retrospective evaluation	Larger clinical datasets	Moderate translational potential
Level 4	Advanced clinical validation	External validation across cohorts	Multi-center / population-based datasets	Near-translational readiness
Level 5	Real-world deployment	Prospective and implementation evaluation	Representative screening populations	Clinically deployable screening system

In summary, achieving clinically viable ECG-based dysglycemia detection requires a shift from fragmented methodological approaches toward integrated, system-level solutions. The convergence of standardized multilead acquisition, physiologically informed feature extraction, robust machine learning, and rigorous validation provides a realistic pathway toward scalable and accessible screening systems. This perspective not only addresses the limitations identified in the current literature but also outlines a practical direction for future research and development. The integrated pathway from ECG acquisition to clinically deployable dysglycemia screening is illustrated in Figure 7.



**Figure 7.** Integrated pathway from ECG acquisition to clinically deployable dysglycemia screening.

This framework emphasizes that reliable screening cannot be achieved through isolated optimization of individual components, but requires coordinated development across the entire pipeline. Standardized multilead ECG acquisition ensures consistent and physiologically meaningful input data, while robust feature representation—combining data-driven and physiologically grounded approaches—enables extraction of clinically relevant patterns. These components must be coupled with machine learning models designed for generalizability rather than dataset-specific performance, and validated through rigorous external and prospective evaluation. Importantly, the framework highlights that translational readiness is determined not by predictive accuracy alone, but by the alignment of data quality, methodological rigor, and system-level integration. This end-to-end perspective provides a practical foundation for advancing ECG-based dysglycemia detection from experimental studies toward scalable, real-world screening applications.

## 5. Limitations of This Review

This review has several methodological limitations that should be considered when interpreting the findings. Although the study was conducted following PRISMA-based principles, no formal review protocol was registered. The literature search was restricted to English-language publications indexed in major databases. Study selection and data extraction were primarily performed by a single reviewer with consistency verification, which, despite efforts to ensure accuracy, may introduce a degree of subjectivity compared to fully independent multi-reviewer procedures.

The synthesis is further constrained by substantial heterogeneity across the included studies, including differences in dataset characteristics, ECG acquisition protocols, feature representation, outcome definitions, and evaluation metrics. This variability precluded the use of quantitative meta-analysis and necessitated a narrative synthesis approach, which may be influenced by interpretative bias. Moreover, the proposed maturity framework represents a conceptual integration of the reviewed evidence rather than a formally validated model, and its generalizability to broader clinical contexts remains to be established.

## 6. Conclusion

This review synthesizes evidence from 17 studies investigating machine learning approaches for ECG-based detection of dysglycemia, highlighting both the promise and the current limitations of this emerging field. Across the reviewed literature, reported model performance is generally high, with accuracy and AUC values frequently exceeding 0.80 and in some cases approaching 0.95–0.99. However, these results are strongly dependent on dataset characteristics, as studies based on small or highly controlled cohorts tend to report substantially higher performance compared to those using larger and more heterogeneous populations. Only a limited number of studies employ large-scale datasets ( $n > 2,000$ ) or external validation, indicating that robust clinical evidence remains scarce.

From a methodological perspective, the analysis reveals pronounced heterogeneity in ECG acquisition protocols, feature representation strategies, and machine learning models, with no consensus regarding optimal approaches. While both deep learning and feature-based methods demonstrate the ability to capture dysglycemia-related patterns, model performance is primarily driven by data quality, dataset scale, and validation strategy rather than algorithmic complexity. Importantly, the majority of existing studies remain at early to intermediate stages of maturity, with relatively few approaches approaching translational readiness.

Taken together, the findings indicate that ECG-based dysglycemia detection is technically feasible but not yet clinically mature. Advancing this field will require a shift from performance-driven model development toward standardized, system-level approaches, including large and representative datasets, reproducible ECG acquisition, hybrid feature frameworks, and rigorous external and prospective validation. The integration of these elements within an end-to-end pipeline provides a realistic pathway toward scalable, non-invasive screening systems capable of deployment in real-world clinical settings.

**Author Contributions:** “Conceptualization and writing—review and editing, Ch. Alimbayev.; literature review and writing—original draft, Zh. Alimbayeva; Supervision, K. Ozhikenov; literature review, A. Ozhikenova; visualization, U. Shylmyrza; literature review, K. Khaidarova.

**Funding:** This research was funded by the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant number AP23485820.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANOVA	Analysis of Variance
AUC	Area Under the Curve
BMI	Body Mass Index
CAN	Cardiac Autonomic Neuropathy
CAD	Coronary Artery Disease
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CGM	Continuous Glucose Monitoring
CMR	Cardiac Magnetic Resonance
CMD	Coronary Microvascular Dysfunction
CNN	Convolutional Neural Network
CV	Cross-Validation
DCM	Diabetic Cardiomyopathy
DL	Deep Learning
DLM	Deep Learning Model
ECG	Electrocardiogram
EHR	Electronic Health Record
EMD	Empirical Mode Decomposition
F1-score	Harmonic Mean of Precision and Recall
FPG	Fasting Plasma Glucose
GBM	Gradient Boosting Machine
GRI	Glycaemia Risk Index
Grad-CAM	Gradient-weighted Class Activation Mapping
HbA1c	Glycated Hemoglobin
HR	Heart Rate
HRV	Heart Rate Variability
IFG	Impaired Fasting Glucose
KNN	k-Nearest Neighbors
LR	Logistic Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
NB	Naïve Bayes
OGTT	Oral Glucose Tolerance Test
PPBG	Postprandial Blood Glucose

AI	Artificial Intelligence
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RF	Random Forest
ROC	Receiver Operating Characteristic
SE	Squeeze-and-Excitation
SVM	Support Vector Machine
T2DM	Type 2 Diabetes Mellitus

## References

1. Nanda, M., Sharma, R., Mubarik, S. et al. Type-2 Diabetes Mellitus (T2DM): Spatial-temporal Patterns of Incidence, Mortality and Attributable Risk Factors from 1990 to 2019 among 21 World Regions. *Endocrine* 77, 444–454 (2022). <https://doi.org/10.1007/s12020-022-03125-5>
2. Atageldiyeva, K., Syssoyev, D., Mussina, K. et al. All-cause hospital admissions and incidence of type 2 diabetes among adolescents in Kazakhstan. *Sci Rep* 15, 20746 (2025). <https://doi.org/10.1038/s41598-025-07451-2>
3. Ziegler D, Herder C, Papanas N. Neuropathy in prediabetes. *Diabetes Metab Res Rev*. 2023;39(8):e3693. <https://doi.org/10.1002/dmrr.3693>
4. Neil H. White, Qing Pan, William C. Knowler, Emily B. Schroeder, Dana Dabelea, Emily Y. Chew, Barbara Blodi, Ronald B. Goldberg, Xavier Pi-Sunyer, Christine Darwin, Mathias Schlägl, David M. Nathan, for the Diabetes Prevention Program Outcome Study (DPPOS) Research Group; Risk Factors for the Development of Retinopathy in Prediabetes and Type 2 Diabetes: The Diabetes Prevention Program Experience. *Diabetes Care* 1 November 2022; 45 (11): 2653–2661. <https://doi.org/10.2337/dc22-0860>
5. Ahmad, A., Lim, LL., Morieri, M.L. et al. Precision prognostics for cardiovascular disease in Type 2 diabetes: a systematic review and meta-analysis. *Commun Med* 4, 11 (2024). <https://doi.org/10.1038/s43856-023-00429-z>
6. Genuth SM, Palmer JP, Nathan DM. Classification and Diagnosis of Diabetes. In: *Diabetes in America*. 3rd ed. National Institute of Diabetes and Digestive and Kidney Diseases (US), Bethesda (MD); 2018. PMID: 33651569.
7. Thomas, A., Shenoy, M. T., Shenoy, K., & George, N. . (2021). Glucometers for Patients with Type 2 Diabetes Mellitus: Are they helpful?. *International Journal of Medical Students*, 9(2), 140–144. <https://doi.org/10.5195/ijms.2021.786>
8. Samuel Seidu, Setor K. Kunutsor, Ramzi A. Ajjan, Pratik Choudhary; Efficacy and Safety of Continuous Glucose Monitoring and Intermittently Scanned Continuous Glucose Monitoring in Patients With Type 2 Diabetes: A Systematic Review and Meta-analysis of Interventional Evidence. *Diabetes Care* 2 January 2024; 47 (1): 169–179. <https://doi.org/10.2337/dc23-1520>
9. Swapna, G., Soman, K.P., Vinayakumar, R. (2020). Diabetes Detection Using ECG Signals: An Overview. In: Dash, S., Acharya, B., Mittal, M., Abraham, A., Kelemen, A. (eds) *Deep Learning Techniques for Biomedical and Health Informatics. Studies in Big Data*, vol 68. Springer, Cham. [https://doi.org/10.1007/978-3-030-33966-1\\_14](https://doi.org/10.1007/978-3-030-33966-1_14)
10. Balcıoğlu AS, Müderrisoğlu H. Diabetes and cardiac autonomic neuropathy: Clinical manifestations, cardiovascular consequences, diagnosis and treatment. *World J Diabetes*. 2015 Feb 15;6(1):80-91. doi: 10.4239/wjd.v6.i1.80. PMID: 25685280; PMCID: PMC4317320.
11. Chirag H. Mandavia, Annayya R. Aroor, Vincent G. DeMarco, James R. Sowers, Molecular and metabolic mechanisms of cardiac dysfunction in diabetes, *Life Sciences*, Volume 92, Issue 11, 2013, Pages 601-608, ISSN 0024-3205, <https://doi.org/10.1016/j.lfs.2012.10.028>.
12. Adegate, E., Singh, J. Structural changes in the myocardium during diabetes-induced cardiomyopathy. *Heart Fail Rev* 19, 15–23 (2014). <https://doi.org/10.1007/s10741-013-9388-5>
13. Jonas L. Isaksen, Christian B. Sivertsen, Christian Zinck Jensen, Claus Graff, Dominik Linz, Christina Ellervik, Magnus T. Jensen, Peter G. Jørgensen, Jørgen K. Kanters, Electrocardiographic markers in patients

- with type 2 diabetes and the role of diabetes duration, *Journal of Electrocardiology*, Volume 84, 2024, Pages 129-136, ISSN 0022-0736, <https://doi.org/10.1016/j.jelectrocard.2024.04.003>.
14. Kuehl, M., Stevens, M. Cardiovascular autonomic neuropathies as complications of diabetes mellitus. *Nat Rev Endocrinol* 8, 405–416 (2012). <https://doi.org/10.1038/nrendo.2012.21>
  15. Filipović N, Marinović Guić M, Košta V, Vukojević K. Cardiac innervations in diabetes mellitus-Anatomical evidence of neuropathy. *Anat Rec (Hoboken)*. 2023 Sep;306(9):2345-2365. doi: 10.1002/ar.25090. Epub 2022 Oct 17. PMID: 36251628.
  16. Sudo, S.Z.; Montagnoli, T.L.; Rocha, B.d.S.; Santos, A.D.; de Sá, M.P.L.; Zapata-Sudo, G. Diabetes-Induced Cardiac Autonomic Neuropathy: Impact on Heart Function and Prognosis. *Biomedicines* 2022, 10, 3258. <https://doi.org/10.3390/biomedicines10123258>
  17. Balcıoğlu AS, Müderrisoğlu H. Diabetes and cardiac autonomic neuropathy: Clinical manifestations, cardiovascular consequences, diagnosis and treatment. *World J Diabetes*. 2015 Feb 15;6(1):80-91. doi: 10.4239/wjd.v6.i1.80. PMID: 25685280; PMCID: PMC4317320.
  18. Evans, A.J.; Li, Y.-L. Remodeling of the Intracardiac Ganglia During the Development of Cardiovascular Autonomic Dysfunction in Type 2 Diabetes: Molecular Mechanisms and Therapeutics. *Int. J. Mol. Sci.* 2024, 25, 12464. <https://doi.org/10.3390/ijms252212464>
  19. Tarvainen MP, Laitinen TP, Lipponen JA, Cornforth DJ and Jelinek HF (2014) Cardiac Autonomic Dysfunction in Type 2 Diabetes – Effect of Hyperglycemia and Disease Duration. *Front. Endocrinol.* 5:130. doi: 10.3389/fendo.2014.00130
  20. Qian LL, Liu XY, Li XY, Yang F, Wang RX. Effects of Electrical Remodeling on Atrial Fibrillation in Diabetes Mellitus. *Rev Cardiovasc Med*. 2023 Jan 3;24(1):3. doi: 10.31083/j.rcm2401003. PMID: 39076858; PMCID: PMC11270397.
  21. Charlotte Coopmans, Tan Lai Zhou, Ronald M.A. Henry, Jordi Heijman, Nicolaas C. Schaper, Annemarie Koster, Miranda T. Schram, Carla J.H. van der Kallen, Anke Wesselius, Robert J.A. den Engelsman, Harry J.G.M. Crijns, Coen D.A. Stehouwer; Both Prediabetes and Type 2 Diabetes Are Associated With Lower Heart Rate Variability: The Maastricht Study. *Diabetes Care* 1 May 2020; 43 (5): 1126–1133. <https://doi.org/10.2337/dc19-2367>
  22. Alam, Krishna Chaitanya; Dasari, Dhanunjaya; Modampuri, Akhil Koundinya. A clinical study of corrected QT interval in type 2 diabetes mellitus patients. *MRIMS Journal of Health Sciences* 12(4):p 268-273, Oct–Dec 2024. | DOI: 10.4103/mjhs.mjhs\_31\_23
  23. Chávez-González E, Calero YME, Harrichand S, Mensah EB. QRS and QT Interval Modifications in Patients with Type 2 Diabetes Mellitus. *Curr Health Sci J*. 2022 Jul-Sep;48(3):270-276. doi: 10.12865/CHSJ.48.03.04. Epub 2022 Sep 30. PMID: 36815079; PMCID: PMC9940933.
  24. Singh, R.M., Waqar, T., Howarth, F.C. et al. Hyperglycemia-induced cardiac contractile dysfunction in the diabetic heart. *Heart Fail Rev* 23, 37–54 (2018). <https://doi.org/10.1007/s10741-017-9663-y>
  25. Bakkar, N.-M.Z.; Dwaib, H.S.; Fares, S.; Eid, A.H.; Al-Dhaheri, Y.; El-Yazbi, A.F. Cardiac Autonomic Neuropathy: A Progressive Consequence of Chronic Low-Grade Inflammation in Type 2 Diabetes and Related Metabolic Disorders. *Int. J. Mol. Sci.* 2020, 21, 9005. <https://doi.org/10.3390/ijms21239005>
  26. Kiruthika Balakrishnan, Durgadevi Velusamy, Karthikeyan Ramasamy, Hana E. Hinkle, Holly J. Hudson, Ram Bilas Pachori, Hikmat Khan, Artificial intelligence approaches for non-invasive diabetes prediction using ECG signals: A systematic review, *Computer Methods and Programs in Biomedicine*, Volume 278, 2026, 109264, ISSN 0169-2607, <https://doi.org/10.1016/j.cmpb.2026.109264>.
  27. Alimbayev, C.; Alimbayeva, Z.; Ozhikenov, K.; Karibayev, K.; Orynbay, Z.; Igembay, Y.; Daniyalov, M.; Nurdanali, A. Electrocardiographic Signatures of Dysglycaemia: Mechanistic Foundations, Digital Biomarkers, and Artificial Intelligence for Non-Invasive Diabetes Risk Stratification. *Appl. Sci.* 2026, 16, 2902. <https://doi.org/10.3390/app16062902>
  28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int J Surg*. 2021 Apr;88:105906. doi: 10.1016/j.ijsu.2021.105906. Epub 2021 Mar 29. PMID: 33789826.

29. Lin, C.-S.; Lee, Y.-T.; Fang, W.-H.; Lou, Y.-S.; Kuo, F.-C.; Lee, C.-C.; Lin, C. Deep Learning Algorithm for Management of Diabetes Mellitus via Electrocardiogram-Based Glycated Hemoglobin (ECG-HbA1c): A Retrospective Cohort Study. *J. Pers. Med.* 2021, 11, 725. <https://doi.org/10.3390/jpm11080725>
30. Kulkarni AR, Patel AA, Pipal KV, et al Machine-learning algorithm to non-invasively detect diabetes and pre-diabetes from electrocardiogram *BMJ Innovations* 2023;9:32-42, doi: 10.1136/bmjinnov-2021-000759
31. K. Gupta and V. Bajaj, "A Robust Framework for Automated Screening of Diabetic Patient Using ECG Signals," in *IEEE Sensors Journal*, vol. 22, no. 24, pp. 24222-24229, 15 Dec.15, 2022, doi: 10.1109/JSEN.2022.3219554.
32. Cordeiro, R.; Karimian, N.; Park, Y. Hyperglycemia Identification Using ECG in Deep Learning Era. *Sensors* 2021, 21, 6263. <https://doi.org/10.3390/s21186263>
33. S. Z. H. Naqvi, S. Aziz, M. U. Khan, M. Abbas, A. Haider and H. A. Hashmi, "Electrocardiography based System for Characterization of Diabetes," 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey, 2020, pp. 1-6, doi: 10.1109/ICECCE49384.2020.9179452.
34. Jingzhen Li, Jingyi Lu, Igbe Tobore, Yuhang Liu, Abhishek Kandwal, Lei Wang, Jian Zhou, Zedong Nie, Towards noninvasive and fast detection of Glycated hemoglobin levels based on ECG using convolutional neural networks with multisegments fusion and Varied-weight, *Expert Systems with Applications*, Volume 186, 2021, 115846, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2021.115846>.
35. J. Li, I. Tobore, Y. Liu, A. Kandwal, L. Wang and Z. Nie, "Non-invasive Monitoring of Three Glucose Ranges Based On ECG By Using DBSCAN-CNN," in *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3340-3350, Sept. 2021, doi: 10.1109/JBHI.2021.3072628.
36. Mohsen, F., Safa, A. & Shah, Z. ECG features improve multimodal deep learning prediction of incident T2DM in a Middle Eastern cohort. *Sci Rep* 15, 27164 (2025). <https://doi.org/10.1038/s41598-025-12633-z>
37. Fengade VS, Swati H, Chandak M, Rattan R, Singhal A, Kamble P, Phatak M, John N. Development of Enhanced Machine Learning Models for Predicting Type 2 Diabetes Mellitus Using Heart Rate Variability: A Retrospective Study. *Cureus*. 2025 Mar 21;17(3):e80933. doi: 10.7759/cureus.80933. PMID: 40255847; PMCID: PMC12009493.
38. Koga, D., Kaneda, R., Komiya, C. et al. Artificial intelligence identifies individuals with prediabetes using single-lead electrocardiograms. *Cardiovasc Diabetol* 24, 415 (2025). <https://doi.org/10.1186/s12933-025-02982-4>
39. Jethani, Neil & Manas Puli, Aahlad & Zhang, Hao & Garber, Leonid & Jankelson, Lior & Aphinyanaphongs, Yindalon & Ranganath, Rajesh. (2022). New-Onset Diabetes Assessment Using Artificial Intelligence-Enhanced Electrocardiography. 10.48550/arXiv.2205.02900.
40. Wang, L.; Mu, Y.; Zhao, J.; Wang, X.; Che, H. IGRNet: A Deep Learning Model for Non-Invasive, Real-Time Diagnosis of Prediabetes through Electrocardiograms. *Sensors* 2020, 20, 2556. <https://doi.org/10.3390/s20092556>
41. A. Site, J. Nurmi and E. S. Lohan, "Machine-Learning-Based Diabetes Prediction Using Multisensor Data," in *IEEE Sensors Journal*, vol. 23, no. 22, pp. 28370-28377, 15 Nov.15, 2023, doi: 10.1109/JSEN.2023.3319360.
42. D. Santhakumar, K. Dhana Shree, M. Buvanesvari, A. Saran Kumar, Ayodeji Olalekan Salau, HD-MVCNN: High-density ECG signal based diabetic prediction and classification using multi-view convolutional neural network, *Egyptian Informatics Journal*, Volume 28, 2024, 100573, ISSN 1110-8665, <https://doi.org/10.1016/j.eij.2024.100573>.
43. Chiu, I.-M.; Cheng, C.-Y.; Chang, P.-K.; Li, C.-J.; Cheng, F.-J.; Lin, C.-H.R. Utilization of Personalized Machine-Learning to Screen for Dysglycemia from Ambulatory ECG, toward Noninvasive Blood Glucose Monitoring. *Biosensors* 2023, 13, 23. <https://doi.org/10.3390/bios13010023>
44. R. Musale and A. N. Paithane, "Design and develop an algorithm for a diabetic detection using ECG signal," 2017 International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2017, pp. 961-966, doi: 10.1109/ICCMC.2017.8282610.
45. Song, H.-J.; Han, J.-H.; Cho, S.-P.; Im, S.-I.; Kim, Y.-S.; Park, J.-U. Predicting Dysglycemia in Patients with Diabetes Using Electrocardiogram. *Diagnostics* 2024, 14, 2489. <https://doi.org/10.3390/diagnostics14222489>

46. Xiaowei Zhang, Changning Liu, Yang Sun, Liangzhen You, Xiaoyu Zhang, Hongcai Shang, Clinical research on artificial intelligence medical diagnostic devices: A scoping review, *EngMedicine*, Volume 3, Issue 1, 2026, 100120, ISSN 2950-4899, <https://doi.org/10.1016/j.engmed.2026.100120>.
47. Fahim YA, Hasani IW, Kabba S, Ragab WM. Artificial intelligence in healthcare and medicine: clinical applications, therapeutic advances, and future perspectives. *Eur J Med Res*. 2025 Sep 23;30(1):848. doi: 10.1186/s40001-025-03196-w. PMID: 40988064; PMCID: PMC12455834.
48. Bartusik-Aebisher, D.; Justin Raj, D.R.; Aebisher, D. Artificial Intelligence in Medical Diagnostics: Foundations, Clinical Applications, and Future Directions. *Appl. Sci.* 2026, 16, 728. <https://doi.org/10.3390/app16020728>
49. Kelly, C.J., Karthikesalingam, A., Suleyman, M. et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 17, 195 (2019). <https://doi.org/10.1186/s12916-019-1426-2>
50. American Diabetes Association Professional Practice Committee; 17. Diabetes Advocacy: Standards of Care in Diabetes—2024. *Diabetes Care* 1 January 2024; 47 (Supplement\_1): S307–S308. <https://doi.org/10.2337/dc24-S017>
51. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019 Jan;25(1):65-69. doi: 10.1038/s41591-018-0268-3. Epub 2019 Jan 7. Erratum in: *Nat Med*. 2019 Mar;25(3):530. doi: 10.1038/s41591-019-0359-9. PMID: 30617320; PMCID: PMC6784839.
52. Neri, L.; Oberdier, M.T.; van Abeelen, K.C.J.; Menghini, L.; Tumarkin, E.; Tripathi, H.; Jaipalli, S.; Orro, A.; Paolucci, N.; Gallelli, I.; et al. Electrocardiogram Monitoring Wearable Devices and Artificial-Intelligence-Enabled Diagnostic Capabilities: A Review. *Sensors* 2023, 23, 4805. <https://doi.org/10.3390/s23104805>
53. Alimbayev, Chingiz and Alimbayeva, Zhadyra and Ozhikenov, Kassymbek and Bodin, Oleg and Mukazhanov, Yerkat, Development of Measuring System for Determining Life-Threatening Cardiac Arrhythmias in a Patient's Free Activity (February 29, 2020). *Eastern-European Journal of Enterprise Technologies*, 1(9 (103)), 12-22, 2020, doi: 10.15587/1729-4061.2020.197079, Available at SSRN: <https://ssrn.com/abstract=3703319>
54. Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat Mach Intell*. 2019 May;1(5):206-215. doi: 10.1038/s42256-019-0048-x. Epub 2019 May 13. PMID: 35603010; PMCID: PMC9122117.
55. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019 Jan;25(1):44-56. doi: 10.1038/s41591-018-0300-7. Epub 2019 Jan 7. PMID: 30617339.
56. Shaffer F and Ginsberg JP (2017) An Overview of Heart Rate Variability Metrics and Norms. *Front. Public Health* 5:258. doi: 10.3389/fpubh.2017.00258
57. Rieke, N., Hancox, J., Li, W. et al. The future of digital health with federated learning. *npj Digit. Med.* 3, 119 (2020). <https://doi.org/10.1038/s41746-020-00323-1>
58. Tuija Leinonen, David Wong, Antti Vasankari, Ali Wahab, Ramesh Nadarajah, Matti Kaisti, Antti Airola, Empirical investigation of multi-source cross-validation in clinical ECG classification, *Computers in Biology and Medicine*, Volume 183, 2024, 109271, ISSN 0010-4825, <https://doi.org/10.1016/j.combiomed.2024.109271>.
59. Nasef, D.; Nasef, D.; Basco, K.J.; Singh, A.; Hartnett, C.; Ruane, M.; Tagliarino, J.; Nizich, M.; Toma, M. Clinical Applicability of Machine Learning Models for Binary and Multi-Class Electrocardiogram Classification. *AI* 2025, 6, 59. <https://doi.org/10.3390/ai6030059>
60. Attia IZ, Tseng AS, Benavente ED, Medina-Inojosa JR, Clark TG, Malyutina S, Kapa S, Schirmer H, Kudryavtsev AV, Noseworthy PA, Carter RE, Ryabikov A, Perel P, Friedman PA, Leon DA, Lopez-Jimenez F. External validation of a deep learning electrocardiogram algorithm to detect ventricular dysfunction. *Int J Cardiol*. 2021 Apr 15;329:130-135. doi: 10.1016/j.ijcard.2020.12.065. Epub 2021 Jan 2. PMID: 33400971; PMCID: PMC7955278.
61. Kalmady, S.V., Salimi, A., Sun, W. et al. Development and validation of machine learning algorithms based on electrocardiograms for cardiovascular diagnoses at the population level. *npj Digit. Med.* 7, 133 (2024). <https://doi.org/10.1038/s41746-024-01130-8>

62. Ong Ly C, Unnikrishnan B, Tadic T, Patel T, Duhamel J, Kandel S, Moayedi Y, Brudno M, Hope A, Ross H, McIntosh C. Shortcut learning in medical AI hinders generalization: method for estimating AI model generalization without external data. *NPJ Digit Med*. 2024 May 14;7(1):124. doi: 10.1038/s41746-024-01118-4. PMID: 38744921; PMCID: PMC11094145.
63. Quer G, Arnaout R, Henne M, Arnaout R. Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2021 Jan 26;77(3):300-313. doi: 10.1016/j.jacc.2020.11.030. PMID: 33478654; PMCID: PMC7839163.
64. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. *Sci Transl Med*. 2015 Apr 15;7(283):283rv3. doi: 10.1126/scitranslmed.aaa3487. PMID: 25877894; PMCID: PMC4748838.
65. Alimbayev, C.; Alimbayeva, Z.; Ozhikenov, K.; Karibayev, K.; Orynbay, Z.; Igembay, Y.; Daniyalov, M.; Nurdanali, A. Development and Pilot Evaluation of a Wearable 12-Lead ECG System for Multilead Feature Analysis in Individuals with Different Glycemic Status. *Sensors* 2026, 26, 1598. <https://doi.org/10.3390/s26051598>.
66. Yang Y, Gao W. Wearable and flexible electronics for continuous molecular monitoring. *Chem Soc Rev*. 2019 Mar 18;48(6):1465-1491. doi: 10.1039/c7cs00730b. PMID: 29611861.
67. Lin CS, Liu WT, Chen YH, Lin SH, Lin C. Artificial intelligence-enabled electrocardiography from scientific research to clinical application. *EMBO Mol Med*. 2026 Jan;18(1):22-40. doi: 10.1038/s44321-025-00351-y. Epub 2025 Dec 1. PMID: 41326714; PMCID: PMC12808761.
68. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health*. 2020 Oct;2(10):e537-e548. doi: 10.1016/S2589-7500(20)30218-1. Epub 2020 Sep 9. PMID: 33328048; PMCID: PMC8183333.
69. Alimbayeva, Z.; Alimbayev, C.; Ozhikenov, K.; Bayanbay, N.; Ozhikenova, A. Wearable ECG Device and Machine Learning for Heart Monitoring. *Sensors* 2024, 24, 4201. <https://doi.org/10.3390/s24134201>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.