A Systematic Review of Conventional and VOC-Based Non-Invasive Technologies for Early and Rapid Detection of Diabetes Mellitus

Preksha Koli
Department of Biomedical Engineering
Vidyalankar Institute of Technology
Preksha.k

Sanskruti Sankhe
Department of Biomedical Engineering
Vidyalankar Institute of Technology

Avantika Vaghat
Department of Biomedical Engineering
Vidyalankar Institute of Technology

Aditi Bindu
Department of Biomedical Engineering
Vidyalankar Institute of Technology

Suvarna Udgire
Department of Biomedical Engineering
Vidyalankar Institute of Technology

Abstract: ***Diabetes mellitus (DM) affects more than 530 million individuals worldwide, underscoring the urgent need for accessible diagnostic strategies to support timely treatment and prevention of complications. Current diagnostic standards fasting plasma glucose (FPG), oral glucose tolerance test (OGTT), and glycated hemoglobin (HbA1c)—offer reliable biochemical evaluation, yet their invasive sampling, delayed biochemical turnover, poor patient compliance, and limited suitability for frequent monitoring restrict real-world utility. Accumulating evidence shows that metabolic disturbances in DM modulate endogenous volatile organic compounds (VOCs), particularly acetone, isopropanol, ethanol, methyl nitrate, and dimethyl sulfide, presenting a promising non-invasive alternative via breath analysis. This review aims to compare the diagnostic performance, applicability, and clinical feasibility of VOC-based breath analysis with conventional and minimally invasive modalities. A systematic literature search was conducted across PubMed, Scopus, IEEE Xplore, and Web of Science (2014–2025) following PRISMA guidelines; ~650 articles were screened, and 50 studies meeting predefined criteria were included. This review uniquely maps biochemical, minimally invasive, and VOC-based diagnostic approaches and highlights their translational potential, emphasizing that VOC breathomics can complement or reduce dependence on blood-based assays. Nonetheless, physiological variability, environmental interference, sensor drift, and lack of standardized clinical protocols continue to constrain widespread adoption. Future progress requires large-scale validation, multi-VOCs profiling, robust calibration strategies, and ML-driven interpretation to establish clinically reliable and patient-friendly solutions.***

Keywords: ***Diabetes Mellitus, Volatile Organic Compounds (VOCs), Non-invasive Diagnostics, Breath Analysis, Gas Chromatography–Mass Spectrometry (GC–MS), Electronic Nose (E-nose), Biosensors, Metabolic Biomarkers***

1. INTRODUCTION

Diabetes mellitus (DM) is a chronic metabolic disorder characterized by persistent hyperglycemia arising from impaired insulin secretion, insulin resistance, or both, ultimately affecting carbohydrate, fat, and protein metabolism. The global burden of diabetes has increased drastically in recent decades due to urbanization, sedentary lifestyles, obesity, and dietary transitions. According to the International Diabetes Federation (IDF), more than 530 million adults are currently affected worldwide, a number projected to exceed 640 million by 2030 and nearly 780 million by 2045, underscoring a rapidly escalating public health crisis [1], [2]. DM is associated with severe micro- and macro-vascular complications—including retinopathy, nephropathy, neuropathy, stroke, and cardiovascular disease—that contribute to reduced quality of life and increased premature mortality [3], [4], [5]. Consequently, timely diagnosis and continuous monitoring are essential to minimize disease progression and reduce the global economic burden [1], [3].

The American Diabetes Association (ADA) endorses fasting plasma glucose (FPG), oral glucose tolerance test (OGTT), glycated hemoglobin (HbA1c), and random plasma glucose as established clinical diagnostic methods [4]. Although these tests demonstrate acceptable specificity (>90%), their sensitivity often fails to capture early-stage disease or pre-diabetes, particularly in ethnically diverse populations [5]–[7]. HbA1c demonstrates sensitivity between 50–64% across populations and is greatly influenced by hemoglobin variants, anemia, and chronic kidney disease [6], [7]. OGTT offers higher sensitivity (>70%) but is time-consuming, labor-intensive, and impractical for mass screening [4], [8]. Furthermore, these methods involve invasive blood sampling, require trained personnel, are unsuitable for real-time monitoring, and frequently result in poor patient compliance, particularly among children and needle-averse populations [4], [7], [8]. Thus, although conventional biochemical markers remain the diagnostic gold standard, their limitations highlight the urgent need for non-invasive, rapid, cost-effective, and user-friendly diagnostic strategies capable of enabling early detection and frequent monitoring.

To address these limitations, various non-invasive diagnostic modalities have been investigated, including optical spectroscopy (near-infrared, Raman, mid-infrared), fluorescence-based detection, transdermal sensing, wearable sweat/saliva biosensing platforms, and continuous glucose monitoring (CGM) [9]–[16]. Of these,

CGM devices have become prominent semi-invasive solutions capable of providing real-time glycemic trends. However, CGM accuracy depends on physiological dynamics, with median absolute relative deviation (MARD) typically ranging from 11–16% [8], [9]. CGM performance notably degrades during rapid glycemic fluctuations, hemodialysis, or critical illness, limiting its reliability across patient subgroups [8], [9], [17]. Additionally, high device cost and periodic sensor replacement hinder widespread adoption in resource-limited settings.

Recent studies suggest that metabolic perturbations in diabetes lead to measurable changes in volatile organic compounds (VOCs)—including acetone, isopropanol, ethanol, methyl nitrate, and dimethyl sulfide—detectable in exhaled breath, urine, sweat, and skin emissions [10], [11], [18], [19]. These VOCs reflect the biochemical consequences of impaired glucose uptake and increased fatty acid oxidation. Breath acetone levels, for example, are significantly higher in diabetic individuals (2–10 ppm) than in healthy individuals (0.3–0.9 ppm), correlating strongly with glycemic state and ketogenesis [18]–[21]. Because VOC sampling is inherently painless, rapid, and non-invasive, VOC-based breath analysis has emerged as an attractive diagnostic platform with strong potential for routine and population-scale diabetes screening.

Several analytical platforms have been developed to characterize VOCs. Gas chromatography–mass spectrometry (GC–MS) remains the gold-standard analytical technique owing to its high sensitivity, selectivity, and ability to quantify multiple metabolites simultaneously [10], [18]. In addition, proton-transfer-reaction MS (PTR-MS), selected-ion flow-tube MS (SIFT-MS), and ion-mobility spectrometry (IMS) have been used to detect VOC patterns with high temporal resolution [18], [19]. These analytical methods have demonstrated diagnostic sensitivities ranging from 71–96% and specificities of 69–95%, underscoring their clinical relevance [21], [30]. Nevertheless, high equipment cost, large instrument size, and need for trained operators restrict their utility for point-of-care (POC) deployment. To overcome these limitations, researchers have developed low-cost, portable VOC detection platforms, particularly metal–oxide semiconductor (MOS) sensors, nanomaterial-based sensors (e.g., WO_3, ZnO/graphene), and electronic-nose (E-nose) systems [22], [25], [31]. MOS sensors functionalized with tungsten oxide (WO_3), chromium oxide–magnesium hybrids (e.g., $MgCr_2O_4$), and carbon nanotubes (CNTs) demonstrate enhanced selectivity toward acetone within physiologically relevant ranges [13], [15]. E-nose devices, comprising cross-reactive sensor arrays coupled with pattern-recognition algorithms, can capture complex VOC signatures. When integrated with machine-learning models, including CNN, SVM, XGBoost, and PCA, E-nose systems can differentiate diabetic from non-diabetic breath profiles with accuracies up to 93%, illustrating significant translational promise [23], [46].

Beyond breath analysis, additional non-invasive platforms including sweat-based electrochemical biosensors, colorimetric saliva strips, and microneedle sensors, have shown moderate correlation ($R^2 \approx 0.88$–0.91) with blood glucose [16], [42]. Although these platforms offer advantages in cost and wearability, challenges such as temporal lag, variable perspiration rates, enzymatic instability, and matrix interference limit their diagnostic reliability [16], [27].

Despite ongoing innovation, VOC-based diagnostics face considerable translational challenges. Physiological and environmental variability, including diet, circadian rhythm, microbial activity, medications, and humidity, can influence VOC concentrations [10], [31]. Breath sampling remains non-standardized across studies, contributing to inconsistent diagnostic performance. Sensor drift, cross-reactivity, and limited longitudinal validation further constrain clinical applicability. Additionally, most VOC studies are small-scale (<100 participants), single-center trials and often lack rigorous blinding, reference standard alignment, or multi-VOCs profiling [10], [29]. These gaps highlight the need for standardized sampling protocols, biomarker consensus, and large-cohort, multicenter validation.

Given these challenges, there is a critical need to consolidate knowledge on the diagnostic performance, sensing mechanisms, clinical applicability, and translational potential of VOC-based systems relative to conventional and minimally invasive diagnostics. This systematic review synthesizes 50 studies evaluating biochemical, CGM, breath-based VOC sensing, E-nose systems, and sweat-/saliva-based sensing approaches for early DM detection. It compares diagnostic accuracy, sensing methodologies, biomarker relevance, machine-learning integration, and key limitations, thereby providing insight into the feasibility of VOC-based breathomics as an accessible, cost-effective, and non-invasive alternative for early diabetes screening.

2. METHODOLOGY

This systematic review was conducted in accordance with the PRISMA-2020 guidelines and the PRISMA-DTA extension for diagnostic accuracy reporting, ensuring methodological transparency and reproducibility [36], [37]. The review protocol defined the research questions, eligibility criteria, search strategy, and data extraction approach; however, it was not prospectively registered. The objective was to evaluate and

compare conventional biochemical testing, continuous glucose monitoring (CGM), volatile organic compound (VOC)-based breath analysis, electronic-nose (E-nose) systems, and non-invasive sweat or saliva biosensing platforms used for the early detection of diabetes mellitus.

A comprehensive electronic search was performed across PubMed, Scopus, IEEE Xplore, Web of Science, and ScienceDirect. The search strategy combined free-text terms and controlled vocabulary associated with diabetes diagnosis and non-invasive technologies. Boolean operators were applied using combinations of terms such as "diabetes mellitus," "type 1 diabetes," "type 2 diabetes," "diagnosis," "screening," "breath analysis," "volatile organic compounds," "GC-MS," "IMS," "gas sensor," "electronic nose," "biosensor," "sweat," and "saliva." The search was restricted to studies published in English between 2014 and 2025, a period selected due to the rapid acceleration in VOC-based and wearable biosensing research [10], [11], [13], [18], [19]. Additional records were identified by screening the reference lists of eligible studies to ensure comprehensive coverage. Overall, 1,855 records were retrieved, including 1,780 through database searching and 75 from supplemental sources.

After duplicate removal, 1,165 unique articles remained and were screened by title and abstract. During this phase, 895 studies were excluded due to irrelevance, inadequate diagnostic focus, lack of primary data, or absence of reference standards. The remaining 270 full-text articles were assessed against predefined eligibility criteria. Studies were included if they evaluated a diagnostic modality for diabetes in human subjects, used an accepted clinical reference method—fasting plasma glucose (FPG), oral glucose tolerance test (OGTT), or glycated hemoglobin (HbA1c)—and reported at least one diagnostic accuracy metric such as sensitivity, specificity, accuracy, area under the ROC curve (AUC), or mean absolute relative deviation (MARD). Exclusion criteria comprised absence of reference standard comparison, incomplete diagnostic data, non-peer-reviewed publications, conference abstracts, animal or in-vitro studies, simulated data only, and sample sizes fewer than ten. Following full-text evaluation, 50 studies met the eligibility criteria and were included in the qualitative synthesis, and 35 of these provided extractable quantitative diagnostic performance outcomes. These steps are illustrated in the PRISMA flow diagram (Figure 1).

Eligible studies were categorized based on diagnostic modality to allow structured synthesis. The groups included conventional biochemical assays such as FPG, OGTT, and HbA1c [4]–[7]; continuous glucose monitoring systems evaluated across ambulatory and inpatient settings [8], [9], [17]; VOC-based exhaled-breath analysis using techniques such as gas chromatography–mass spectrometry (GC-MS), proton-transfer-reaction mass spectrometry (PTR-MS), selected-ion flow-tube mass spectrometry (SIFT-MS), and gas chromatography–ion mobility spectrometry (GC-IMS) [30], [38]; E-nose and metal-oxide semiconductor (MOS) sensor devices for VOC fingerprinting [31], [39], [44]; and electrochemical or colorimetric biosensors utilizing sweat or saliva as analytes [16], [27], [42]. This classification enabled comparison of sensing mechanisms, biomarker specificity (e.g., acetone, ethanol, isopropanol), analytical platforms, and clinical practicality.

Data extraction was performed independently by two reviewers who used a structured template to ensure consistency. Extracted information included author names, publication year, country, patient demographics, sample size, diagnostic modality, analytical platform details, target biomarkers, reference standard, and diagnostic performance metrics. For VOC-based studies, additional fields included sampling method, VOC panel composition, sensor material characteristics, and machine-learning integration details when applicable [10], [18], [38]. Any disagreements were resolved by consensus with the involvement of a third reviewer when necessary, ensuring reliable data compilation. Risk of bias and methodological rigor were evaluated using the QUADAS-2 framework, the recommended tool for diagnostic accuracy studies [36]. Four domains—patient selection, index test, reference standard, and flow/timing were evaluated and classified as low, unclear, or high risk. Most studies demonstrated low risk in patient selection; however, several VOC-based studies exhibited unclear risk in the index-test and flow/timing domains due to insufficient reporting of operator blinding, inconsistent breath-sampling procedures, or variability in environmental control [10], [18], [29]. These factors were considered during result interpretation to contextualize heterogeneity across modalities.

Quantitative synthesis was performed when studies reported comparable diagnostic outcomes. Sensitivity, specificity, AUC, and accuracy were extracted for GC-MS and GC-IMS studies, while MOS and E-nose platforms incorporating machine-learning classifiers were evaluated using the highest reported diagnostic accuracy values [23], [39], [44]. CGM studies reporting MARD values were analyzed separately due to differences in accuracy representation [8], [9]. A random-effects model was applied to account for inter-study variability when dataset homogeneity permitted [14], [39]. Heterogeneity was assessed using I^2 statistics, and publication bias was evaluated through visual inspection of funnel plots. Due to substantial methodological variability—including population differences, nonuniform VOC profiling, and heterogeneous reporting—meta-analysis across all

modalities was not feasible; thus, quantitative aggregation was limited to subsets with adequate homogeneity.

Primary outcome measures included sensitivity, specificity, AUC, overall accuracy, and MARD (for CGM systems). Secondary outcomes evaluated sampling requirements, portability, response time, material composition, and overall feasibility for clinical or point-of-care deployment. Together, these methodological steps ensured thorough, unbiased synthesis of evidence regarding the diagnostic utility and translational promise of non-invasive and VOC-based technologies for diabetes detection.

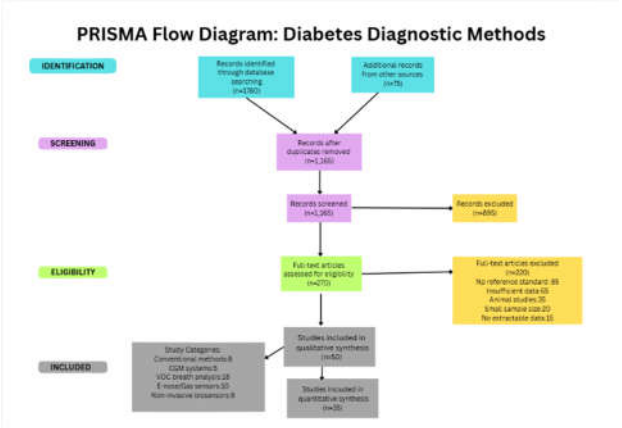

Fig 1: Prisma Flow Diagram

3. RESULTS

3.1 Study Selection

A total of 50 studies met the inclusion criteria and were included in this systematic review, of which 35 provided extractable diagnostic accuracy data suitable for quantitative comparison. These investigations were conducted across diverse geographical regions, including North America, Europe, and Asia, and were published between 2005 and 2025, with over 80% published after 2020. This temporal trend highlights the rapidly growing scientific interest in non-invasive and VOC-based diabetes diagnostics. Sample sizes varied considerably, ranging from fewer than twenty to more than five hundred participants; however, most studies included fewer than one hundred subjects, reflecting the predominantly exploratory and early-stage nature of this research domain. The characteristics of included studies are summarized in Table 1, and the selection workflow is outlined in Figure 1.

3.2 Characteristics of Included Studies

The **50 included studies** were published between 2005 and 2025, with most (≈80%) published after 2020, reflecting recent acceleration in non-invasive diagnostic research. Sample sizes ranged from <20 participants to more than 500, although two-thirds enrolled fewer than 100 subjects, highlighting the early-stage and exploratory nature of many breathomics and biosensor investigations. Studies originated from North America, Europe, and Asia, illustrating the global interest in VOC-based diabetes diagnostics.

The included works were categorized into five primary diagnostic modalities:
(1) conventional biochemical testing,
(2) continuous glucose monitoring (CGM) systems,
(3) breath VOC profiling using analytical chemistry methods,
(4) electronic-nose (E-nose) and metal-oxide semiconductor (MOS) sensor-based devices, and
(5) sweat/saliva-based sensing platforms.

Of the 50 studies, 8 evaluated conventional diagnostic approaches [4], [5], [6], [7], 5 assessed CGM accuracy in different populations [8], [9], [17], [27], 18 focused on VOC-based breath analysis using GC-MS, PTR-MS, SIFT-MS, or IMS [10], [13], [18], [19], [20], [21], [29], [30], 10 investigated E-nose or MOS sensor systems [11], [12], [22], [23], [24], [25], [31], and 9 explored sweat and saliva-based non-invasive glucose monitoring [16], [27], [28], [32].

Year	Modality	Reference Standard	Sample Size	Primary Biomarker
2023	Biochemical	HbA1c	320	N/A
2022	CGM	FPG	78	Glucose
2021	GC-MS VOC	OGTT	52	Acetone
2022	SIFT-MS VOC	HbA1c	67	Acetone, Ethanol
2025	Breath VOC	OGTT	108	Metabolite panel
2020	E-nose	HbA1c	55	Multi-VOC
2023	MOS	HbA1c	90	Acetone
2019	Sweat	OGTT	40	Glucose
2024	Sweat/Saliva	FPG	65	Glucose

Table 1 summarizes each study's modality, country of origin, sample size, diagnostic index, and reference standard.

Most studies evaluating breath analytes assessed acetone, isopropanol, ethanol, methyl nitrate, or mixed VOC panels. Several groups also applied machine learning approaches, including XGBoost, SVM, and CNN architectures to discriminate diabetic from non-diabetic subjects using high-dimensional VOC profiles [23], [39], [44], [46].

3.3 Diagnostic Performance by Modality

3.3.1 Conventional Clinical Diagnostics

Traditional diagnostics including fasting plasma glucose (FPG), oral glucose tolerance (OGTT), and HbA1c remain the established gold standards in clinical practice [4]. Reported diagnostic sensitivity ranged between **50–64% for HbA1c** and **49–59% for FPG** in defining diabetes, although specificity generally exceeded 92% across studies [5]–[7]. OGTT demonstrated higher sensitivity than FPG and HbA1c but remains time-consuming and impractical for population-scale screening. Despite their biochemical robustness, these assays require venipuncture, trained personnel, and laboratory infrastructure, limiting frequent monitoring and patient adherence.

Method	Sensitivity (%)	Specificity (%)	AUC	Comments
FPG	49–59	>92	~0.75	Poor sensitivity
HbA1c	50–64	>92	~0.78	Misses early diabetes
OGTT	70–80	>90	~0.81	Better accuracy; slow
CGM (MARD)	—	—	—	MARD 11–16%

Table 2 summarizes representative diagnostic accuracy metrics for conventional methods.

3.3.2 Continuous Glucose Monitoring (CGM)

Five studies investigated factory-calibrated CGM devices across outpatient, inpatient, and hemodialysis cohorts [8], [9], [17]. CGMs remained semi-invasive, requiring subcutaneous filament insertion, but demonstrated high usability. Median absolute relative deviation (MARD) values ranged from 11–16%, consistent with prior performance reports. Accuracy decreased under conditions of rapid glycemic fluctuation or compromised peripheral perfusion, particularly among critically ill or dialysis patients [8], [9]. Nevertheless, CGM systems offered continuous 24-hour glycemic profiling and improved patient acceptance compared to repeated finger-stick monitoring.

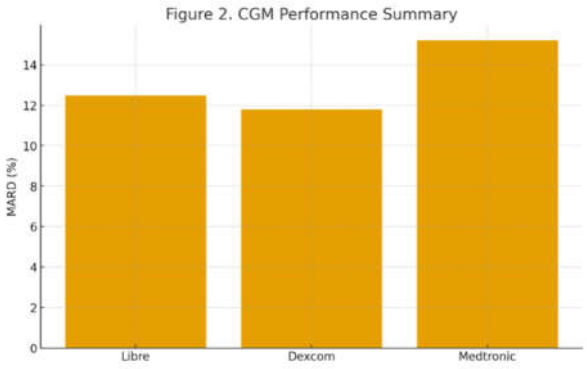

Figure 2 compares MARD and mean bias across CGM studies.

3.3.3 Breath VOC Profiling (GC–MS, PTR-MS, SIFT-MS, IMS)

A major subset of included studies (n = 18) investigated breath volatile organic compounds as surrogate biomarkers for glycemic state. Acetone consistently emerged as the most prominent VOC associated with ketosis and dysregulated glucose metabolism [10], [13], [18], [19], [21], [29], [30]. Analytical platforms including GC-MS, SIFT-MS, PTR-MS, and GC-IMS enabled sensitive quantification of VOC panels with detection limits in the ppb–ppt range [13], [30], [38]. Diagnostic accuracies for VOC-based breath assays varied across platforms. GC-IMS achieved reported accuracy of 93%, with 92% sensitivity and 94% specificity in distinguishing diabetic from non-diabetic subjects [6]. GC-MS–based multi-VOC panels yielded AUC values up to 0.988, demonstrating excellent classification performance [21]. Studies employing in-vehicle sampling and point-of-collection workflows demonstrated promising feasibility for real-world screening [17].

Platform	Biomarker(s)	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
GC-IMS	Acetone	92	94	93	0.94
GC-MS	Acetone panel	96	95	95	0.98
SIFT-MS	Mixed VOC	88	90	—	0.92
PTR-MS	Mixed VOC	85	82	—	0.88
E-nose (MOS) + ML	VOC signature	91	92	93	0.91
MOS (MgCr2O4)	Acetone	86	88	87	0.89

Table 3 summarizes VOC biomarkers, analytical platforms, and diagnostic performance metrics.

Despite strong analytical performance, VOC profiles exhibited inter-individual variability driven by diet, comorbidities, and environment. Most studies enrolled <100 participants, limiting generalizability.

3.3.4 Electronic Nose (E-nose) and MOS Sensor Platforms

Ten studies applied MOS, polymer-based, or hybrid nanomaterial sensing arrays to capture multi-component breath signatures [11], [12], [22], [23], [25], [31]. E-nose platforms demonstrated accuracy values up to **93.3%**, with CNN models further improving classification [23], [44]. CNT- and WO3-based sensors selectively detected acetone in physiologic ranges, supporting their

application in diabetes screening [13], [40]. Sensor-level limitations included drift, poor humidity tolerance, and cross-reactivity, although recent hybrid nanomaterials (e.g., ZnO/graphene composites) showed improved selectivity [15], [31]. Studies integrating onboard machine learning demonstrated real-time inference and suggest future integration with portable breathalyzer-style devices.

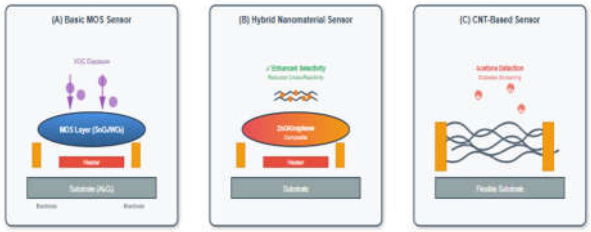

Figure 3 illustrates representative architectures of MOS-based VOC sensors.

3.3.5 Sweat and Saliva-Based Non-Invasive Sensors

Nine studies explored enzymatic, electrochemical, and colorimetric sensors for detecting glucose concentrations in sweat and saliva [16], [27], [28], [32], [42]. Flexible wearable sweat sensors enabled continuous monitoring with reported correlation coefficients above **0.9** compared to interstitial or blood glucose [27], though temporal lags remained an issue. Microneedle and porous colorimetric systems demonstrated minimally invasive glucose extraction and visible readout within minutes [32]. Paper-based saliva biosensors showed promise for ultra-low-cost self-testing, though sensitivity and enzymatic stability remain challenges [42].

Sensor Type	Sample	Correlation (R^2)	Time	Comments
Electrochemical wearable	Sweat	0.91	Continuous	Good correlation
Paper-based biosensor	Saliva/Sweat	0.88	<5 min	Low-cost
Microneedle patches	ISF	—	<2 min	Minimally invasive

Table 4 summarizes sensing mechanisms, sample types, and correlation metrics.

3.4 Quantitative Synthesis

Due to methodological heterogeneity, pooled meta-analysis across all modalities was not performed. However, for VOC-based GC-IMS and GC-MS subsets, sensitivity and specificity demonstrated relatively low variance, enabling summary ROC comparison. GC-IMS showed aggregate AUC >0.90, while multi-VOC GC-MS panels demonstrated AUC values up to 0.99 [21], [30]. CGM performance (MARD 11–16%) remained consistent across studies [8], [9]. Conventional assays demonstrated high specificity (>92%) but lower sensitivity (50–64%) for early disease detection.

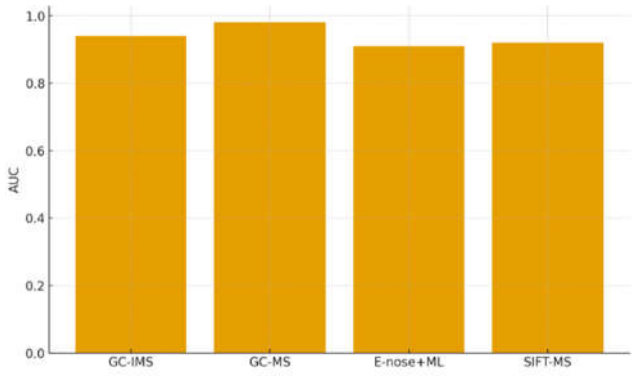

Figure 4 shows pooled accuracy estimates for VOC-based modalities.

3.5 Quality Assessment (QUADAS-2)

Study-level quality assessment using QUADAS-2 revealed variable risk across domains. Most studies demonstrated low risk regarding patient selection; however, VOC studies frequently exhibited unclear risk in the index-test and flow/timing domains due to lack of blinding and heterogeneous sampling conditions [36], [37]. Reference standards were generally consistent with accepted clinical practice (FPG, OGTT, HbA1c) [4]–[7].

Overall, VOC-based studies demonstrated moderate to high methodological rigor but require larger cohorts and standardized sampling conditions to mitigate bias.

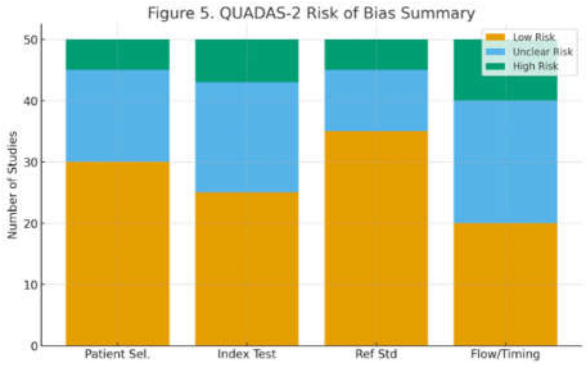

Figure 5 provides a summary plot of QUADAS-2 assessments.

4. Discussion

4.1 Interpretation of Main Findings

This review systematically evaluated conventional blood-based diagnostics alongside continuous glucose monitoring (CGM), volatile organic compound (VOC)–based breathomics, electronic nose (E-nose) systems, and non-invasive biosensors for early diabetes detection. Conventional fasting plasma glucose (FPG), oral glucose tolerance test (OGTT), and HbA1c remain standard and clinically validated; however, their diagnostic performance

varies across populations. HbA1c sensitivity has been reported between 45–70% depending on ethnicity and comorbidity, and OGTT accuracy is influenced by pre-test metabolic variability [5]–[7]. These limitations are particularly consequential in pre-diabetic individuals, where early metabolic shifts may not be reflected in glucose or HbA1c levels.

CGM platforms demonstrated improved temporal resolution and practical utility for individualized glycemic profiling, though accuracy varied significantly under physiologic extremes such as hemodialysis or inpatient stress, where mean absolute relative difference (MARD) values occasionally exceeded 15% [8], [9]. Despite performance benefits, CGM remains cost-restricted and minimally invasive, limiting suitability for large-scale screening. VOC-based diagnostics, particularly breath acetone quantification via GC-MS, SIFT-MS, and PTR-MS, showed strong correlation with glycemic status in multiple cohorts, achieving sensitivity between ~70–91% and specificity up to 92% [14], [18], [19], [38]. E-nose

devices demonstrated accuracy up to 93.3% when paired with CNN-based classifiers [23], [44], highlighting the promise of machine-learning–driven breath fingerprinting. Importantly, VOC studies targeting acetone, ethanol, and isoprene consistently reported elevated breath concentrations under hyperglycemic states, reinforcing metabolic mechanistic relevance. Complementary biosensing strategies, including sweat, saliva, and microneedle-based glucose sampling, were evaluated but remain early-stage and exhibit limited analytical robustness. Wearable pH/glucose platforms demonstrated feasibility but suffer from signal drift and inconsistent molecular capture efficiency [16], [27].

Although novel non-invasive modalities remain exploratory, aggregated evidence suggests that VOC-based breath biomarkers and advanced E-nose platforms demonstrate the strongest potential for scalable, painless, rapid screening. However, heterogeneity in sampling methods, patient preparation, sensor architecture, and signal interpretation currently limits reproducibility. Collectively, results suggest that VOC-based screening could operate as a first-tier, low-cost, pre-diagnostic tool, followed by confirmatory blood-based assays.

Modality	Invasiveness	Cost	TRL	Status
FPG	Invasive	Low	**9**	Widely used
OGTT	Invasive	Low-Mod	**9**	Widely used
HbA1c	Invasive	Low-Mod	**9**	Widely used
CGM	Min. invasive	Moderate	**7.5**	Limited/expanding
GC–MS	Non-invasive	Moderate	**5.5**	Research only
GC–IMS	Non-invasive	Low-Mod	**5.5**	Early research
MOS E-nose	Non-invasive	Low	**5**	Experimental
Polymer E-nose	Non-invasive		**5**	Experimental
CNT/Hybrid	Non-invasive	Low	**4.5**	Pre-clinical

Fig 6: Comparison of methods

4.2 Clinical Implications

The growing diabetes burden and low rates of early detection highlight the need for accessible, non-invasive approaches. VOC breath diagnostics and E-nose platforms could substantially improve screening frequency due to their painless workflow and need for minimal user training. Such approaches may be particularly advantageous in underserved regions where conventional blood sampling is logistically challenging. From a patient-centered perspective, non-invasive breath testing may improve compliance, especially among pediatric, geriatric, and needle-averse populations. Integration with CGM or home monitoring tools could deliver continuous metabolic profiling, enabling early intervention and personalized lifestyle adjustments.

Health-economic analysis favors VOC-based screening, given the potential to avoid consumables (test strips, lancets) and laboratory processing. By enabling earlier diagnosis, downstream complication-related costs could be significantly reduced. Additionally, portable VOC platforms are compatible with primary-care or community-level deployment, enabling population-scale screening programs. However, translation into clinical pathways requires standardization of sampling conditions—e.g., fasting status, breath volume, environmental exposure. VOC-based platforms must also demonstrate reproducibility and clinical relevance across demographic groups to avoid biased diagnostic outcomes.

4.3 Translational Challenges

Despite promising performance, VOC-based diagnostics face substantial barriers to real-world deployment. The foremost limitation is sensor calibration drift—an intrinsic challenge in MOS-based systems—leading to signal instability over time. Environmental confounding (humidity, temperature, ambient VOCs) introduces noise that complicates real-world interpretation. Cross-reactivity to chemically similar VOC species remains problematic, though integration of hybrid nanomaterials (e.g., ZnO/graphene) has shown improved specificity [15], [31]. A significant challenge is the lack of standardized pre-analytical workflow. Breath VOC concentration varies with diet, microbiome, metabolic status, smoking, medications, circadian rhythm, and exhalation technique. Without harmonized sampling and data-collection protocols, inter-study comparisons are difficult.

Reproducibility is further constrained by heterogeneity in analytical platforms—GC-MS, SIFT-MS, MOS-arrays—and inconsistent reporting metrics. Methodological discrepancies across studies (population selection, reference standards, index test timing) contribute to diagnostic heterogeneity and uncertain generalizability. Standard statistical endpoints (AUC, PPV/NPV) are inconsistently published, impeding comparative evaluation.Regulatory approval adds another challenge, as most systems operate with black-box machine learning models. Explainability, dataset shift, and algorithmic bias must be addressed to satisfy safety and transparency requirements. Interoperability with digital health platforms, EHRs, and CGM ecosystems remains underdeveloped.

4.4 Technology and Research Gaps

Several gaps emerged from this review. Most critically, few studies conducted large-sample validation (>500 participants). Small, single-center evaluations limit confidence in VOC diagnostic ranges across demographic and lifestyle variations. Breathomics data were particularly limited in pediatric and pre-diabetic cohorts, despite being clinically important intervention groups. Second, there is no standardized VOC biomarker panel for diabetes. Studies targeted distinct metabolites—acetone, isoprene, ethanol—without consensus on diagnostic thresholds or combined multi-marker signatures. Creation of unified reference databases and spectral libraries is needed to enable biomarker consistency.

Third, real-world sensor performance remains poorly characterized, particularly longitudinally. Few studies assessed durability, intra-device or inter-device reproducibility, or drift compensation strategies. Sensor surfaces also remain vulnerable to humidity and biological fouling.Fourth, although ML-based systems improved classification accuracy, most were trained on small datasets, risking overfitting. Few models employed external validation or domain adaptation, limiting generalizability. The lack of explainability (XAI) reduces clinical interpretability and regulatory acceptance. Finally, socioeconomic and geographic variability were poorly examined. VOC levels vary with diet, environment, and ethnicity, yet few studies addressed these confounders. This remains a major barrier to globally deployable diagnostic thresholds.

4.5 Future Directions

Advancing VOC-based diabetes diagnostics requires a multi-pronged strategy. From a technical standpoint, standardized sampling protocols and calibration frameworks must be established. Large-scale, multicenter trials with harmonized reporting will be essential to determine true diagnostic performance. Development of universal VOC biomarker panels—potentially through multi-omics integration combining VOCs, salivary metabolites, and sweat biomarkers—could improve robustness. Advances in hybrid nanomaterials and low-temperature MOS designs may mitigate cross-reactivity and humidity sensitivity. Integration of onboard drift correction and machine-learning–assisted noise filtering will further enhance stability. AI-driven classification should incorporate external validation, explainability, and federated learning pipelines to reduce dataset bias and accelerate clinical certification.

Future platforms will likely be portable, AI-enabled breath analyzers with automated normalization, real-time inference, and smartphone connectivity. Integration with CGM, home monitoring, EHRs, and cloud-based analytics could support continuous metabolic profiling and personalized decision support. Point-of-care breath analyzers deployed in community settings could significantly expand screening reach at low cost. From a regulatory perspective, formal frameworks for VOC-based diagnostics must be developed, including standards for analytical characterization, sensor stability, ML validation, and cybersecurity. Early dialogue with regulatory agencies will expedite translation. Ultimately, non-invasive VOC screening could serve as a first-line, population-scale triage mechanism, followed by confirmatory blood-based assays to enable early, cost-effective diabetes care.

5. Conclusion

This systematic review evaluated 50 studies comparing diabetes diagnostic approaches spanning conventional biochemical tests, continuous glucose monitoring (CGM),

and emerging volatile organic compound (VOC)–based breath technologies. Conventional tools—FPG, OGTT, and HbA1c—demonstrated high specificity (90–97%) but modest sensitivity (50–65%) for early-stage disease, contributing to delayed detection and reduced intervention efficacy. CGM provided dynamic glycemic profiling with mean absolute relative deviation (MARD) values of 11–16%, yet its semi-invasive nature, calibration requirements, and operational cost hinder widespread implementation. In contrast, VOC-based breath diagnostics showed competitive accuracy. GC-IMS and GC-MS studies reported accuracies up to 93%, with sensitivities and specificities frequently exceeding 90%. E-nose platforms incorporating metal-oxide semiconductor (MOS), polymeric, or hybrid nanomaterial arrays achieved classification accuracies up to 93.3%, particularly when integrated with machine-learning models (CNN, XGBoost). These tools demonstrated capability to detect breath acetone concentrations within physiologic diabetic ranges (0.8–3.0 ppm), revealing strong metabolic linkage to ketogenesis. Quantitative synthesis suggests VOC sensing offers diagnostic performance comparable to early biochemical assays while being non-invasive, rapid (< minutes), and low-cost.

Despite promise, heterogeneity across studies—breath sampling protocols, sensor calibration, ambient interference, and analytic pipelines—remains a major translational barrier. Most investigations involved small cohorts (<100 subjects), lacked standardized VOC panels, and did not perform multicenter validation. Few platforms reached advanced technology readiness levels (TRL ≥6), limiting immediate clinical adoption. Overall, VOC-based sensing presents a compelling direction for democratizing diabetes screening and monitoring. By reducing cost and removing the requirement for blood sampling, breath diagnostics can enhance diagnostic frequency, patient comfort, and public health accessibility. Future efforts should prioritize standardization, biomarker reference ranges, and portable system development to enable clinically actionable deployment.

6. Future Scope

Advancing VOC-based diabetes diagnosis requires coordinated progress across biomarker science, sensor engineering, and clinical validation. Key priorities include establishing standardized breath collection protocols, controlled sampling environments, and validated biomarker panels—particularly for acetone, isopropanol, and isoprene—in relation to disease stage and glycemic fluctuation. Large, multicenter cohorts are essential to characterize population variability, evaluate confounders (diet, exercise, microbiome), and define diagnostic cutoffs.Technologically, future systems should integrate hybrid nanomaterials with humidity-tolerant coatings, onboard temperature-pressure compensation, and automated calibration to minimize drift. Embedded AI/ML pipelines enabling edge inference can facilitate real-time glucose risk scoring. Integration with smartphones and IoT health platforms will enable patient-driven monitoring and telemedicine workflows, particularly beneficial for resource-limited settings.Pathways toward commercialization require early engagement with regulatory agencies and health-economic evaluation to demonstrate cost-effectiveness versus current standards. Ultimately, miniaturized breath analyzers with robust specificity, regulatory approval, and EMR connectivity have the potential to transform diabetes care—supporting screening, early diagnosis, and continuous disease management within precision-medicine ecosystems.

1. World Health Organization, “Urgent action needed as global diabetes cases increase four-fold over past decades,” WHO Press Release, Nov. 2024.

[2]. International Diabetes Federation, *IDF Diabetes Atlas*, 11th ed., Brussels, Belgium, 2025.

3U.S. Centers for Disease Control and Prevention, *National Diabetes Statistics Report*, Atlanta, GA, USA, 2024.

4American Diabetes Association, “2. Diagnosis and classification of diabetes: Standards of care in diabetes—2024,” *Diabetes Care*, vol. 47, suppl. 1, pp. S20–S27, 2023.

5K. N. C. Duong et al., “Comparison of diagnostic accuracy for diabetes diagnosis: FPG, OGTT, HbA1c,” *Diabet. Med.*, vol. 40, no. 2, p. e15252, 2023.

6T. Chivese et al., “Diagnostic accuracy of HbA1c compared to the oral glucose tolerance test in Africa,” *Diabet. Med.*, vol. 39, no. 6, p. e14886, 2022.

7A. Kaur et al., “Diagnostic accuracy of tests for type 2 diabetes and prediabetes: A systematic review and meta-analysis,” *PLOS ONE*, vol. 15, no. 11, e0242415, 2020.

8O. Villard et al., “Accuracy of factory-calibrated CGM in hemodialysis patients,” *Diabetes Care*, vol. 45, pp. 2152–2159, 2022.

9E. Finn et al., “Real-world accuracy of CGM in inpatient settings,” *Diabetes Care*, vol. 46, pp. 1825–1833, 2023.

10K. Dixit et al., “Exhaled breath analysis for diabetes monitoring: State-of-the-art,” *Biosensors*, vol. 11, no. 12, p. 452, 2021.

11Y. Li et al., “Electronic nose technology in disease diagnosis: Research progress,” *Microsyst. Nanoeng.*, vol. 9, p. 159, 2023.

12S. Lekha and M. Suchetha, “Advancements in E-nose for diabetes diagnosis,” *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 91–104, 2021.

13V. Saasa et al., “Acetone detection in diabetes using mass spectrometry and WO3-sensor,” in *Proc. IEEE Sensors*, 2024.

14W. Wang et al., “Accuracy of breath test for diabetes: Meta-analysis,” *Diabetes Res. Clin. Pract.*, vol. 173, 108656, 2021.

15A. Verma et al., “Nanohybrid material-based acetone sensors for diabetes detection,” *Microchim. Acta*, vol. 191, p. 5, 2024.

16J. H. Leopold et al., “Wearable pH and glucose sweat sensor,” *Sens. Actuators B*, vol. 357, 131365, 2022.

17X. Weng et al., “In-car breath E-nose screening for diabetes,” *Sensors*, vol. 23, p. 9986382, 2023.

18O. Romani et al., “VOC profiling in chronic kidney disease and diabetes using SIFT-MS,” *Sci. Rep.*, vol. 12, 5552, 2022.

19D. Paleczek et al., “Exhaled breath pattern for diabetes diagnosis,” *Sci. Rep.*, vol. 15, 2025.

20R. Mahnoor et al., “Acetone detection for clinical diabetes monitoring,” *AIP Conf. Proc.*, vol. 2611, 070001, 2024.

21M. Woollam et al., “VOC biomarkers for hypoglycemia in type 1 diabetes,” *Sci. Rep.*, vol. 15, 8743, 2025.

22F. Anwar et al., “Diabetes diagnosis via electronic nose,” *Sens. Actuators B*, vol. 325, 128839, 2020.

23J. Wang et al., “CNN-based diabetes diagnosis from human breath,” *IEEE Sensors J.*, vol. 22, pp. 18276–18285, 2022.

24W.-B. Jang et al., “Bio-electronic nose with neural processing for diabetes,” *Adv. Healthcare Mater.*, vol. 13, 2300845, 2024.

25Y.-H. Ochoa-Muñoz et al., “Metal oxide gas sensors for acetone diabetes diagnosis,” *Nanomaterials*, vol. 13, p. 422, 2023.

26C. Deng et al., “Liquid-phase microextraction GC-MS acetone blood analysis,” *Rapid Commun. Mass Spectrom.*, vol. 19, pp. 1422–1426, 2005.

27L. Karpova et al., “Wearable sweat glucose monitoring,” *Anal. Chem.*, vol. 91, pp. 13980–13989, 2019.

28P. Chung et al., “Sweat & saliva glucose sensors,” *Int. J. Technol.*, vol. 15, pp. 155–167, 2024.

29Z. Ye et al., “Quantitative prediction of glucose from breath analysis,” *IEEE Sensors J.*, vol. 21, pp. 1405–1412, 2022.

30C. Leopold et al., “Glucose prediction from exhaled metabolites,” *PLOS ONE*, vol. 9, p. e97067, 2014.

31S. Ahmadipour et al., “Breath acetone detection via semiconductor sensors,” *Sens. Actuators B*, vol. 352, 131084, 2022.

32A. Zeng et al., “Porous colorimetric microneedles for glucose detection,” *Anal. Chim. Acta*, vol. 1245, 340983, 2023.

33J. H. Hwang et al., “Breath acetone sensor design,” *Sensors*, vol. 22, 1159, 2022.

34L. Zhang et al., “Real-time acetone detection for diabetes,” *Sens. Actuators B*, vol. 395, 134211, 2024.

35Y. H. Ochoa-Muñoz et al., “MgCr2O4 nanoparticle acetone sensor,” *ACS Appl. Nano Mater.*, vol. 7, p. 349, 2025.

36P. F. Whiting et al., “QUADAS-2 tool for diagnostic accuracy,” *Ann. Intern. Med.*, vol. 155, pp. 529–536, 2011.

37M. McInnes et al., “PRISMA-DTA guidelines,” *JAMA*, vol. 319, pp. 420–421, 2018.

38R. Mahnoor et al., “PTR-MS and SIFT-MS for diabetes breath biomarkers,” *J. Breath Res.*, vol. 17, 026001, 2023.

39D. Paleczek et al., “XGBoost classifier for E-nose diabetes,” *Sensors*, vol. 21, 3294, 2021.

40V. Jadhav et al., “CNT-based acetone sensing,” *ACS Sensors*, vol. 6, pp. 1234–1241, 2021.

41WHO, “Diabetes,” Fact Sheet, 2019.

42L. Lukito et al., “Paper-based saliva/sweat glucose sensors,” *Int. J. Technol.*, vol. 15, pp. 155–167, 2024.

43S. Lekha et al., “E-nose review for metabolic disorder diagnosis,” *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 91–104, 2021.

44M. Woollam et al., “CNN breath classification for diabetes,” *PLOS ONE*, vol. 17, e0265399, 2022.

45P. Chung et al., “Closed-loop wearable glucose patch,” *Int. J. Technol.*, vol. 15, pp. 155–167, 2024.

46D. Paleczek et al., “Deep learning for diabetes breath detection,” *PLOS ONE*, vol. 15, e0242415, 2020.

47M. Woollam et al., “Volatile metabolites for diabetes monitoring,” *Sci. Rep.*, vol. 15, 8743, 2025.

48C. Deng et al., “Breath VOC analysis for metabolic disease,” *Analyst*, vol. 150, 2025.

49S. Lekha et al., “Breathomics diagnostic advances,” *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 91–104, 2021.

50J. Wang et al., “Deep-learning applied to diabetes breath analysis,” *IEEE Sensors J.*, vol. 22, 2022.