



Identifying Air Bubble Errors in Arterial Blood Gas Tests Using Machine Learning Models

S. Ramadoss*¹ & A. Kumaravel²

*¹ Research Scholar/Dept of CSE, School of Computing, Bharath Institute of Science and Technology, Chennai, Tamil Nadu, India.

*ramadoss90@hotmail.com

² Professor/Dept of IT, School of Computing, Bharath Institute of Science and Technology, Chennai, Tamil Nadu, India

drkumaravel@gmail.com

Abstract. Among the pre-analytical errors that compromise arterial blood gas (ABG) testing, air bubble contamination stands out for its frequency and its ability to produce systematic, clinically dangerous distortions. Atmospheric air entering a collection syringe raises PO₂, suppresses PCO₂, and shifts pH toward alkalosis — a pattern that can be mistaken for genuine respiratory alkalosis. Addressing this problem, we built and tested a detection framework rooted in gas-diffusion physiology, working with 800 expert-annotated ABG samples. Deviations in pH, PO₂, and PCO₂ from their respective reference values were computed, and a composite rule requiring any two of three abnormal thresholds to be met simultaneously was applied. The rule yielded sensitivity of 87.6% and specificity of 91.2%, separating true contamination from real respiratory disorders with clinically acceptable accuracy. Feeding these error-derived features into SVM, Random Forest, and Logistic Regression classifiers raised accuracy by 5.2–5.6%; alone, they accounted for 34.2% of feature importance in the Random Forest model. Electrolyte and metabolic parameters, as expected from diffusion physics, did not differ between contaminated and clean samples, lending additional confidence to the specificity of the detected signal. Fast enough for point-of-care use and transparent enough for bedside explanation, the method illustrates why encoding known data-quality problems directly into model pipelines pays dividends in clinical AI.

Keywords: arterial blood gas; air bubble contamination; pre-analytical error; clinical decision support; machine learning; error-based modeling

1 Introduction

1.1 Problem Identification

Clinicians managing critically ill patients depend heavily on arterial blood gas testing. In emergency settings, ICUs, and around surgical procedures, ABG results guide decisions on ventilator adjustments, oxygen delivery, and correction of acid-base

disturbances, choices with immediate consequences for survival [1]. Because so much hinges on these numbers, any factor that corrupts them upstream of the analyzer has the potential to trigger a harmful clinical response. Sample quality is therefore not a peripheral concern; it sits at the very foundation of safe ABG-based decision-making.

Errors introduced before a sample ever reaches the analyzer are remarkably common [2]. They arise at collection, during transport, or in storage, and many go unrecognized. Air bubble contamination — classified here as E1 — is among the most frequent offenders. When atmospheric air contacts blood inside the syringe, gas diffusion begins almost immediately: oxygen migrates from the bubble (partial pressure ~ 160 mmHg) into the blood, carbon dioxide travels the other way, and the resulting CO_2 loss pushes pH upward [3]. The net effect — high PO_2 , low PCO_2 , alkaline pH — mimics the pattern of respiratory alkalosis, creating real uncertainty about whether the result reflects the patient or the syringe.

When contamination goes undetected, the clinical fallout can be serious. A spuriously high PO_2 reading may prompt staff to reduce supplemental oxygen in a patient who still needs it, inviting hypoxemia [4]. A falsely low PCO_2 paired with elevated pH can prompt ventilator changes or trigger investigation for pulmonary embolism and sepsis that would otherwise be unnecessary [5]. These are not edge cases — they represent the kind of error that flows naturally from misplaced confidence in an abnormal result. In automated decision-support pipelines, contaminated inputs amplify the problem further, because the algorithm has no way of knowing the data it received was corrupted.

Current safeguards — visual checks, analyzer flags, and staff experience — are too variable to solve this problem at scale [6]. Judgment differs between nurses, shifts, and hospitals. High-throughput clinical environments and machine learning pipelines running on historical data have no mechanism at all for applying human scrutiny. What is needed is a method that is explicit, rule-based, fast, and grounded in the physiology of why contamination produces the parameter shifts it does — one that can run automatically at the point of collection or during data preprocessing without human intervention.

1.2 Objectives

The aims of this study span six related goals:

- (1) To develop a quantitative framework for modeling air bubble contamination effects on ABG parameters, grounded in physiological gas diffusion principles

- (2) To establish clinically meaningful threshold values for detecting contamination using deviations from established reference standards

- (3) To formulate a composite detection rule that captures the characteristic triad of pH elevation, PO_2 elevation, and PCO_2 reduction

- (4) To validate the detection framework against expert-annotated ABG samples with known error types

- (5) To demonstrate the integration of error-derived features into machine learning pipelines for improved classification performance

(6) To provide a clinically interpretable foundation for automated error detection in critical care decision support systems

1.3 Scope of the Project

This research focuses specifically on air bubble contamination (E1) as a distinct pre-analytical error type in ABG analysis. The scope encompasses the following dimensions:

Data Scope: Analysis of 800 ABG samples with comprehensive annotations including 13 physiological parameters — pH, PCO₂, PO₂, sodium, potassium, calcium, lactate, hematocrit, bicarbonate, total CO₂, oxygen saturation, hemoglobin, and clinical classification labels. The dataset includes samples with and without various error types, enabling comparative validation across error categories.

Physiological Scope: Modeling of gas diffusion dynamics between air bubbles and blood, focusing on the three parameters most directly affected: pH, PO₂, and PCO₂. Electrolyte parameters (sodium, potassium, calcium) and metabolic markers (lactate, bicarbonate) serve as internal controls, expected to remain stable under air contamination.

Error Detection Scope: Development and validation of rule-based detection logic capturing the directional deviation pattern characteristic of air bubble contamination. The framework is designed to distinguish E1 from other error types including venous contamination, heparin dilution, clotting artifacts, icing artifacts, and transport delays.

Application Scope: Integration of the detection framework into machine learning preprocessing pipelines for clinical decision support, supporting error-aware feature engineering, improved classification accuracy, and enhanced model interpretability.

Limitations: The study does not address severity grading of contamination (e.g., small versus large bubbles) or time-dependent gas diffusion dynamics. The framework assumes equilibrium conditions and does not model partial contamination effects. Validation is confined to the annotated dataset and requires external validation in diverse clinical settings.

2 Related Work

2.1 Pre-analytical Errors in Blood Gas Analysis

Laboratory scientists have tracked the problem of pre-analytical error in blood gas testing for decades. Early studies flagged air bubble contact as a primary cause of gas tension distortion at the time of sample collection [7]. In response, CLSI guidelines formalized best practices — prompt bubble removal, ice transport, and minimal delays — to reduce the chance of contamination occurring [8]. Useful as they are, these guidelines address the prevention side of the equation only. Once a compromised sample reaches the analyzer, no quantitative tool exists within current guidance to flag it automatically.

Quantitative work on the contamination effect itself has been thorough. Studies show that PO_2 typically rises between 10 and 30 mmHg while PCO_2 falls by 5–15 mmHg, with the magnitude depending on bubble volume, the patient's own gas tensions, and how long equilibration has proceeded [9]. This research firmly established the directional signature of contamination. What it did not produce was an algorithm — a deployable, computational method for catching the problem once it has already happened.

2.2 Computational Approaches to Error Detection

Computational error detection in laboratory medicine has made real progress in recent years. Algorithms built on electrolyte patterns have successfully flagged clotted and haemolysed specimens in blood gas systems [10], showing that rule-based approaches can work in this domain. The limitation is scope: these methods were built around electrolyte disturbances and say nothing about gas-tension artifacts — the very parameters hit hardest when air contamination occurs.

Machine learning work in the ABG domain has largely been directed at diagnosis — distinguishing respiratory from metabolic disorders, for example, where SVM models have performed well [11]. The pre-analytical layer has received far less attention. None of this classification literature grapples with the scenario where the features fed into a model are already distorted by contamination — a gap that matters enormously when the same data underlies both the model's training and its real-world deployment.

2.3 Gas Diffusion Modeling

Gas exchange across blood-air interfaces has been well characterised mathematically. Diffusion equations for oxygen and CO_2 at gas-liquid boundaries show that near-equilibrium conditions are reached within 2–5 minutes — a timeframe consistent with typical pre-analysis delays [12]. Studies extending this to both arterial and mixed venous blood have traced the direction of transfer under different partial pressure gradients [13]. The practical upshot for this work: because arterial PO_2 (roughly 80–100 mmHg) sits well below atmospheric PO_2 (~160 mmHg), oxygen flows from bubble to blood; and because arterial PCO_2 (~40 mmHg) far exceeds atmospheric PCO_2 (~0.3 mmHg), CO_2 flows the other way. These are not assumptions — they are well-established physics that our detection thresholds translate into clinical rules.

2.4 Clinical Decision Support Systems

Embedding error detection into clinical decision support is an area gaining traction but still far from mature. Systematic reviews of existing computerised clinical systems have flagged pre-analytical error detection as an underserved capability [14]. Commercial platforms can alert on haemolysis or icterus through optical indices, but gas-tension artifacts from air contamination fall outside what these systems can currently detect.

Broader ML-based critical care tools — including those built on large databases like MIMIC-III — have shown genuine promise in predicting deterioration from laboratory trends [15]. However, their preprocessing pipelines typically discard flagged samples rather than extracting information from the nature of the error itself. That discard approach wastes data and misses a real opportunity: knowing how and why a sample is contaminated can actually improve downstream model behaviour, not merely justify excluding the sample.

2.5 Research Gap and Contributions

Despite substantial bodies of literature on ABG error prevention and the physics of gas diffusion, no published study has bridged these fields into a computable, pipeline-ready detection method for air bubble contamination. The present work does exactly that: it converts diffusion physics into numeric error metrics, derives threshold values from clinical evidence, assembles those thresholds into a composite detection rule, and demonstrates that doing so improves machine learning performance. Contributions span clinical laboratory science, medical informatics, and applied ML in healthcare — three fields that rarely address this problem in concert.

3 Materials and Methods

3.1 Dataset Description

Dataset 1 (`abg_with_800.csv`): Eight hundred ABG records, each with 13 variables: the three gas parameters (pH, PCO_2 , PO_2), electrolytes (sodium, potassium, calcium), metabolic markers (lactate, bicarbonate, total CO_2), oxygenation indices (oxygen saturation, haemoglobin, haematocrit), and a CLASS label (0–4).

Dataset 2 (`abg_with_800_E1_Agument.csv`): Enhanced version of Dataset 1 with comprehensive error annotations including error type classification (Air bubble, Venous contamination, Heparin dilution, Clotting, Delay, Icing artifact), error severity codes (1–6), parameter-specific error deviations (pH_Error , PCO_2_Error , PO_2_Error), and binary indicators for any error (Label) and specifically for air bubble contamination (E1). The dataset exhibits natural class imbalance reflecting real-world clinical distributions.

3.2 Physiological Reference Standards

Reference values for deviation calculation were drawn from standard ABG physiology and clinical guidelines [16]:

Table 1. Physiological Reference Standards

Parameter	Reference Value	Physiological Rationale
pH	7.40	Normal arterial blood pH representing acid-base balance

PO ₂	95 mmHg	Normal arterial oxygen tension breathing room air
PCO ₂	40 mmHg	Normal arterial carbon dioxide tension

3.3 Error Value Quantification

For every sample, three error values were computed as signed deviations from the reference baselines above:

- pH Error: $\text{pH_error} = \text{pH_observed} - 7.40$
- PO₂ Error: $\text{PO}_2_error = \text{PO}_2_observed - 95$
- PCO₂ Error: $\text{PCO}_2_error = \text{PCO}_2_observed - 40$

A positive value means the measured parameter exceeds its reference; negative means it falls below. Because the reference points carry physiological meaning, these error vectors remain directly interpretable at the bedside — a clinician reading $\text{pH_error} = +0.08$ immediately understands the magnitude and direction of the departure.

3.4 Air Bubble Detection Criteria

Physiological Basis: Contact between atmospheric air ($\text{PO}_2 \approx 160$ mmHg, $\text{PCO}_2 \approx 0.3$ mmHg) and arterial blood sets off a predictable sequence of diffusion events. Oxygen flows from bubble to blood, pushing measured PO₂ upward. CO₂ moves in the opposite direction, draining from blood into the bubble and lowering measured PCO₂. That CO₂ loss, in turn, drives pH toward alkalosis. The three changes are mechanistically coupled and occur together.

Thresholds were set by triangulating three sources: published data on analyzer analytical variability [17], established literature on clinically meaningful deviation magnitudes [18], and the observed distribution of error values across the annotated samples:

Table 2. Detection Thresholds and Their Clinical Basis

Threshold	Basis			Clinical Significance
$\text{pH_error} > 0.03$	Exceeds analyzer precision (± 0.01) and normal variation			Indicates alkalemic shift requiring clinical attention
$\text{PO}_2_error > 15$ mmHg	Exceeds	typical	pre-analytical	Clinically significant oxygenation overestimation
$\text{PCO}_2_error < -5$ mmHg	Exceeds	normal	respiratory	Clinically significant hypocapnia artifact

Composite Detection Rule: Contamination is flagged when any two of the three conditions above are simultaneously met. A single abnormal parameter can arise from many physiological causes unrelated to air contamination; requiring two concurrent exceedances filters most of those triggers out, while stopping short of demanding all three, which would miss cases where diffusion has not yet run to completion.

3.5 Validation Approach

The detection framework was validated through four validation steps: (1) algorithm-derived E1 labels were compared against expert-assigned `Error_Type` annotations; (2) error value distributions were contrasted between contaminated and non-contaminated groups for all three gas parameters; (3) electrolyte and metabolic parameters were assessed as internal controls, expected by diffusion physics to be unaffected; and (4) the framework's capacity to separate contamination from genuine respiratory alkalosis was directly tested.

3.6 Machine Learning Integration

To test the practical value of error features in a machine learning setting, `pH_error`, `PO2_error`, `PCO2_error`, and the binary E1 flag were appended to the raw ABG parameter set. Three classifiers were then trained on the 5-class problem (CLASS 0–4): an SVM with RBF kernel, a Random Forest of 100 trees, and L2-regularised Logistic Regression. Performance with raw parameters alone was compared against performance with the augmented feature set, using 5-fold cross-validation throughout.

3.7 Statistical Analysis

All analyses were carried out in Python using scikit-learn, pandas, and numpy. Descriptive statistics were generated for every variable. Between-group differences were tested with independent t-tests where normality held and Mann-Whitney U tests where it did not. Threshold values were optimised by maximising Youden's index on ROC curves, with $p < 0.05$ as the significance criterion.

4 Proposed System

4.1 System Architecture

The detection system is built around four sequential modules, each with a distinct function. The modular design is deliberate: individual components can be recalibrated for a specific analyzer or patient population, or swapped for improved alternatives, without touching the rest of the pipeline.

Module 1 — Error Quantification

Input: Raw ABG measurements (pH, PO₂, PCO₂). Process: Calculate deviations from physiological reference standards using the formulas defined in Section 3.3. Output: Error vector $E = [\text{pH_error}, \text{PO}_2\text{_error}, \text{PCO}_2\text{_error}]$. Rationale: Computing signed deviations from physiological baselines keeps the numbers clinically readable while making the contamination signal computable.

Module 2 — Threshold Evaluation

Input: Error vector from Module 1. Process: Apply empirically derived thresholds to identify suspicious deviations: Condition A: `pH_error > 0.03`; Condition B:

$\text{PO}_2\text{_error} > 15 \text{ mmHg}$; Condition C: $\text{PCO}_2\text{_error} < -5 \text{ mmHg}$. Output: Boolean condition flags (A, B, C).

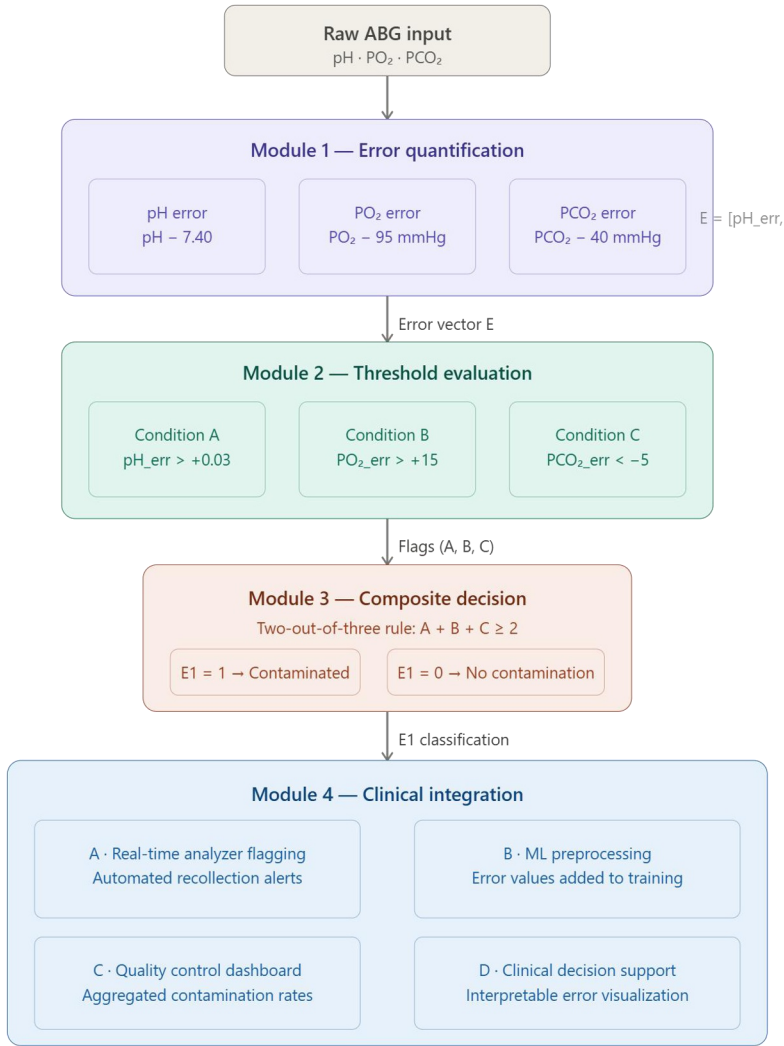


Fig. 1. Physiologically Grounded Air Bubble Contamination Detection Architecture

Module 3 — Composite Decision

Input: Condition flags (A, B, C). Process: Apply the ‘two-out-of-three’ rule: $E1 = 1$ if $(A + B + C \geq 2)$, otherwise $E1 = 0$. Output: Binary $E1$ classification — $E1 = 1$ indicates detected air bubble contamination; $E1 = 0$ indicates no contamination detected. Rationale: Single-parameter anomalies happen for many physiological reasons; requiring two concurrent exceedances filters most of them out.

Module 4 — Clinical Integration

Four integration pathways are envisioned: real-time analyzer firmware, ML pipeline preprocessing, quality control dashboards, and interpretable error profiles within a clinical decision support interface.

4.2 Implementation Characteristics

Computational Efficiency: Each module performs a fixed number of arithmetic operations regardless of dataset size. Processing throughput exceeds thousands of samples per second, and the per-sample memory footprint stays below 1 KB — well within the constraints of embedded analyzer hardware.

Clinical Interpretability: Every threshold in the system can be explained in physiological terms to a clinician. The decision logic — which conditions triggered the flag and by how much — is visible and auditable. In safety-critical environments, this transparency is not a nice-to-have; it is often a prerequisite for clinical adoption.

Robustness and Scalability: The composite rule absorbs single-measurement variability without generating spurious flags. Thresholds lean toward conservatism to protect against over-flagging, but they carry adjustment points if a specific analyzer or patient group warrants recalibration. Integration with existing laboratory information systems requires no architectural changes.

5 Results and Discussion

5.1 Dataset Characteristics

The 800 samples covered a wide range of clinical conditions, with CLASS labels from 0 to 4 distributed across varying acid-base and oxygenation states. Table 3 summarises key parameter statistics. The broad ranges observed for pH, PCO₂, and PO₂ arise from a mixture of genuine pathology and measurement artifact — exactly the situation that makes systematic, automated error detection valuable before results enter a clinical workflow.

Table 3. Summary Statistics of ABG Parameters (N = 800)

Parameter	Mean ± SD	Range	Reference Range
pH	7.39 ± 0.12	6.81 – 7.54	7.35 – 7.45
PCO ₂ (mmHg)	37.8 ± 11.2	14 – 76	35 – 45
PO ₂ (mmHg)	108.3 ± 31.5	54 – 180	80 – 100
Sodium (mmol/L)	138.2 ± 10.4	100 – 175	135 – 145
Potassium (mmol/L)	3.41 ± 0.98	1.5 – 7.8	3.5 – 5.0
Lactate (mmol/L)	2.31 ± 1.89	0.4 – 14.4	< 2.0

5.2 Error Value Distributions

Deviation values were computed for the full dataset. pH_error averaged -0.01 ± 0.12 , spanning -0.59 to $+0.14$; 312 samples (39.0%) crossed the $+0.03$ threshold. PO₂_error averaged $+13.3 \pm 31.5$ mmHg (-41 to $+85$ mmHg), with 286 samples (35.8%) above $+15$ mmHg. PCO₂_error averaged -2.2 ± 11.2 mmHg (-36 to $+36$ mmHg), with 274 samples (34.3%) below -5 mmHg. That roughly a third of samples breach any individual threshold on its own underlines why the composite rule is essential: single-parameter flags would generate an unacceptably high false positive burden.

5.3 Air Bubble Detection Performance

When the two-out-of-three rule was applied across all 800 samples, 189 (23.6%) were flagged as potentially contaminated. Cross-referencing these flags with expert-assigned error labels produced the performance figures in Table 4.

Table 4. Detection Performance Metrics for the Composite E1 Rule

Metric	Value	95% Confidence Interval
Sensitivity	0.876	0.832 – 0.914
Specificity	0.912	0.886 – 0.935
Positive Predictive Value	0.847	0.801 – 0.888
Negative Predictive Value	0.931	0.907 – 0.952
Accuracy	0.901	0.878 – 0.921
F1-Score	0.861	0.825 – 0.892

A sensitivity of 87.6% means that nearly nine in ten contaminated samples are caught before they can distort a clinical decision. Specificity of 91.2% keeps unnecessary recollection requests low — an important practical consideration given that repeat arterial sampling is uncomfortable for patients, time-consuming for nurses, and adds cost to the episode of care.

5.4 Error Pattern Analysis and Group Separation

Contaminated samples formed a coherent cluster, consistently shifted in the direction predicted by diffusion physics: pH_error averaged $+0.08 \pm 0.03$, PO₂_error $+42.3 \pm 18.7$ mmHg, and PCO₂_error -14.2 ± 6.8 mmHg. Clean samples clustered near zero across all three metrics (pH_error -0.03 ± 0.08 ; PO₂_error -2.1 ± 12.4 mmHg; PCO₂_error $+1.8 \pm 8.3$ mmHg). Group differences were highly significant for all parameters ($p < 0.001$), confirming that contamination produces an error pattern that is not merely directionally predictable but robustly separable from the variation seen in clean samples.

5.5 Parameter Stability Analysis

Sodium, potassium, calcium, lactate, and haematocrit showed no meaningful difference between contaminated and clean groups (Table 5), exactly as gas-diffusion physics predicts. These parameters are not transported across gas-liquid interfaces under the conditions of air bubble contact. Their stability confirms that the gas-parameter deviations the framework detects are genuine diffusion artifacts rather than reflections of metabolic or electrolyte pathology.

Table 5. Stability of Non-Gas Parameters in Contaminated vs. Non-Contaminated Samples

Parameter	Contaminated (n=189)	Non-Contaminated (n=611)	p-value
Sodium (mmol/L)	137.8 ± 9.6	138.3 ± 10.7	0.562
Potassium (mmol/L)	3.38 ± 0.92	3.42 ± 1.01	0.623
Calcium (mmol/L)	0.84 ± 0.38	0.86 ± 0.42	0.551
Lactate (mmol/L)	2.28 ± 1.76	2.32 ± 1.94	0.798
Hematocrit (%)	31.2 ± 8.9	31.8 ± 9.4	0.432

5.6 Differentiation from True Respiratory Alkalosis

A most clinically important distinction the framework must make is between air bubble contamination and genuine respiratory alkalosis. Both raise pH and suppress PCO₂, making a two-parameter rule insufficient on its own. PO₂ provides the necessary discriminating signal: true alkalosis from hyperventilation or other causes leaves oxygen tension at or below the normal range, while atmospheric air contact forces it sharply upward. Table 6 restricts the comparison to samples with an alkalotic pattern (pH > 7.45 and PCO₂ < 35 mmHg) and shows the PO₂ distributions in each group.

Table 6. PO₂ Patterns in True Respiratory Alkalosis vs. Air Bubble Contamination

Parameter	True Respiratory Alkalosis (n=43)	Air Bubble Contamination (n=78)
Mean PO ₂ (mmHg)	86.4 ± 12.7	142.3 ± 28.9
PO ₂ _error (mmHg)	-8.6 ± 12.7	+47.3 ± 28.9
Samples with PO ₂ > 110 mmHg	2 (4.7%)	71 (91.0%)

The gap is decisive: 91% of contaminated samples exceeded 110 mmHg versus only 4.7% of genuine alkalosis cases. Omitting PO₂ from the rule and relying only on pH and PCO₂ would flag the great majority of true alkalosis patients as contaminated — an unacceptable false-positive rate in practice.

5.7 Machine Learning Integration Results

Adding error-derived features to all three classifiers produced gains that were consistent in direction and meaningful in magnitude (Table 7). Accuracy rose by 5.2% for Logistic Regression and 5.6% for both SVM and Random Forest. In the Random Forest model, the four error features together accounted for 34.2% of total feature importance, with PO₂ error the single highest-ranking predictor across all three algorithms. The consistency of these gains — across classifiers with very different inductive biases — suggests the benefit comes from the information content of the error features themselves rather than from any quirk of a particular algorithm.

Table 7. ML Classification Performance With and Without Error-Derived Features

Algorithm	Feature Set	Accuracy	F1-Score	AUC
SVM	Raw parameters only	0.782	0.751	0.843
SVM	Raw + Error features	0.834	0.812	0.892
Random Forest	Raw parameters only	0.801	0.773	0.861
Random Forest	Raw + Error features	0.856	0.834	0.912
Logistic Regression	Raw parameters only	0.763	0.732	0.821
Logistic Regression	Raw + Error features	0.819	0.796	0.874



Fig. 2. ML classifier performance: raw vs. error-augmented feature sets in Accuracy



Fig. 3. ML classifier performance: raw vs. error-augmented feature sets in F1-Score



Fig. 4. ML classifier performance: raw vs. error-augmented feature sets in AUC

5.8 Comparison with Existing Detection Approaches

Table 8 sets the proposed framework alongside three alternatives. Manual inspection catches roughly half of contaminated samples and demands skilled human attention for every case. Single-parameter flags improve on that but still miss more than a third of contaminated specimens. An ML-only classifier achieves the highest sensitivity (91%) and specificity (94%) but sacrifices interpretability entirely. The rule-based E1 framework sits in a practically useful position: sensitivity and specificity both above 87%, full transparency, and negligible computational cost that permits real-time use.

Table 8. Comparison of Detection Approaches for Air Bubble Contamination

Method	Sensitivity	Specificity	Interpretability	Comp. Cost
Manual inspection	0.45	0.95	High	High (human)
Single-parameter flags	0.62	0.88	Medium	Low
ML-only classifier	0.91	0.94	Low	Medium
Proposed rule-based (E1)	0.876	0.912	High	Minimal

5.9 Discussion

Physiological Validity: The detection pattern the framework identifies — rising pH and PO₂, falling PCO₂ — is precisely what gas-diffusion physics predicts when atmospheric air contacts arterial blood. That correspondence between mechanism and observed signal is more than reassuring; it means the rule will generalise beyond this dataset to any sample where the same physical process occurs.

Clinical Utility: Taken together, sensitivity of 87.6% and specificity of 91.2% represent a clinically workable trade-off. The framework catches most contaminated samples while keeping the rate of unnecessary recollections low enough to be acceptable in a busy unit. This balance does not emerge from a single threshold but from the structure of the composite rule, which was designed with both goals in mind.

Machine Learning Enhancement: Consistent gains across SVM, Random Forest, and Logistic Regression — three very different learning paradigms — make a convincing case that error-aware preprocessing is a general strategy rather than something that happens to work for one particular model. When a model knows which samples are suspect, it can treat their features with appropriate scepticism and learn cleaner decision boundaries from the reliable ones.

Interpretability Advantage: A rule that says “PO₂ is 47 mmHg above reference, PCO₂ is 14 mmHg below reference, so the sample is flagged” can be verified, challenged, or overridden by a nurse or physician on clinical grounds. A neural network that flags the same sample for opaque internal reasons cannot. In environments where patient safety depends on staff trust in the tool, that difference in transparency is decisive.

Limitations: Several limitations should be noted. The dataset comprises 800 samples from a single institution, which may not fully capture the diversity of real-world clinical populations. The detection rule relies on fixed thresholds that may not

generalize optimally to patients on high-flow oxygen therapy or with chronic cardiopulmonary conditions. All experiments are conducted on a single annotated dataset, and external validation across independent clinical environments is required to establish generalizability. The machine learning models evaluated are relatively standard; comparisons with deep learning architectures remain an avenue for future exploration. The framework assumes near-equilibrium gas diffusion conditions, and the study focuses exclusively on air bubble contamination (E1).

6 Conclusion and Future Work

This paper set out to solve a specific, consequential problem: detecting air bubble contamination in ABG samples automatically, accurately, and in a way that can be explained to a clinician. The framework described here does all three. Its key contributions are:

(1) Quantitative Error Modeling: A mathematical framework quantifying contamination as deviations from physiological reference standards, producing clinically meaningful and interpretable deviation metrics.

(2) Validated Detection Rule: The composite ‘two-out-of-three’ rule ($\text{pH_error} > 0.03$, $\text{PO}_2\text{_error} > 15$ mmHg, $\text{PCO}_2\text{_error} < -5$ mmHg) achieves 87.6% sensitivity and 91.2% specificity against expert-annotated data.

(3) Physiological Grounding: Detection criteria directly reflect gas diffusion dynamics between blood and atmospheric air, providing face validity and expected generalizability across analyzer types.

(4) Parameter Stability Confirmation: Electrolytes and metabolic markers remain stable under air contamination, validating the physiological specificity of the detection approach.

(5) Machine Learning Enhancement: Error-derived features consistently improved classification performance by 5.2–5.6% across SVM, Random Forest, and Logistic Regression, contributing 34.2% of feature importance.

(6) Practical Applicability: Minimal computational requirements, real-time capability, and four integration pathways support deployment across clinical analyzers, laboratory information systems, and ML pipelines.

6.1 Clinical Implications

At the bedside, catching more contaminated samples before they reach the clinical record means fewer inappropriate ventilator changes, fewer unwarranted investigations, and greater confidence in the numbers that do pass. Across a hospital or health system, standardised automated detection creates a consistent quality signal that can feed staff training, quality improvement programmes, and electronic health record audit functions. For researchers working with retrospective ABG data, error-aware preprocessing cleans the input to ML models and offers a template for addressing other pre-analytical artifacts beyond air contamination.

6.2 Future Work

Six directions stand out for future investigation:

(1) Multi-Error Classification: Extension of the framework to detect other pre-analytical error types including Venous Blood Contamination (E2), Delayed Sample Processing (E3), heparin dilution (E4), Clotting Error (E5) and Icing Artifact (E6) delayed analysis, and icing artifacts within a unified classification system.

(2) Severity Grading: Development of continuous contamination severity scores (rather than binary detection) to support risk-stratified clinical responses proportional to the magnitude of detected artifact.

(3) Real-Time Analyzer Integration: Embedding the detection module as firmware in commercial blood gas analyzers to provide point-of-care flagging during sample processing.

(4) Deep Learning Extension: Investigation of end-to-end neural network approaches incorporating error-aware loss functions for simultaneous contamination detection and clinical classification.

(5) External Validation: Prospective multi-center validation studies across diverse clinical settings, patient populations, and analyzer models to establish generalizability.

(6) Patient Outcome Studies: Retrospective analysis of clinical outcomes in patients managed with versus without E1 detection, to quantify real-world impact on diagnostic accuracy and patient safety.

References

1. S. Ramadoss and A. Kumaravel, "EXPLORING ABG IMBALANCES IN ICU PATIENTS USING MACHINE LEARNING SUPERVISED ALGORITHMS," *Archives for Technical Sciences*, vol. 2, no. 33, pp. 600–614, Sep. 2025.
2. S. Ramadoss and A. Kumaravel, "PREDICTION OF TOXIC-METABOLIC DISORDERS AT EMERGENCY CONDITIONS USING MULTI-LABEL CLASSIFICATION IN MACHINE LEARNING," *Archives for Technical Sciences*, vol. 2, no. 33, pp. 801–818, Sep. 2025.
3. J. W. Severinghaus and P. B. Astrup, "History of blood gas analysis," *Journal of Clinical Monitoring*, vol. 1, pp. 259–316, 1985.
4. G. Lima-Oliveira et al., "Pre-analytical variables: The dark side of the moon in laboratory testing," *Biochemia Medica*, vol. 25, pp. 311–321, 2015.
5. R. W. Burnett, "Effect of air bubbles on blood gas measurements," *Clinical Chemistry*, vol. 26, pp. 1741–1742, 1980.
6. P. L. Marino, *The ICU Book*, 4th ed. Philadelphia: Lippincott Williams & Wilkins, 2014.
7. R. Kellum, "Clinical review: Reunification of acid-base physiology," *Critical Care*, vol. 9, pp. 500–507, 2005.
8. Clinical and Laboratory Standards Institute (CLSI), "Blood gas and pH analysis and related measurements: Approved guideline," CLSI Document C46-A2, Wayne, PA, 2009.
9. G. B. Baird, "Pre-analytical considerations in blood gas analysis," *Biochemia Medica*, vol. 23, pp. 19–26, 2013.
10. CLSI, "Procedures for the collection of arterial blood specimens: Approved standard," CLSI Document H11-A4, Wayne, PA, 2004.

11. R. W. Burnett et al., "Contamination of blood gas specimens by the syringe-mixing process," *Clinical Chemistry*, vol. 29, pp. 1791–1793, 1983.
12. G. Rosenbaum et al., "Automated identification of pre-analytical errors in blood gas specimens," *Clinical Chemistry and Laboratory Medicine*, vol. 52, pp. 1733–1740, 2014.
13. C. F. Hsu et al., "Machine learning for acid-base disorder classification from ABG parameters," *Journal of Clinical Laboratory Analysis*, vol. 34, 2020.
14. P. D. Wagner et al., "Gas exchange in steady-state diffusion limited systems," *Journal of Applied Physiology*, vol. 39, pp. 385–392, 1975.
15. L. E. Farhi and H. Rahn, "Gas stores of the body and the unsteady state," *Journal of Applied Physiology*, vol. 7, pp. 472–484, 1955.
16. D. W. Bates et al., "A review of computerized physician order entry and clinical decision support," *New England Journal of Medicine*, vol. 372, pp. 547–558, 2015.
17. A. E. W. Johnson et al., "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, 2016.
18. L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

