ISSN 1840-4855
e-ISSN 2233-0046

Original Scientific Article
http://dx.doi.org/10.70102/afts.2025.1833.600

EXPLORING ABG IMBALANCES IN ICU PATIENTS USING MACHINE LEARNING SUPERVISED ALGORITHMS

S. Ramadoss[1], Dr.A. Kumaravel[2]

[1]Research Scholar, School of Computing, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India. e-mail: ramadoss90@hotmail.com, orcid: https://orcid.org/0009-0004-2688-2803
[2]Supervisor, School of Computing, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India. e-mail: drkumaravel@gmail.com, orcid: https://orcid.org/0000-0002-1278-2308

Received: August 03, 2025; Revised: September 23, 2025; Accepted: November 05, 2025; Published: December 20, 2025

SUMMARY

Arterial Blood Gas (ABG) analysis is an important diagnostic tool in intensive care unit (ICU) settings that provides valuable information about the patient's respiratory and metabolic status. However, in the absence of predictive information, when testing is over-utilized or not planned, it can cause discomfort to the patient, costs to the healthcare system, and overtax already burdened resources. This work develops a predictive model that employs machine learning to classify acid-base imbalances and guides testing, which will advance diagnosis and efficiency in practice. The primary data source for model development was a dataset that included ABG profiles of ICU patients along with parameters of pH, $PaCO_2$, HCO_3^-, PaO_2, lactate, and clinical indications of hemodynamic stability, respiratory support, and therapeutic interventions. Data pre-processing included: normalization, missing value imputation, and feature scaling, and the Synthetic Minority Over-sampling Technique (SMOTE) was used to create better class balance to improve generalization. The predictive utility used a family of Support Vector Machine (SVM) classifiers with linear, polynomial, and radial basis function (RBF) kernels, which were tuned using a grid search and 10-fold cross-validation. The implementation framework was created in Python 3.11 using Scikit-learn, NumPy, and Panda's libraries. The optimized SVM classifier achieved a maximum accuracy of 93.02%, F-measure of 92.8%, precision of 93%, and an area under the ROC curve (AUC) of 0.97, for test data. The incorporation of SMOTE resulted in better class balance. This is the first application of its kind, exploring machine learning algorithms to achieve such high-performance metrics in the analysis of clinical ABG data obtained in the ICU, supporting and enhancing healthcare diagnostics.

Key words: *arterial blood gas, respiratory imbalance prediction, machine learning, supervised classification, support vector machine*.

INTRODUCTION

In the management of patients with acute and critical illness. ABG testing provides accurate, real time information on pH, partial pressure of oxygen (PaO_2), partial pressure of carbon dioxide ($PaCO_2$), and bicarbonate (HCO_3^-) levels, which are used to evaluate respiratory function, metabolic status, and acid-base balance of critically ill patients [16].

In critically ill patients in an intensive care unit (ICU) or general hospital setting, ABG testing is frequently done to direct management decisions for patients, changing ventilatory parameters, prescribing medications, or changing fluid therapy regimens accordingly. However, frequent ABG testing can lead to increased patient discomfort, risk of hospital-acquired anaemia, and increased costs because of increased laboratory use [32]. Therefore, determining the best frequency of ABG testing is crucial for the balance between assessment of adequate monitoring and patient discomfort and decreased resource use. ABG interpretation is an important, timely diagnostic tool in an intensive care unit (ICU) for providing necessary information regarding a patient's respiratory status/metabolic status, acid-base balance, and oxygenation [27] [34] [35] [36].

ABGs are regularly applied in the appropriate management of critically ill patients, particularly those requiring mechanical ventilation, those in respiratory failure or derivations from normal metabolic status. Despite it being an important, necessary quality of care when managing critically ill patients, the frequency and timing of performing ABGs in the clinical setting have been contentious in critical care settings. Furthermore, excessive testing contributes to higher healthcare costs, inefficient use of laboratory resources, and an increase in workload for ICU staff [8] [25] [33]. Despite these concerns, there is often no clear protocol guiding the frequency of ABG testing, and decisions tend to be based on habit, subjective clinical judgment, or institutional practice rather than evidence-based guidelines.

The contents of the paper are organized as the sequence of related works in sections 2, material and methods in section 3, data description in section 4, proposed work in section 5, results and discussion in section 6 and finally the conclusions with future remarks in section 7, augmented with a list of references.

RELATED WORKs

The main issues in this theme of blood tests at the ICU deal with not only patients' further health control but also clinicians making optimal decisions [22].

Minimizing the cost and discomfort for the patients is done by Nadkarni et al [1]. It helps us to identify the risk of complications such as infection or hematoma. Marik et al [3] show the decision-support models employed in critical care to improve resource management, clinical workflow efficiency, and patient outcomes. The models employ clinical data to facilitate medical personnel in assessing the appropriateness of test frequency and the need for diagnostic testing such as ABG. Langley, Wong. [5] allege that ABG tests might help to optimize the test frequency and timing, consequently avoiding any unnecessary testing. Kallstrom et al [7] discuss using ABG tests for various clinical conditions in critical contexts. Blum et al., [9] suggest that eliminating unnecessary laboratory tests could be a simple way to reduce costs without having any adverse effects on patient safety.

According to Capovilla et al [11], studies of technologies to continuously monitor or test blood gases with less invasiveness may influence how frequently testing via arterial blood gas sampling will be required. Their research found that a need for a testing approach, which was more deliberately considered, would provide the same clinical relevance while decreasing the frequency of unnecessary testing cited by Benjamin Cunanan et al., [13]. In addition, Delvaux et al., [14] claimed that clinical decision support systems (CDSS) could increase efficiency by recommending an ABG test only if clinically indicated. Thus, a CDSS would allow testing to decrease while maintaining patient care. As reported by Stanski et al., [15], predictive models utilized in the ICU would allow clinicians to predict clinical deterioration of patients, leading to diagnostic testing efficiencies, including ABG tests. These models utilize predictive algorithms incorporating data such as, for example, pH and oxygen saturation in conjunction with one another, and they would help identify patients who were likely to require ABG evaluation, in which they would assist in reducing potential over-testing by Abd et al., [10].

Kallstrom et al., [7]. Explained in their guidelines that ABG tests should be performed based on clinical need and recommended based on non-invasive methods, when possible, to reduce the frequency of invasive ABG testing. It highlighted that algorithm-driven approaches to ABG testing, which focus on clinical need, could be more cost-effective by Wilinska & Hovorka, [17]. Castro et al., et al [18] found that these technologies could help reduce the frequency of invasive ABG tests, especially in stable

patients. Verma & Kapoor, 2021 [6] found that machine learning (ML) models could significantly reduce unnecessary tests, including ABGs, while enhancing the detection of clinical deterioration.

Kumaravel et al.,[19] applied machine learning algorithms with cost-sensitive classifiers to train and test the IVF dataset, observing their influence on the resulting loss. Kajanan et al. [20] applied supervised machine learning approaches to ABG in emergency care units (ECU) and intensive care units (ICU). Doctors and nurses often face difficulties identifying the type of respiratory failure using ABG test results.

METRIAL AND METHOD WORK

In this section, we provide the explanations for the terminology used in the context of relating the ABG data to the prediction methods. Firstly, we skim through the performance of the matrix of the ML methods, the next step involves utilizing SVM with different kernel types to better understand how the proposed model can be optimized [30]. Thirdly, describing the data and analyzing its distribution is essential for understanding the dataset's underlying structure [4]. Finally, performing an analysis of the data structure, particularly (ABG: Attribute-Based Grouping), is crucial for identifying patterns or groupings within the dataset that may not be immediately obvious. Table 1 contains three statistical measures: mean, standard deviation, and variance.

Table 1. Attribute description with statistical analysis

Attribute Name	Description	Mean	Standard Deviation
S.No	Serial No	NA	NA
NAME	The name of the patients	NA	NA
SEX	Gender of the patient (Male/Female)	NA	NA
pH	Measures the acidity or alkalinity of blood. Normal range is 7.35 - 7.45	0.772	0.129
$spCO_2$	Indicates respiratory function. Normal range is 35 - 45 mmHg	0.236	0.135
Na (Sodium)	Measures serum sodium levels. Normal range is 135 - 145 mmol/L	0.793	0.103
K (Potassium)	Measures serum potassium levels. Normal range is 3.5 - 5.0 mmol/L	0.135	0.084
Ca (Calcium)	Measures serum calcium levels. Normal range is 8.5 - 10.5 mg/dL	0.013	0.089
Lac (Lactate)	Indicates tissue oxygenation and metabolic status. Normal range is 0.5 - 2.2 mmol/L	0.144	0.133
HCT (Hematocrit)	Measures the proportion of red blood cells. Normal range for men is 40-54%, and for women is 36-48%	0.503	0.142
HCO_3	Reflects metabolic function. Normal range is 22 - 26 mmol/L	0.339	0.115
TCO_2	Includes bicarbonate and dissolved CO_2. Normal range is 23 - 30 mmol/L	0.407	0.166
SO_2C	Measures the percentage of haemoglobin saturated with oxygen. Normal range is 95 - 100%.		
	0.906	0.154	0.023716
THBC	Reflects the total amount of haemoglobin in blood	0.514	0.152
Label	Different class	NA	NA

In Table 1 gives various physical attributes measured in patients, together with their descriptions, mean values, and standard deviations. pH levels specify blood acidity, with a mean of 0.772 and a standard deviation of 0.129, though these values appear normalized. $spCO_2$, reflecting respiratory function, has a

lesser mean of 0.236. Electrolytes such as sodium (Na) and potassium (K) are more important for cellular function, with means of 0.793 and 0.135, respectively. Calcium (Ca), crucial for bones and cellular activity, has a lower mean of 0.013. Lactate (Lac), shows metabolic status, averaging 0.144. Hematocrit (HCT), measuring red blood cell ratio, has a mean of 0.503. Bicarbonate (HCO_3) and total CO_2 (TCO_2), reflecting metabolic and respiratory balance, show means of 0.339 and 0.407. SO_2C, measuring oxygen saturation, averages 0.906, while THBC, representing total haemoglobin, has a mean of 0.514. Standard deviations highlight variability in patient data.

Clinical Parameters

There are three types of chemical experiments as clinical parameters as follows.

I. Respiratory Status: Monitored through parameters such as respiratory rate, oxygen saturation (SpO_2), and $PaCO_2$. II. Metabolic Status: Evaluated using bicarbonate levels, base excess, and pH. III. Hemodynamic Stability: Assessed using blood pressure, heart rate, and lactate levels.

Patient-Specific Factors

I. Underlying Conditions: Chronic respiratory diseases, renal impairment, and metabolic disorders. II. Recent Changes in Therapy: Introduction or adjustment of mechanical ventilation, vasoactive drugs, or renal replacement therapy. III. Clinical Deterioration: Signs of clinical instability, such as altered mental status or sudden changes in vital signs.

Patterns of collecting the data

I. Standard schedule: Establish a baseline frequency for ABG testing (e.g., every 4-6 hours) for stable patients. II. Dynamic Adjustments: Increase testing frequency if significant changes in clinical parameters are detected, and decrease frequency if stability is confirmed. III. Procedure for monitoring: Procedures and processes are in place for continuous monitoring of data and output recommendations for ABG testing intervals.

SVM (Support Vector Machine) Classifier Algorithm

SVM is a powerful supervised machine learning algorithm primarily used for classification tasks, though it can also be applied to regression [2] [12]. It is based on the concept of finding a hyperplane that best separates the data points of different classes in the feature space [29]. The goal of SVM is to maximize the margin between data points of different classes [24]. Here's a step-by-step breakdown of how the algorithm works:

Steps in SVM Algorithm:

1. Input Data Preparation:

The algorithm takes a labelled dataset with n features and a corresponding target variable (class labels) for training [21]. Each data point can be represented as a vector $Xi = \{Xi1, Xi2, \ldots X_{in}\}$ where n=12, with class labels $Y_i \in$ {AC-RC, AC-RENC, AL-RENC, AL-RC, NORMAL}.

2. Choose a Kernel Function:

If the data is linearly separable, the linear kernel can be used. For non-linear data, SVM uses kernel functions (e.g., polynomial, RBF) to transform the original feature space into a higher-dimensional space where the data becomes linearly separable.

Popular kernels include:

- Linear Kernel: Best for linearly separable data.

- Polynomial Kernel: For more complex boundaries.
- Radial Basis Function (RBF) Kernel: A common choice for non-linear problems.

3. Find the Optimal Hyperplane.

The algorithm searches for the hyperplane that best divides the data into two classes. Mathematically, this hyperplane can be expressed as:

$$\omega^T x + b = 0 \quad (1)$$

where ω is the weight vector and b is the bias term.

The key is to maximize the margin among the five classes, which is the distance between the hyperplane and the nearest data points (called support vectors) [31]. The margin is calculated as $\frac{\mathbf{2}}{||\boldsymbol{\omega}||}$ for each pair of targets and aggregating these calculations

4. Maximizing the Margin:

The goal of SVM is to maximize the margin between the two classes while minimizing classification errors. This is formulated as an optimization problem:

$$\min \frac{1}{2}||\omega||^2 \quad subeject\ to\ y_{i(\omega^T x_i + b=0)} \geq 1 \quad (2)$$

The support vectors are the data points that lie closest to the hyperplane and influence its position.

5. Handling Non-Linearly Separable Data (Soft Margin and C Parameter):

If the data is not perfectly separable, SVM allows for some misclassifications by introducing a soft margin. This introduces a penalty term controlled by a regularization parameter C. Hence, the optimization problem becomes:

$$\min \frac{1}{2}||\omega||^2 + C \sum_i \varepsilon_i \quad subeject\ to\ y_{i(\omega^T x_i + b=0)} \geq 1 - \varepsilon_i \quad where\ \ \varepsilon_i \geq 0 \quad (3)$$

represents $\varepsilon_i$the degree of misclassification.

6. Kernel Trick for Non-Linear Data:

If the data is not linearly separable, the kernel trick is applied to map the data into a higher-dimensional space where a linear hyperplane can be found. SVM avoids explicitly computing this transformation by using a kernel to compute the dot product in the transformed space.

7. Solving the Optimization Problem:

SVM uses quadratic programming (QP) or specialized optimization techniques like the SMO (Sequential Minimal Optimization) algorithm to find the optimal hyperplane. Once the optimization problem is solved, we obtain the support vectors, which define the optimal hyperplane.

8.Prediction:

After training, the model uses the learned hyperplane to classify new data points. For a given test point x_{test} the decision function is:

$$F(x_{\{test\}}) = \omega^T x_{\{test\}} + b = 0 \quad (4)$$

If $F(x_{test}) \geq 0$, the test point is classified as +1, otherwise, it is classified as −1.

DATA DESCRIPTION

Data used in this article was collected from the lab Biochemistry testing lab at Sree Baalaji Medical Hospital Chomped Chennai, Tamil Nadu, India. The dataset used in this study includes ABG measurements from 202 ICU patients, classified into five categories:

Table 2. Five classification

Five Categories(classes)	Description
ACIDOSIS-RC	Respiratory Compensated Acidosis is characterized by low pH and elevated pCO_2
ACIDOSIS-RENAL-C	Renal Compensated Acidosis is characterized by low pH and low HCO_3.
ALKALOSIS-RC	Respiratory Compensated Alkalosis is characterized by elevated pH and low pCO_2.s
ALKALOSIS-RENAL-C	Renal Compensated Alkalosis is characterized by elevated pH and high HCO_3.
NORMAL	Absence of the above conflicts.

These five classifications in the data in Figure 1 show the nature of acid-base imbalances in patients, whether they stem from respiratory or metabolic origins, and whether the body is compensating for the disturbance as described in Table 2.

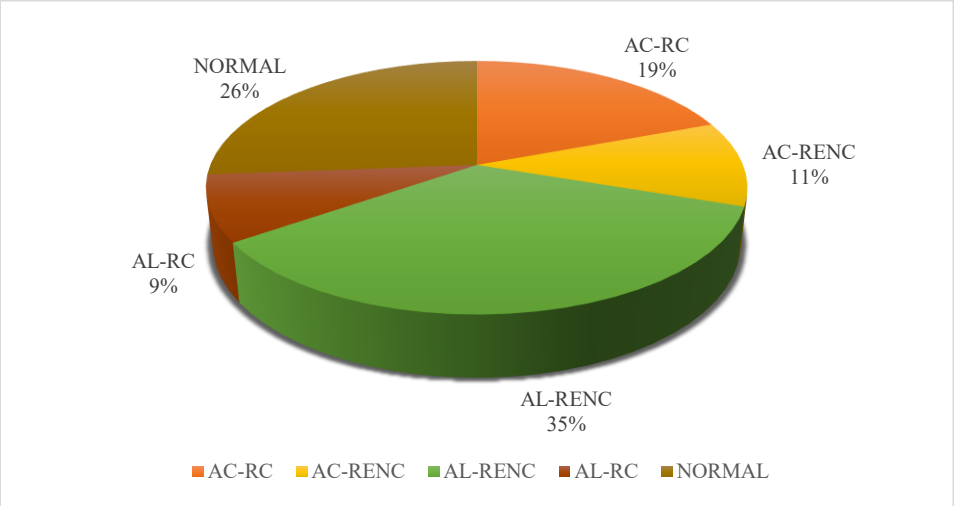

Figure 1. Data distribution

1. AL-RENC (35%): This is the largest portion of the dataset, accounting for 35% of the total data. It indicates that this category, "AL-RENC," is the most common or frequent in the dataset.
2. NORMAL (26%): The second largest category is "NORMAL," making up 26% of the data. This could represent a control or baseline group.
3. AC-RC (19%): The "AC-RC" category represents 19% of the data, which is a moderate portion compared to the others.
4. AC-RENC (11%):"AC-RENC" contributes 11% to the dataset, making it a smaller but still significant portion of the total data.
5. AL-RC (9%): The smallest category in this distribution is “AL-RC," which accounts for 9% of the dataset.

Figure 1 shows a fairly balanced dataset, with some skew towards AL-RENC and NORMAL, which together make up more than half of the dataset (61%). The other categories are lower in ratio, with AL-RC being the smallest represented. Understanding this distribution is important for knowing whether the

dataset is balanced or imbalanced, which can affect model performance, particularly in classification tasks.

PROPOSED FRAMEWORK

We provide a flowchart-based structure for our proposed system for predicting models from the Arterial Blood Gas (ABG) dataset [23]. This system aims to improve clinical decision-making using a Support Vector Machine (SVM) classifier method and multiple kernel methods, along with fine-tuned numerical parameters. The framework begins with data collection, in which pre-cleaned ICU ward samples (202 records) are collected and organized for analysis. The next step is preprocessing and feature selection, where we ensure that only the most relevant points in the dataset inform the classification task. Five target variables are determined that will inform the prediction [26]. A significant feature of the system is the iterative training step that applies each kernel method (linear kernel, radial basis function (RBF) kernel, polynomial kernel, and PUK kernel) to the SVM classifier. Concurrently, we will fine-tune numerical parameters such as the margin, which determines the distance of the decision boundary and the support vectors from the decision boundary and the exponent, which determines polynomial transformations. In this iterative cycle, we examined different training arrangements to find the ideal performing model. The next phase is to select the model with the best accuracy, generalization, and computational efficiency. Selecting this improved model becomes the most accurate predictor and provides clinical reasoning, as well as assisting healthcare professionals in their decision-making. The way in which we presented the proposed system, as illustrated in Figure 2, ensures we propose a systematic, data-driven method to the analysis of ABG's which has improved predictive performance and propinquity to real-world settings encountered in the delivery of healthcare.

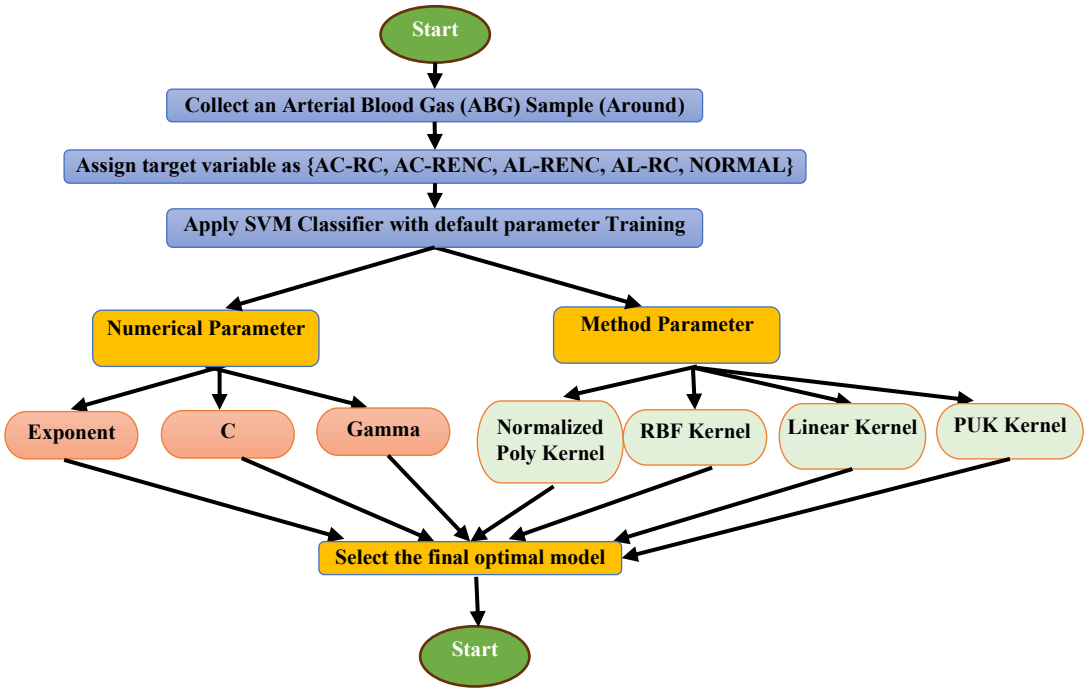

Figure 2. Proposed architecture diagram

RESULTS AND DISCUSSION

We got results for 202 collected sample inputs for our proposed model. The two different types of parameter results are provided below.

Patient-Specific Factors

Different kernel methods are used to modify the feature space to allow the data to be linearly separable:

1. Normalized Poly Kernel: A polynomial kernel function or method that is possibly normalized for better performance in particular specific cases.

2. RBF Kernel: A Radial Basis Function kernel, widely used in non-linear data classification.

3. Linear Kernel: A basic kernel used for linearly separable data.

4. Puk Kernel: This could refer to a specific kernel function, possibly used in advanced machine learning methods.

Table 3. Different kernel types with accuracy and ROC

KERNEL TYPE	ACCURACY	ROC
Linear Kernel	93.1507	0.976
Normalized Poly Kernel	84.4749	0.787
Puk kernel	76.2557	0.868
RBF Kernel	32.4201	0.23

Table 3 shows the performance of different kernel types in terms of accuracy and ROC (Receiver Operating Characteristic) for a classification task. The Linear Kernel has the highest accuracy of 93.15% and ROC of 0.976, making it the best kernel for the dataset. The Normalized Polynomial Kernel has the next accuracy of 84.47% with an ROC inferior to the Linear Kernel, with a score of 0.787 (which is still decent). The PUK Kernel (Pearson Universal Kernel) performed moderately with 76.26% accuracy and an ROC of 0.868. In contrast, the RBF Kernel performs poorly with a significantly low accuracy of 32.42% and ROC of 0.230, making it the least suitable for this classification task.

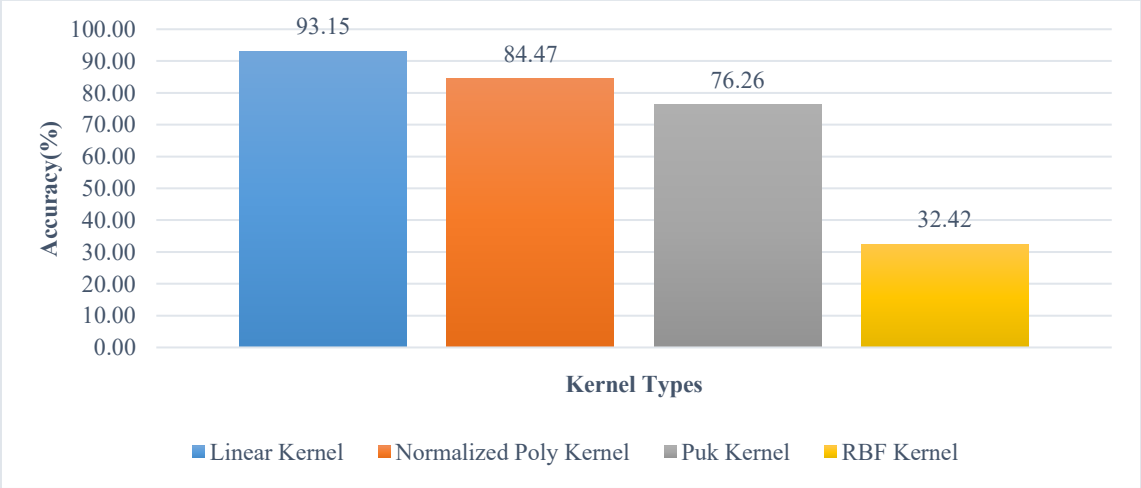

Figure 3. Accuracy vs different kernel types

Figure 3, Here is a bar chart displaying the accuracy of different kernel types. It visually highlights that the Linear Kernel has the highest accuracy (93.15%), followed by the Normalized Polynomial Kernel (84.47%) and the PUK Kernel (76.26%). The RBF Kernel performs the worst with an accuracy of 32.42%

Figure 4, Here is the bar chart visualizing the ROC (Receiver Operating Characteristic) values for the different kernel types. The Linear Kernel has the highest ROC at 0.976, followed by the PUK Kernel at 0.868, and the Normalized Polynomial Kernel at 0.787. The RBF Kernel has the lowest ROC at 0.230, indicating the poorest performance.

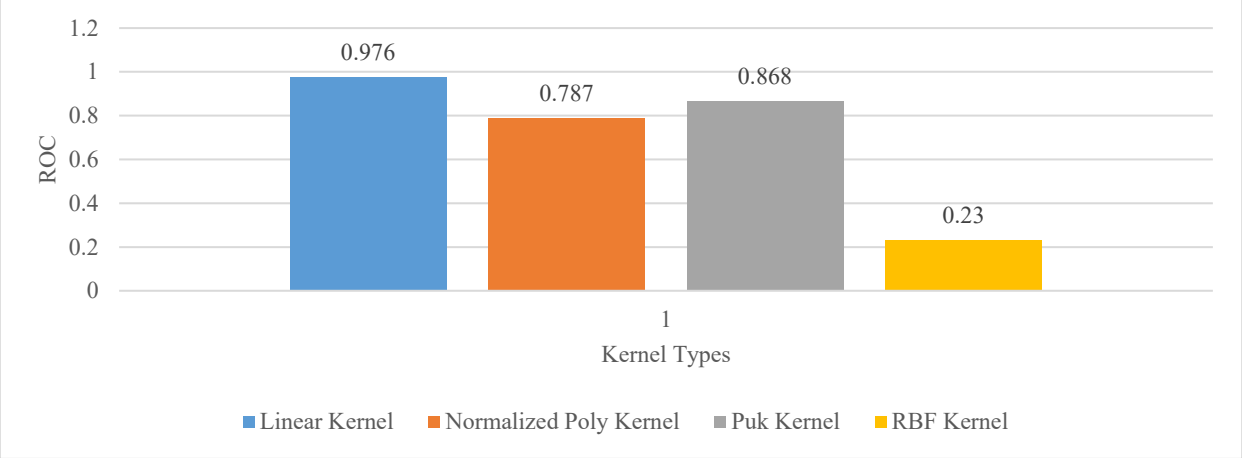

Figure 4. ROC vs different kernel types

Numerical Parameter

This chapter are explained about hyperparameters of the SVM model that need to be optimized. "C" controls the trade-off between maximizing the margin and allowing some misclassification (soft margin parameter). "Exponent" may refer to a specific parameter in a kernel method or another regularization term.

Exponent

This chapter refers to the power applied to a particular parameter or feature in a model (for example, in an SVM with a polynomial kernel, the exponent could refer to the degree of the polynomial).

Table 4 provides lists of various exponent values and their corresponding performance metrics, including True Positive Rate (TP Rate), False Positive Rate (FP Rate), Precision, and Recall.

Table 4. Different exponent(E) values with TP, FP, precision and recall

Exponent	TP Rate	FP Rate	Precision	Recall
1	0.921	0.022	0.922	0.921
1.2	0.911	0.023	0.909	0.911
1.4	0.921	0.022	0.922	0.92
1.6	0.916	0.019	0.915	0.916
1.8	0.916	0.019	0.915	0.916

Table 4 shows that the performance of the model is fairly stable across different exponent values, with the 1.0 and 1.4 exponent values providing the best balance of high TP Rate (or Recall), Precision, and low FP Rate. Increasing the exponent beyond 1.4 leads to a small decline in performance, but the results remain consistent and reliable. The consistently low false positive result across all exponent values shows that the model is highly successful at avoiding false positive classifications.

The two figures (i.e. Figure 5 and Figure 6) show the results between the exponent value (E) and key performance metrics: The True Positive Rate (TP Rate) or Recall is generally higher across all of the exponent values, ranging between 0.906 and 0.921, indicating that the model is consistently identifying most of the actual positives.

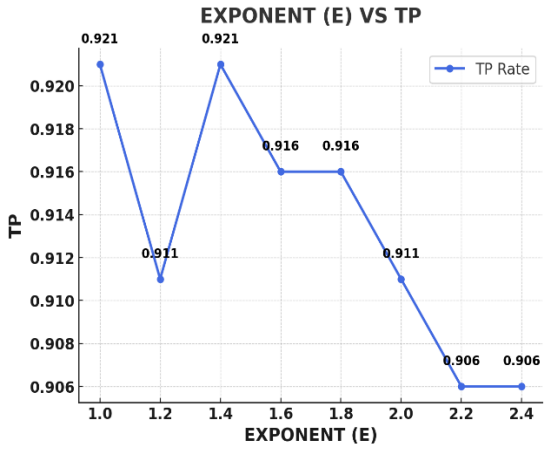

Figure 5. Exponent vs TP

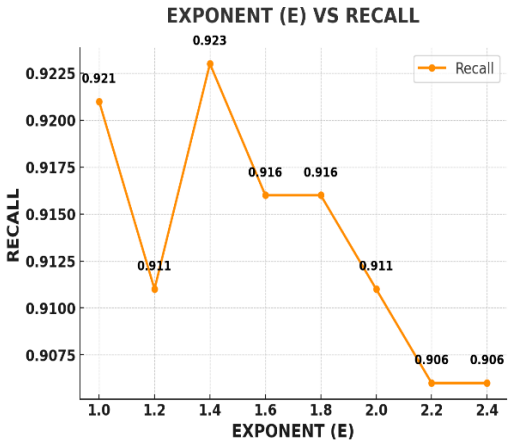

Figure 6. Exponent vs Recall

Overall, the FP Rate stays low, within the 0.019 to 0.024 range, meaning there are few false positives in Figure 7, which is good.

In Figure 8, the Precision is close to the TP rate, meaning that when the model makes a positive prediction, it is usually correct. Precision fluctuates slightly, but it stays within the 0.905 to 0.922 range overall.

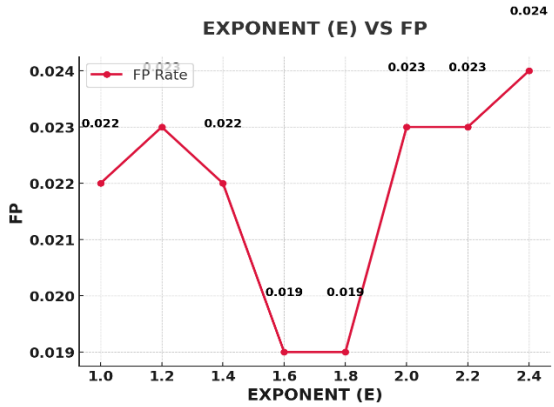

Figure 7. Exponent vs FP

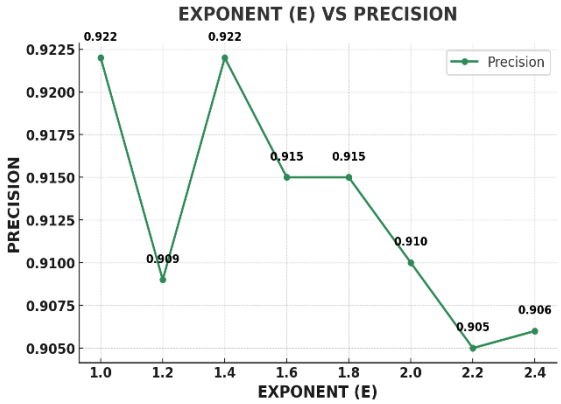

Figure 8. Exponent vs Precision

Margin(C)

Table 5 below presents key performance metrics, including Margin (C), Accuracy, Precision, F-Measure, and ROC.

Table 5. Different margin(C) values with accuracy, precision, F-Measure and ROC

Margin(C)	Accuracy	Precision	F-Measure	ROC
0.1	57.0776	0.792	0.524	0.803
0.2	86.3014	0.878	0.859	0.94
0.3	88.1279	0.886	0.877	0.946
0.4	88.5845	0.889	0.881	0.955
0.5	89.9543	0.897	0.895	0.959

Table 5 represents the classification performance for various values of the parameter "C" (likely from a support vector machine or a similar classifier). As "C" increases, the percentage of correctly classified instances improves, while the percentage of imperfectly classified instances decreases.

- At C = 0.1, the classifier correctly classified only about 57% of instances, with a relatively high 43% incorrectly classified. As C increases, the accuracy improves dramatically, attainment over 86% for C = 0.2 and about 88% for C = 0.3 and C = 0.4. The incorrectly classified percentage correspondingly decreases.
- From C = 0.5 to C = 0.8, the accuracy continues to increase somewhat, and the incorrectly classified percentage decreases little by little. For example, at C = 0.5, the classifier correctly classifies nearly 90% of instances and incorrectly classifies about 10%. By C = 0.8, the correctly classified percentage increases to over 91%, while the incorrectly classified percentage drops under 9%.
- Finally, at C = 0.9 and C = 1, the classifier achieves its highest performance, correctly classifying around 93% of instances, with only about 7% misclassified. These results suggest that larger values of C improve the model's ability to correctly classify instances, although there may be diminishing returns after C = 0.9, where the performance plateaus.

Here is the line graph. Figure 9 illustrates the relationship between the parameter 'C' and the percentage of correctly classified instances. As margin(C) increases from 0.1 to 1.0, the percentage of correctly classified instances rises sharply at first, then plateaus near the 90% mark. This suggests that higher values of 'C' generally lead to better classification performance, although the improvement levels off after a certain point (around C = 0.9).

Even though the initial accuracy is around 57% at the initial point 0.1 in the thin margin of the SVM classifier, at the very next iteration, it sharply rises to 86 % and maintains a higher band of values in the rest of the points in the margin variations.

Figure 10 represents the precision values for different "C" margins. Precision measures the proportion of true positive predictions out of all positive predictions made by the model.

- At C = 0.1, the precision is 0.792, indicating that about 79.2% of the predicted positives are true positives.
- As C increases, precision improves significantly, reaching 0.878 at C = 0.2 and 0.886 at C = 0.3.
- From C = 0.4 to C = 0.6, there is a steady rise in precision, culminating at C = 0.6 with a precision of 0.903.
- The highest precision is obtained at C = 1, where the model got 93.1% precision, meaning that most of its positive predictions are accurate.

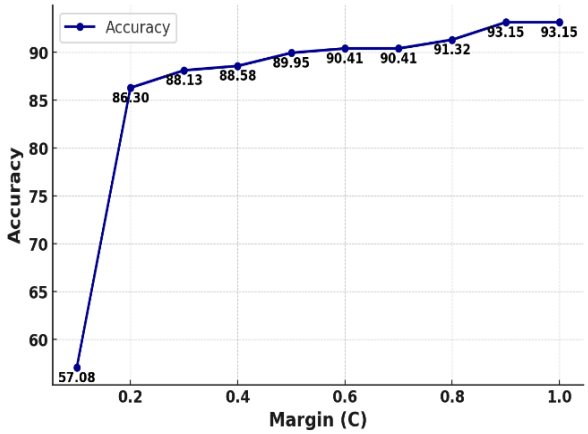

Figure 9. Margin vs Accuracy

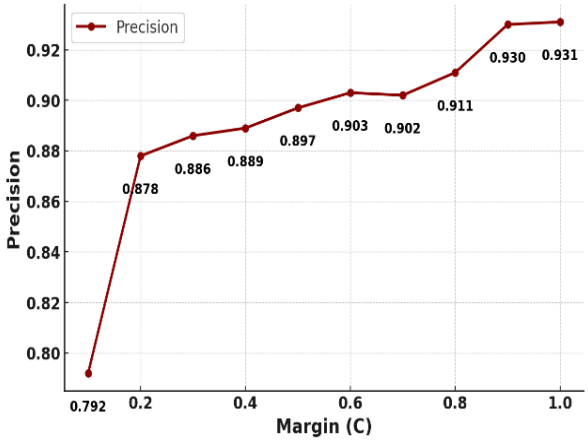

Figure 10. Margin vs Precision

Table 5 suggests that as the value of C increases, the model becomes more precise in its predictions, correctly identifying more true positives with fewer false positives.

The graph in Figure 11 shows the F-Measure (or F1 Score) values for different "C" margins. The F Measure is the consonant mean of precision and recall, providing a balance between the two, especially when the distribution between positive and negative classes is uneven.

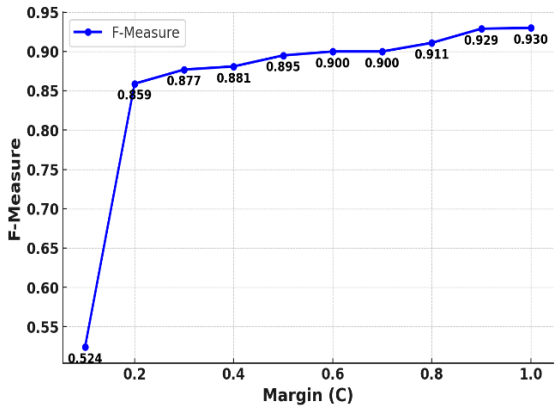

Figure 11. Margin vs F-Measure

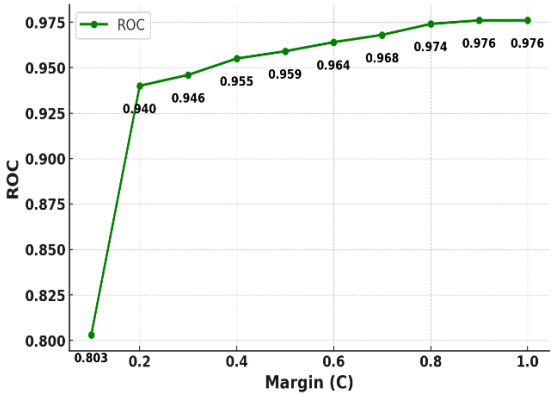

Figure 12. Margin vs ROC

- At C = 0.1, the F-Measure is 0.524, which is comparatively low. This shows that the balance between precision and recall is not optimal at this margin.
- As C increases to 0.2, the F-Measure significantly improved to 0.859, showing a much better performance of the model.
- From C = 0.3 to C = 0.5, the F-Measure continues to improve, reaching 0.895 at C = 0.5.
- From C = 0.6 onwards, the F-Measure stabilizes around 0.9, representing that the model maintains a high balance between precision and recall at these margins.
- The highest F-Measure is observed at C = 0.9 and 1, where the model reaches 0.929 and 0.93, respectively.

Table 5 suggests that as the margin C increases, the model becomes better at balancing precision and recall, with an optimal performance obtained at higher values of C.

The graph Figure 12 presents the ROC (Receiver Operating Characteristic) values for different MARGIN(C) values, which measure the model's ability to classify between classes.

- At C = 0.1, the ROC is 0.803, indicating moderate classification performance.
- As C increases to 0.2, the ROC jumps to 0.94, showing a significant improvement in the model's discriminatory power.
- The ROC continues to increase steadily from 0.3 to 0.7, reaching 0.968 at C = 0.7, which indicates a high level of accuracy in distinguishing between positive and negative classes.
- From C = 0.8 to 1, the ROC value plateaus between 0.974 and 0.976, suggesting that further increasing the margin doesn't significantly improve the model's performance beyond this point.

Overall, as C increases, the ROC values improve, showing that higher C values allow the model to better differentiate between classes, with optimal performance being reached around C = 0.9 and C = 1.

CONCLUSION

ICU patients require swift decisions by clinicians, who in turn demand support from software tools for this purpose. Such tools are based on frameworks for optimizing models to yield improved performance, supporting the ICU clinicians. In this paper, we designed the hyperparameters for the framework represented by 'Support Vector Machine' and implemented the optimal model to classify the imbalances in the blood tests to take accurate decisions governing patients' conditions, and to recommend further treatments accordingly. Hence, in this paper it has been proved the existence of an ML model framework for predicting the ABG imbalances in an emergency situation prevailing in the ICU. It should be extended to other types of testing samples to a similar type of classification algorithmic approaches. The experimental results demonstrate that a decision-making model for predicting ABG testing for deciding the normality of imbalances of gases of an ICU patient can yield up to a maximum of 0.93 accuracy, F-measure 0.93, precision is 0.931, and ROC is 0.97. This being the first of its kind, it may be improved further by increasing the data size and applying other types of learning techniques.

Author Contributions: S. Ramadoss conducted the experiment and tabulated the results, compiled the findings, and prepared the original article. Dr A. Kumaravel had investigated the feasibility of the application over ABG data and the selection of a classifier.

Funding: This work was carried out without any external funding.

Data availability: The authors can provide the data upon a justifiable request. Requests for access to the data should be directed to S. Ramadoss at ramadoss90@hotmail.com.

Code availability: Weka environment is considered, and the steps are shown in Figure 2 for deployment and implementation.

Conflict/Competing: There is a conflict/competing of interest among the authors.

Declarations:

Informed consent: There was no involvement of human participants in this research for the purpose of data collection. However, the data used were anonymized and handled with strict confidentiality. The institution's ethics committee approved the study.

Institutional review board statement: Not applicable.

Consent to participate, authors are very much interested in research, especially in applying machine learning algorithms, and they participate in the relevant projects [28].

Consent: for publication, the authors are ready to allow their works to be published.

Acknowledgements: The authors would like to thank the authorities of Bharath Institute of Higher Education and Research for their support and encouragement.

REFERENCES

[1] Nadkarni A, Besic N, Yap J, Micik S, Chapple LA, Gnanamanickam E, Reddi B, Farquharson M. Rationalising arterial blood gas sampling analysis in the intensive care unit: A before-and-after study. Australian Critical Care. 2025 Jul 1;38(4):101237. https://doi.org/10.1016/j.aucc.2025.101237

[2] Rosnelly R, Riza BS, Suparni S. Comparative analysis of support vector machine and convolutional neural network for malaria parasite classification and feature extraction. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications. 2023;14(3):194-217. https://doi.org/10.58346/JOWUA.2023.I3.015

[3] Marik PE. Arterial blood gas analysis. In: Marik PE, editor. Evidence-Based Critical Care. Cham: Springer International Publishing; 2014. p. 329-347.

[4] Salman R, Banu AA. DeepQ residue analysis of computer vision dataset using support vector machine. Journal of Internet Services and Information Security, 2023;13(1):78-84. https://doi.org/10.58346/JISIS.2023.I1.008

[5] Langley RJ, Wong HR. Predictive and prognostic tools for the management of sepsis in the intensive care unit: a work in progress. Clin Chest Med. 2013;34(3):587-601.

[6] Verma S, Kapoor H. Machine learning for predictive maintenance: a cloud computing architecture and lessons for a healthcare context. International Academic Journal of Science and Engineering. 2021;8(2):1-5. https://doi.org/10.71086/IAJSE/V8I2/IAJSE0808

[7] American Association for Respiratory Care. AARC clinical practice guideline. Sampling for arterial blood gas analysis. Respir Care. 1992;37(8):913-7.

[8] Jalali Z, Shaemi A. The impact of nurses' empowerment and decisionmaking on the care quality of patients in healthcare reform plan. Human Resource Management. 2015;2(9):33-9.

[9] Blum FE, Lund ET, Hall HA, Tachauer AD, Chedrawy EG, Zilberstein J. Reevaluation of the utilization of arterial blood gas analysis in the Intensive Care Unit: effects on patient safety and patient outcome. Journal of Critical Care. 2015 Apr 1;30(2):438-e1. https://doi.org/10.1016/j.jcrc.2014.10.025

[10] Abd IS, Farazdaq H, Khudair AN. Integrative evaluation of adeA and adeS efflux gene expression, biofilm production, and antimicrobial resistance in clinical isolates of Acinetobacter baumannii. Natural and Engineering Sciences. 2025;10(2):342-351. https://doi.org/10.28978/nesciences.1744935

[11] Capovilla J, VanCouwenberghe C, Miller WA. Noninvasive blood gas monitoring. Critical Care Nursing Quarterly. 2000 Aug 1;23(2):79-86.
[12] Chinnasamy. A blockchain and machine learning integrated hybrid system for drug supply chain management for the smart pharmaceutical industry. Clinical Journal for Medicine, Health and Pharmacy. 2024;2(2):29-40.
[13] Cunanan B, Muppa H, Orellana L, Bates S, McGain F. Blood gas sampling in the intensive care unit: A prospective before-and-after interventional study on the effect of an educational program on blood gas testing frequency. Australian Critical Care. 2024 Sep 1;37(5):755-60. https://doi.org/10.1016/j.aucc.2024.01.009
[14] Delvaux N, Piessens V, Burghgraeve TD, Mamouris P, Vaes B, Stichele RV, Cloetens H, Thomas J, Ramaekers D, Sutter AD, Aertgeerts B. Clinical decision support improves the appropriateness of laboratory test ordering in primary care without increasing diagnostic error: the ELMO cluster randomized trial. Implementation Science. 2020 Nov 4;15(1):100.
[15] Stanski NL, Wong HR. Prognostic and predictive enrichment in sepsis. Nature Reviews Nephrology. 2020 Jan;16(1):20-31.
[16] American Association for Respiratory Care (AARC). Sampling for arterial blood gas analysis. Respir Care. 1992;37:891-897.
[17] Wilinska ME, Hovorka R. Glucose control in the ICU using continuous glucose monitoring: What level of the measurement error is acceptable?. In Diabetes Technology & Therapeutics 2015 Feb 1 (Vol. 17, Pp. A3-A3). 140 Huguenot Street, 3rd Fl, New Rochelle, Ny 10801 Usa: Mary Ann Liebert, Inc.
[18] Castro D, Patil S, Zubair M, Keenaghan M. Arterial blood gas. StatPearls. 2024 Jan 8.
[19] Kumaravel A, Vijayan T. Comparing cost sensitive classifiers by the false-positive to false-negative ratio in diagnostic studies. Expert Systems with Applications. 2023 Oct 1;227:120303. https://doi.org/10.1016/j.eswa.2023.120303
[20] Kajanan S, Kumara BS, Banujan K, Prasanth S, Manitheepan K. Classify the outcome of arterial blood gas test to detect the respiratory failure using machine learning. In2022 International Conference on Decision Aid Sciences and Applications (DASA) 2022 Mar 23 (pp. 1139-1143). IEEE. https://doi.org/10.1109/DASA54658.2022.9765012
[21] Ramadoss S, Kumaravel A. Medical dataset for arterial blood gas analysis [dataset]. Kaggle. 2025.
[22] Ayala-De la Cruz S, Arenas-Hernández PE, Fernández-Herrera MF, Quiñones-Díaz RA, Llaca-Díaz JM, Díaz-Chuc EA, Robles-Espino DG, San Miguel-Garay EA. Human-in-the-Loop Performance of LLM-Assisted Arterial Blood Gas Interpretation: A Single-Center Retrospective Study. Journal of Clinical Medicine. 2025 Sep 22;14(18):6676. https://doi.org/10.3390/jcm14186676
[23] Ozdemir H, Sasmaz MI, Guven R, Avci A. Interpretation of acid–base metabolism on arterial blood gas samples via machine learning algorithms. Irish Journal of Medical Science (1971-). 2025 Feb;194(1):277-87.
[24] Guo J, Wu H, Chen X, Lin W. Adaptive SV-Borderline SMOTE-SVM algorithm for imbalanced data classification. Applied Soft Computing. 2024;150:p.110986. https://doi.org/10.1016/j.asoc.2023.110986
[25] Guido R, Ferrisi S, Lofaro D, Conforti D. An overview on the advancements of support vector machine models in healthcare applications: a review. Information. 2024 Apr 19;15(4):235. https://doi.org/10.3390/info15040235
[26] Yu H, Saffaran S, Tonelli R, Laffey JG, Esquinas AM, de Lima LM, Kawano-Dourado L, Maia IS, Cavalcanti AB, Clini E, Bates DG. Machine learning models compared with current clinical indices to predict the outcome of high flow nasal cannula therapy in acute hypoxemic respiratory failure. Critical Care. 2025 Mar 7;29(1):101.
[27] Mousavinejad SN, Lachouri R, Bahadorzadeh M, Khatami SH. Artificial intelligence for arterial blood gas interpretation. Clinica Chimica Acta. 2025 Oct 29:120691. https://doi.org/10.1016/j.cca.2025.120691
[28] Manoochehri S, et al. Evaluating SMOTE-based machine learning algorithms for clinical imbalanced datasets. Comput Biol Med. 2025;176:108265.
[29] Helleberg J, Sundelin A, Mårtensson J, Rooyackers O, Thobaben R. Beyond labels: determining the true type of blood gas samples in ICU patients through supervised machine learning. BMC Medical Informatics and Decision Making. 2025 Jul 24;25(1):275.
[30] Wibowo A, Masruriyah AF, Rahmawati S. Refining Diabetes Diagnosis Models: The Impact of SMOTE on SVM, Logistic Regression, and Naïve Bayes. Journal of Electronics, Electromedical Engineering, and Medical Informatics. 2025 Jan 11;7(1):197-207. https://doi.org/10.35882/jeeemi.v7i1.596
[31] Khyathi G, Indumathi KP, Jumana Hasin A, Lisa Flavin Jency M, Krishnaprakash G, LISA FJ. Support Vector Machines: A Literature Review on Their Application in Analyzing Mass Data for Public Health. Cureus. 2025 Jan 8;17(1). https://doi.org/10.7759/cureus.77169
[32] Zhou X, Li X, Zhang Z, Han Q, Deng H, Jiang Y, Tang C, Yang L. Support vector machine deep mining of electronic medical records to predict the prognosis of severe acute myocardial infarction. Frontiers in Physiology. 2022 Sep 29;13:991990. https://doi.org/10.3389/fphys.2022.991990

[33] Hu J, Lv S, Zhou T, Chen H, Xiao L, Huang X, Wang L, Wu P. Identification of pulmonary hypertension animal models using a new evolutionary machine learning framework based on blood routine indicators. Journal of Bionic Engineering, 2023;20(2), pp.762-781.

[34] Raoufy MR, Eftekhari P, Gharibzadeh S, Masjedi MR. Predicting arterial blood gas values from venous samples in patients with acute exacerbation chronic obstructive pulmonary disease using artificial neural network. Journal of medical systems. 2011 Aug;35(4):483-8.

[35] Li Y, Yang Y, Song P, Duan L, Ren R. An improved SMOTE algorithm for enhanced imbalanced data classification by expanding sample generation space. Scientific Reports. 2025 Jul 2;15(1):23521.

[36] Qaiser A, Manzoor S, Hashmi AH, Javed H, Zafar A, Ashraf J. Support Vector Machine Outperforms Other Machine Learning Models in Early Diagnosis of Dengue Using Routine Clinical Data. Advances in virology. 2024;2024(1):5588127. https://doi.org/10.1155/2024/5588127