Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 4, August 2026, pp. 3249~3267
ISSN: 2302-9285, DOI: 10.11591/eei.v15i4.12212

Tuberculosis severity classification from exhaled breath using an electronic nose system with Inception-1D and ResNet-1D

Dava Aulia[1], Riyanarto Sarno[1], Muhammad Rivai[2], Muhammad Amin[3,4], Alfian Nur Rosyid[3,4], Kelly Rossa Sungkono[1]
[1]Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Department of Electrical Engineering, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[3]Universitas Airlangga Hospital, Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia
[4]Department of Pulmonology and Respiratory Medicine, Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia

Article Info

Article history:

Received Dec 31, 2025
Revised May 7, 2026
Accepted Jun 19, 2026

Keywords:

Advanced deep learning
Diseases
Electronic nose system
Exhaled breath
Tuberculosis

ABSTRACT

Exhaled breath contains volatile organic compounds (VOCs) that can be analyzed for tuberculosis (TB) detection. Electronic nose systems have demonstrated promise for this application; however, accurately distinguishing between healthy individuals and TB patients with different severity levels, namely, low and high TB, remains challenging and requires advanced deep-learning methods. Unlike previous studies that focused on binary TB detection, this study proposes an electronic nose system combined with one-dimensional deep learning models, including residual network (ResNet), visual geometry group (VGG), EfficientNet, and Inception architectures, for multiclass classification of healthy individuals and TB severity levels using exhaled-breath analysis. The results show that Inception-1D and ResNet-1D achieve the best performance for healthy and TB classification, each attaining an F1-score of 94.99%. For TB severity classification, ResNet-1D outperforms other models with an F1-score of 68.50%. Meanwhile, Inception-1D yields the highest performance on the healthy and TB severity classification dataset, with an F1-score of 74.07%. Moreover, dimensionality reduction using principal component analysis (PCA) reduces the healthy and TB dataset to ten principal components and improves the F1-score to 95.64% with Inception-1D. Overall, the proposed framework successfully captures distinctive gas sensor-response patterns associated with TB presence and severity and may support clinical decision-making in resource-limited healthcare settings.

Corresponding Author:

Riyanarto Sarno
Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Keputih, Sukolilo, Surabaya, 6011 East Java, Indonesia
Email: riyanarto@its.ac.id

1. INTRODUCTION

Tuberculosis (TB) is a pulmonary infectious disease associated with mycobacterium tuberculosis (MTB) [1]–[5]. It is primarily transmitted through aerosolized droplets expelled when an infected individual coughs, sneezes, or speaks [2], [3], [5]. In pulmonary TB, the activity of MTB and the host immune response lead to lung tissue damage [1], which results in a persistent cough with sputum production (with or without blood), chest pain, and shortness of breath [2]. Despite the development of the TB vaccine in the early 20th

century, TB continues to be a major cause of infectious disease-related mortality worldwide [4]. As reported by the World Health Organization in 2024, TB affected an estimated 10.8 million people and caused nearly 1.25 million deaths, mainly in low- and middle-income countries, including Indonesia with nearly 885,000 reported cases [6]. Its high prevalence in populations with poor living conditions and limited healthcare access highlights the urgent need for more effective prevention and treatment strategies.

Standard diagnostic methods for TB rely on clinical evaluation, including physical examination, symptom observation, and assessment of contact history [7]. Nevertheless, these methods are inherently subjective and may lead to variability and delays in diagnosis. The tuberculin skin test (TST) and interferon-gamma release assay (IGRA) are utilized for TB detection by evaluating immune responses to MTB, with TST requiring a skin injection and IGRA using a blood sample [8]. The main limitation of these methods is their invasive nature, which is associated with discomfort or skin irritation at the injection site. Imaging-based methods, such as chest X-rays, are used to detect pulmonary lesions, infiltrates, or cavities suggestive of TB; however, their limitations include nonspecific findings, inability to confirm active disease, and high equipment and maintenance costs [5], [7]. Sputum microscopy is utilized to detect acid-fast bacilli in sputum samples, indicating the presence of MTB; nevertheless, its drawbacks include low sensitivity and dependence on high-quality specimens [9]. The GeneXpert MTB/RIF assay enables rapid detection of MTB deoxyribonucleic acid and rifampicin resistance using real-time polymerase chain reaction [10]. However, its challenges include high costs, sputum sample dependency, and the need for skilled personnel. Additionally, many patients struggle to produce adequate sputum, which complicates sample collection and reduces diagnostic reliability. These shortcomings hinder early TB detection and contribute to continued disease transmission and adverse clinical outcomes, particularly in resource-limited areas.

To resolve these issues, exhaled breath assessment provides a promising non-invasive solution for TB detection [11], [12]. Exhaled breath contains a mixture of volatile organic compounds (VOCs) that reflect metabolic and pathological changes associated with TB infection. These variations are influenced by lung inflammation and bacterial metabolism, which alter the overall breath profile. Gas chromatography (GC) can be used to analyze VOCs, enabling detailed characterization of their chemical composition. Hydrocarbons, alkanes [13], [14], acetone, benzene, esters [15], heptane, propane, and 2-butanone [16], [17] may result from physiological responses to TB infection, oxidative stress, and inflammation, making them potential biomarkers for TB detection. However, GC has several drawbacks, including the need for trained personnel, lengthy procedures, and limited accessibility. As an alternative, electronic nose systems can capture chemical fingerprints from exhaled breath and classify patterns indicative of TB infection by integrating commercial gas sensors with machine learning, thereby offering a low-cost, rapid, and portable solution [17], [18].

A variety of gas sensors have been employed in electronic nose systems to detect TB through exhaled breath. Quartz crystal microbalance (QMB) sensors have been used to differentiate healthy controls from TB patients, achieving an accuracy of 88% using linear discriminant analysis (LDA) [19]. However, their shortcomings include high fabrication costs, implementation challenges, and high sensitivity to humidity and temperature fluctuations. An electronic nose system using modified QMB sensors coated with metalloporphyrin films achieved 93% accuracy for distinguishing healthy individuals from TB patients using the k-nearest neighbors (KNN) model [20]. Nevertheless, its drawbacks include sensor instability, calibration requirements, and system complexity. An electronic nose system using gold nanoparticle and single-walled carbon nanotube sensors has also been developed for healthy and TB classification. By applying receiver operating characteristic (ROC)-based binary classification with Youden's Index, the system achieved an accuracy of 88% [21]. However, this sensor type remains susceptible to limited long-term stability, interference from other breath compounds, and environmental conditions. As an alternative, metal-oxide semiconductor (MOS) gas sensors provide multiple benefits, including low cost, fast response time, high sensitivity, good selectivity, and long-term durability, making them suitable for detecting and monitoring lung diseases such as lung cancer [22], chronic obstructive pulmonary disease [23], and TB [11], [18], [24].

Previous studies have applied electronic nose systems to distinguish healthy controls from TB patients using MOS sensors. A system combined with support vector machine (SVM) achieved a sensitivity of 95% and a specificity of 82% [24]. A similar system using artificial neural networks (ANNs) reported an accuracy of 94.87% for classifying healthy and TB populations [18]. Although these studies demonstrate the potential of electronic nose systems for TB detection, TB severity classification remains an important yet underexplored aspect. Beyond detecting the presence of TB, identifying severity levels is clinically important for determining treatment strategies, monitoring disease progression, and prioritizing patient management. Patients with higher TB severity generally require more intensive treatment and closer clinical monitoring. In addition, TB severity classification may assist in estimating transmission risk, as patients with higher bacterial loads can contribute more significantly to disease spread. It also supports healthcare resource allocation, particularly in low- and middle-income countries where medical resources are limited. Therefore, an accurate and non-invasive method for TB severity classification is highly valuable in clinical practice and

provides significant benefits for clinical decision-making, patient triaging, and early intervention. At present, advanced models such as residual network (ResNet), visual geometry group (VGG), EfficientNet, and Inception have been successfully applied to TB detection using lung imaging with high accuracy [25]–[27]. Therefore, this study employs these models on gas sensor data through one-dimensional (1D) convolutional networks to capture temporal and hierarchical patterns. This framework enables improved generalization, enhanced robustness to noise, and automatic feature learning. Notably, applying these models to gas sensor data for classifying healthy subjects and TB severity levels offers a novel contribution to electronic nose research, particularly for early TB detection in clinical settings.

Despite these advancements, most existing electronic nose studies for TB detection still focus primarily on binary classification between healthy individuals and TB patients, while clinically important TB severity classification remains underexplored. In addition, traditional machine-learning methods rely on manual feature extraction and may fail to capture complex temporal patterns in gas sensor data. Furthermore, there is limited investigation of advanced one-dimensional deep learning architectures for modeling gas sensor time-series data. Moreover, the impact of dimensionality reduction techniques, such as principal component analysis (PCA), on deep learning performance in this context has not been systematically investigated. As a result, existing approaches remain limited in handling multiclass scenarios with subtle inter-class differences. Therefore, there is a need for a more robust method that can automatically learn discriminative features and accurately distinguish between multiple TB severity levels. Accordingly, this study aims to investigate whether advanced 1D deep learning models, including ResNet-1D, VGG-1D, EfficientNet-1D, and Inception-1D, can effectively classify TB severity levels using data collected from an electronic nose system. Additionally, this study develops a MOS-based electronic nose system for TB severity classification, compares the performance of traditional classification models with advanced 1D deep learning architectures, evaluates the impact of PCA on classification performance and robustness, and assesses model generalization using stratified cross-validation.

This study presents a novel electronic nose framework for distinguishing healthy individuals from TB patients with two severity levels, namely low TB and high TB, using exhaled-breath analysis. Moreover, ResNet-1D, VGG-1D, EfficientNet-1D, and Inception-1D are employed to learn discriminative features from healthy and TB severity datasets. This framework contributes to the advancement of electronic nose research for TB detection. The rest of this paper is organized as follows. Section 2 describes the method, section 3 shows the results and discussion, and section 4 summarizes the study and discusses future work.

2. METHOD

2.1. Subjects and experimental design

Breath collection was performed at Airlangga University Hospital, Surabaya, East Java, Indonesia, from 160 volunteers consisting of 80 healthy individuals and 80 TB patients, all of whom were non-smokers aged 30–70 years with no history of acute or chronic illness. Breath samples were gathered using two-liter tedlar bags. The sample size was based on the availability of eligible participants during the study period at the hospital. Due to practical constraints in recruiting TB patients and collecting exhaled breath samples under controlled conditions, a convenience sampling approach was adopted. Although relatively limited, this sample size is comparable to those reported in previous electronic nose–based studies for TB detection and is considered sufficient for evaluating the proposed classification framework. Participants were required to fast and avoid tooth brushing for at least five hours prior to sampling to reduce contamination in VOC profiles and improve measurement consistency. Patients were classified into low TB and high TB severity groups based on GeneXpert MTB/RIF bacterial load categories. These classifications were further supported by clinical assessment using the Bandim TB score and confirmed by pulmonologists to ensure consistency between microbiological burden and clinical severity. This study was approved by the Institutional Review Board of the hospital (Approval No. NA-02-24225; approved on February 14, 2025), and all participants provided informed consent prior to participation.

Figure 1 presents the conceptual framework of the study. All volunteers provided exhaled breath samples that were analyzed using an electronic nose system comprising sixteen gas sensors. The acquired data were processed through feature extraction and data normalization. Because the study involved three distinct classes, the data were organized into three datasets to support both binary and multiclass analyses and were then classified using LDA, SVM, random forest (RF), extreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), ResNet-1D, VGG-1D, EfficientNet-1D, and Inception-1D. After evaluating all datasets, the dataset with the best classification performance was selected for dimensionality reduction using PCA. Model performance was evaluated using confusion matrices with five evaluation metrics, as well as ROC curves and area under the curve (AUC) values to assess class separability.

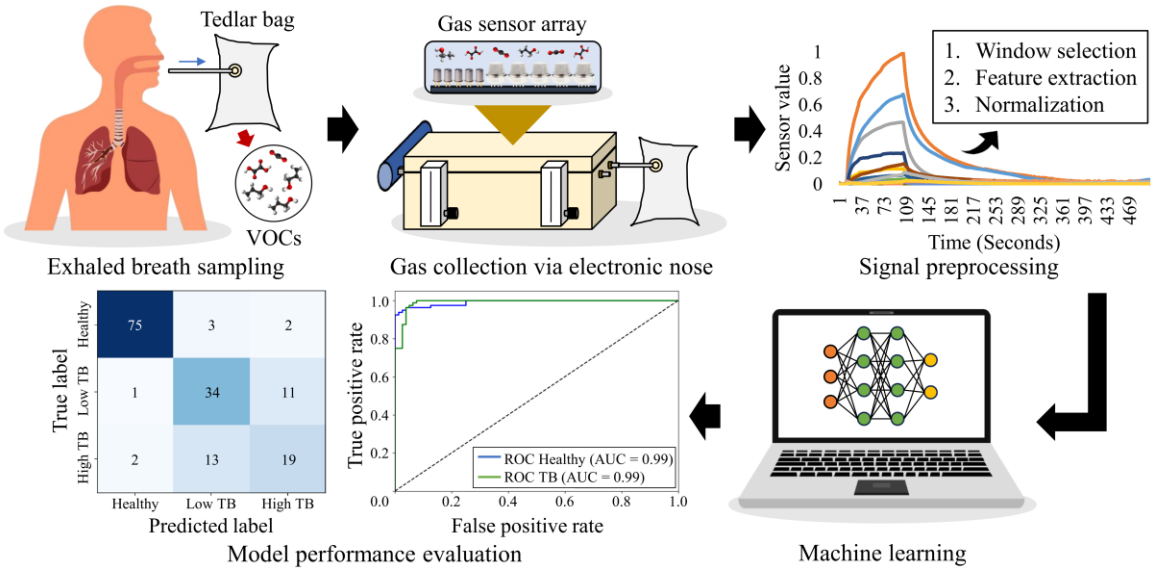
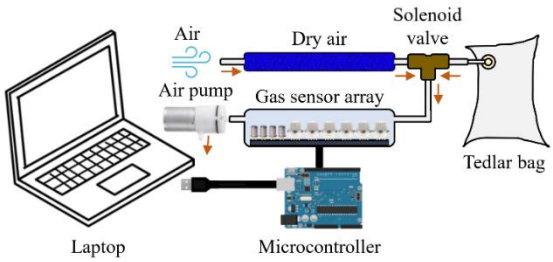


Figure 1. The conceptual framework of the proposed study

2.2. Electronic nose system

The electronic nose system used in this study is shown in Figure 2. Figure 2(a) depicts the design, which comprises three primary components, namely a chamber of 150 mL volume equipped with sixteen MOS-based gas sensors, along with a microcontroller and a laptop. All electronic components were mounted on a printed circuit board to minimize circuit noise. Table 1 lists the gas sensors and their target compounds widely used in exhaled-breath analysis for pulmonary disease detection. The TGS and MQ sensors were manufactured by Figaro Engineering and Hanwei Electronics, respectively. A voltage divider circuit was used to acquire the sensor signals, which produced voltage outputs proportional to the sensor response while limiting current consumption to ensure measurement stability. A supply of filtered dry air was provided, while a solenoid valve managed the selection between dry and sample air streams. A pump controlled by a flowmeter was used to draw air with airflow set to 100 mL/min. Each measurement was conducted for 100 seconds to record the complete gas signal response from the sample, followed by purging the chamber with dry air after each cycle. Sensor voltages were transformed into digital form through the microcontroller and then sent to the laptop to build the dataset. Figure 2(b) shows the system prototype.



(a)

(b)

Figure 2. Electronic nose system schematic diagram; (a) system layout and (b) experimental setup

Table 1. Gas sensors and target compounds used in the electronic nose system

Sensor name	Target gas	Sensor name	Target gas
MQ-2	Hydrogen, propane, methane, and alcohol	MQ-131	Chlorine, nitrogen oxide, and ozone
MQ-3	Benzene and alcohol	MQ-135	Acetone, toluene, ammonium, and alcohol
MQ-4	Methane and propane	MQ-136	Ammonium and hydrogen sulfide
MQ-5	Propane, methane, and hydrogen	MQ-137	Ethanol and ammonia
MQ-6	Methane and propane	TGS813	Ethanol, hydrogen, and propane
MQ-7	Hydrogen and carbon monoxide	TGS822	Acetone, benzene, and ethanol
MQ-8	Hydrogen and alcohol	TGS2600	Hydrogen and ethanol
MQ-9	Carbon monoxide, propane, and methane	TGS2602	Ethanol, toluene, and hydrogen sulfide

2.3. Feature engineering

Data from each sensor were processed by calculating the average ($\bar{x}_n$), as stated in (1). The average sensor response was selected as the primary feature because it provides a robust and consistent representation of the sensor signal within the selected time window. Although the sensor response does not always reach a complete steady-state condition due to measurement time constraints, the mean value captures the dominant signal behavior while reducing the influence of transient fluctuations and noise. Compared with dynamic features such as slope or peak values, the average is less sensitive to signal instability and variability across measurements, making it a more reliable feature for classification under practical experimental conditions. To achieve consistent scaling across sensors, max normalization (x_n^{norm}) was used through the scikit-learn library, where scaling was applied using the maximum absolute value of every feature vector, as stated in (2).

$$\bar{x}_n = \frac{\sum_{T=1}^{k} x_n^T}{k} \quad (1)$$

$$x_n^{norm} = \frac{x_n}{\|x\|_{max}} \quad (2)$$

Here, x_n^T indicates the output of the n-th sensor at time T, k refers to the length of the sample, n corresponds to the gas sensor index, and $\|x\|_{max}$ denotes the maximum absolute magnitude of the feature vector.

2.4. Classification models

This study employed LDA, SVM, RF, and AdaBoost from the scikit-learn library; XGBoost from the XGBoost library; and VGG, ResNet, Inception, and EfficientNet from the Keras library. All model parameters were configured based on the settings provided in the respective libraries. For VGG, ResNet, Inception, and EfficientNet, the architectures were adapted by incorporating 1D convolutional (Conv1D) layers to accommodate one-dimensional time-series sensor signals.

LDA seeks to maximize the ratio of between-class variance to within-class variance to enhance class separability [28]. Two common parameters in LDA are the solver type and shrinkage. In this study, the singular value decomposition solver was used with no shrinkage. This configuration enables efficient computation of the LDA projection while maintaining effective separation among classes in the sensor data.

SVM identifies an optimal hyperplane that maximizes the margin between classes in the feature space. In this study, a radial basis function (RBF) kernel was used, with the regularization parameter C set to 1.0, and gamma (γ) set to "scale". These settings control model complexity and prevent overfitting while allowing the RBF kernel to capture nonlinear relationships among features, supporting discrimination across sensor datasets with multiple classes.

RF is a bagging-based ensemble algorithm that constructs multiple decision trees using randomly selected subsets of samples and features [29]. In this study, RF was configured with 100 trees, the Gini criterion, maximum depth set to "None", maximum number of features set to square root (sqrt), with the minimum split samples set to 2 and the minimum leaf samples set to 1. By randomly sampling features at each split and aggregating predictions across trees, RF captures nonlinear feature interactions and complex relationships in the sensor data, contributing to model stability and robust classification performance for datasets with multiple classes.

XGBoost is a tree-based boosting algorithm that optimizes a loss function through an efficient gradient boosting framework [30]. In this study, XGBoost was configured with a learning rate of 0.3, gamma of 0, maximum depth of 6, minimum child weight of 1, and 80 estimators to balance bias and variance. By combining regularization and complex feature modeling, the algorithm can improve prediction accuracy for sensor data representing different categories.

AdaBoost is an ensemble model that enhances classification accuracy by combining multiple weak learners. In this study, AdaBoost was implemented with a shallow decision tree as the base estimator, 50 trees, a learning rate of 1.0, and stagewise additive modeling using a multiclass exponential loss function.

The algorithm adaptively increases the weights of misclassified samples in each iteration, encouraging subsequent learners to focus on difficult cases. This mechanism supports class separability and can improve model performance for sensor data across classes.

The VGG-1D model employs a structured design for extracting temporal features from sensor signals [31]. Figure 3 shows that the model consists of two convolutional blocks, where the first block contains two Conv1D layers with 64 filters and the second block contains two Conv1D layers with 128 filters. Each layer is followed by a rectified linear unit (ReLU) activation function and a max-pooling operation to reduce temporal dimensions. Flattening is applied to the extracted features and fed into a fully connected layer with 256 neurons and a dropout rate of 0.5, followed by SoftMax classification.

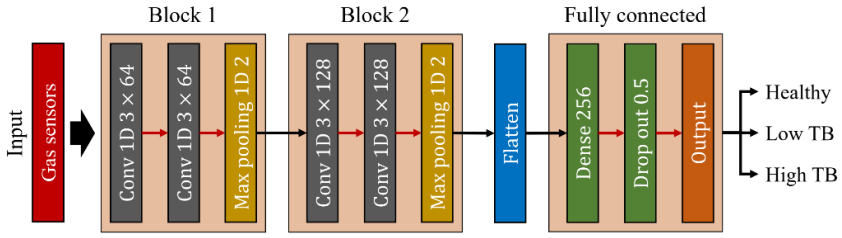


Figure 3. Configuration of the VGG-1D architecture

The ResNet-1D architecture, depicted in Figure 4, employs residual learning to enable deeper neural networks, which mitigates vanishing gradient issues and enables efficient extraction of features from sensor data [32]. An initial Conv1D layer with 64 filters (kernel size 7 and stride 2) is used, followed by three residual blocks with progressively increasing filter sizes of 64, 128, and 256. Each convolutional layer is followed by ReLU activation and batch normalization. Skip connections maintain gradient flow and training stability, while classification is performed using global average pooling and a SoftMax layer.

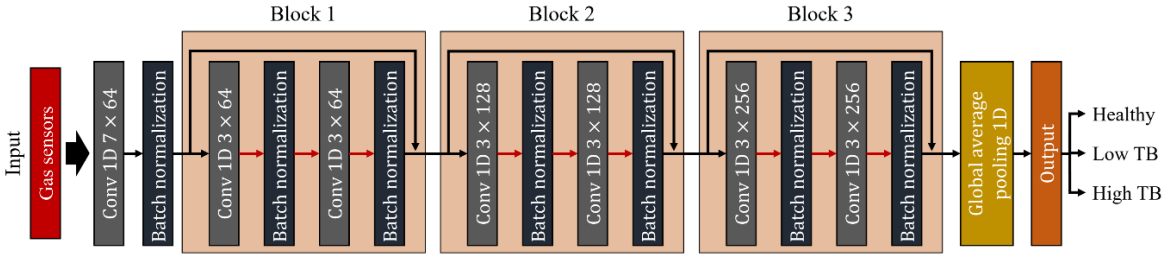


Figure 4. ResNet-1D network architecture design

The Inception-1D model, displayed in Figure 5, performs multi-scale feature extraction from sensor signals via multi-branch convolutional pathways [33]. Each block contains four parallel branches: a 1×1 convolution, a 1×1 convolution followed by a 3×3 convolution, a 1×1 convolution followed by a 5×5 convolution, and a 3×3 max-pooling layer followed by a 1×1 convolution. The concatenated branch outputs are flattened and fed into a dense layer with 128 neurons and 0.5 dropout before SoftMax classification.

The EfficientNet-1D model, presented in Figure 6, uses mobile inverted bottleneck convolution (MBConv) blocks to achieve efficient feature extraction with reduced computational complexity [34]. Three temporal-downsampling MBConv stages with 32, 64, and 128 filters follow an initial convolutional layer with 32 filters. In each MBConv block, input channels are first expanded, and depthwise convolution is then applied, followed by batch normalization and ReLU activation, and the output is projected to the target dimensionality. When the input and output dimensions match, skip connections are applied to enhance training stability. The extracted features undergo global pooling before being classified using a SoftMax layer.

The training configuration of the deep learning models used in this study is summarized in Table 2. The models were trained using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy loss. A batch size of 5 and a maximum of 1000 epochs were used. Early stopping was applied based on the training loss with a patience of 10 epochs to terminate training once the model converged. Although

validation-based early stopping is commonly used, training loss was employed in this study due to the limited dataset size. A weight decay of 1e-5 and a random seed of 6 were used for regularization and reproducibility.

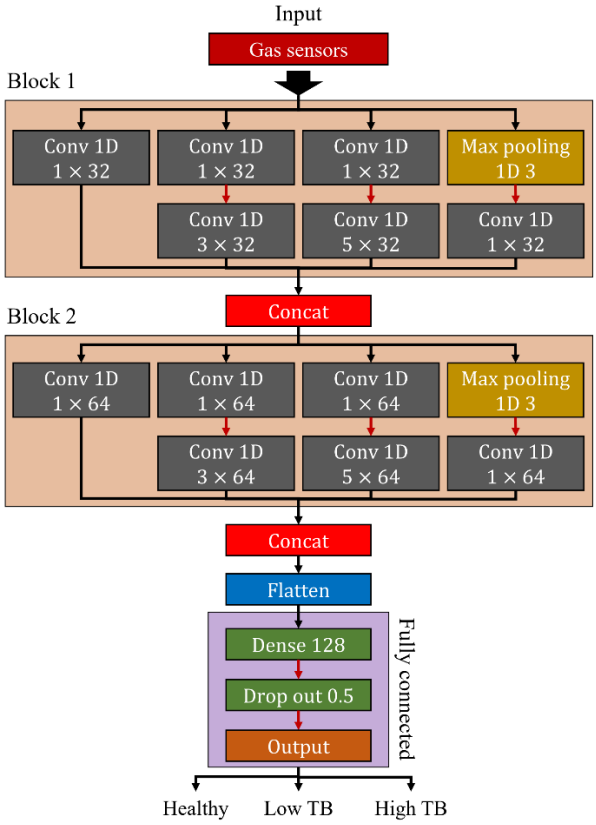


Figure 5. Design of the Inception-1D architecture

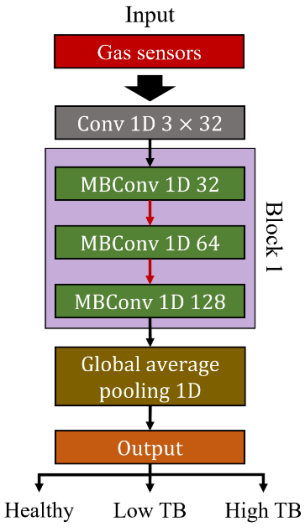


Figure 6. EfficientNet-1D model configuration

2.5. Principal component analysis

PCA is a dimensionality reduction technique that projects data into a principal component (PC) space while retaining the maximum variance from the original dataset. The PCs are ranked according to the amount of variance they explain, with every PC representing a weighted combination of the original features

[35], [36]. PCA was applied in this study to reduce the dimensionality of the dataset, which produced PC scores as inputs for the classification models.

Table 2. Configuration of deep learning model training for this study

Parameter	Value
Optimizer	Adam
Loss	Categorical cross-entropy
Learning rate	0.001
Batch size	5
Epochs	1000
Weight decay	1e-5
Early stopping	Patience=10
Random seed	6

2.6. Evaluation of model performance

Accuracy, precision, recall (sensitivity), specificity, and F1-score were used to evaluate the models, all computed from true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), as stated in (3)–(8), respectively, with precision, recall, specificity, and F1-score obtained through macro averaging. The F1-score was adopted as the primary metric because it captures the trade-off between precision and recall. The discriminative ability of the models was assessed using ROC–AUC based on the true positive rate (TPR) and false positive rate (FPR), where higher values indicate better class separability.

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall\ (Sensitivity/TPR) = \frac{TP}{TP+FN} \quad (5)$$

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

$$F1-score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (7)$$

$$FPR = \frac{FP}{FP+TN} \quad (8)$$

3. RESULTS AND DISCUSSION

3.1. Experiment configuration and data description

A desktop system running Windows 11 Pro (64-bit), equipped with an Intel Core i5-12400F, 32 GB RAM, and an NVIDIA RTX 4060 GPU (8 GB), was employed for the experiments. Python within Jupyter Notebook was used for data preprocessing and classification tasks. LDA applied to the non-normalized dataset of healthy individuals and TB patients was used as the initial configuration. Three evaluation scenarios were considered: classification of healthy and TB, TB severity, and healthy and TB severity levels. Figure 7 shows the gas sensor response patterns for each class, where Figure 7(a) presents healthy subjects, Figure 7(b) depicts low TB cases, and Figure 7(c) illustrates high TB cases. While the low TB and high TB conditions exhibit similar temporal trends, several sensors show differences in response magnitude. In contrast, the healthy group exhibits a distinct response pattern compared with the TB conditions.

Table 3 demonstrates the sample distribution across classes. The healthy and TB dataset is balanced, comprising 80 healthy and 80 TB samples. In contrast, the TB severity dataset is imbalanced, with 46 low TB and 34 high TB samples. To evaluate model performance, stratified 3-fold cross-validation was employed to preserve class distribution across all folds, particularly considering the limited dataset size. A smaller number of folds was selected to ensure sufficient training samples while maintaining reliable evaluation. The procedure was performed once, and the results were averaged across the three folds. The results are presented as mean ± standard deviation values. To prevent data leakage, normalization and PCA were performed exclusively on the training data within each cross-validation fold, and the resulting transformations were applied to the corresponding test data.

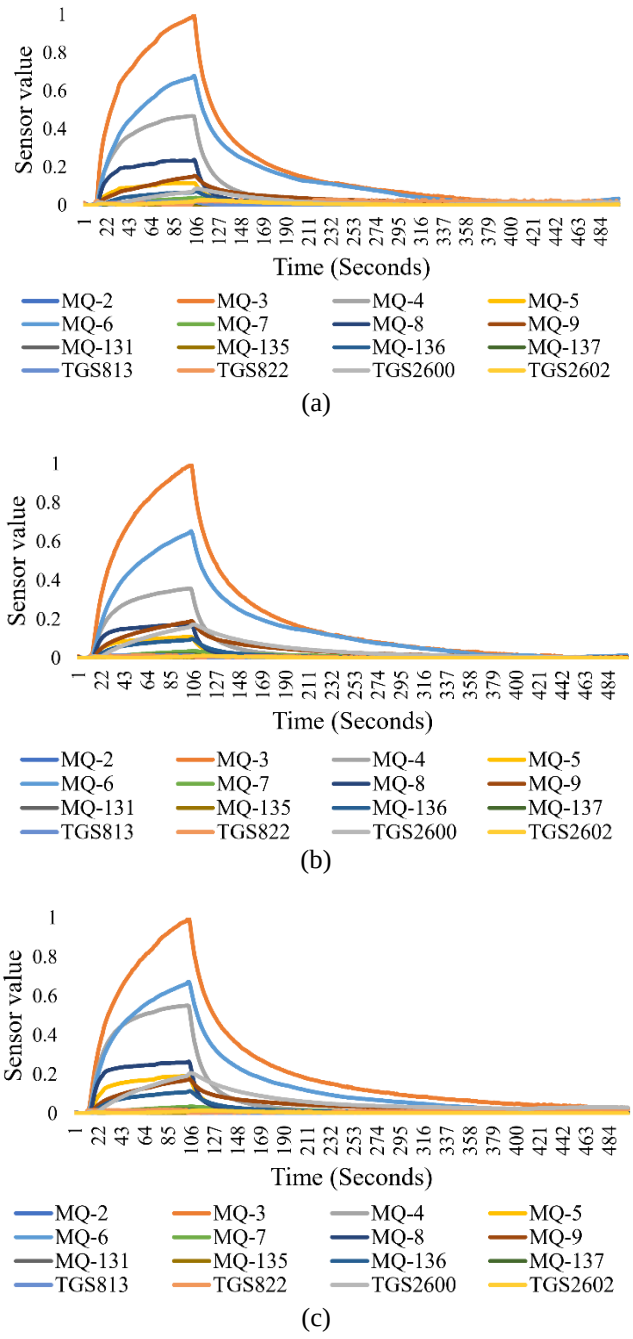


Figure 7. The gas sensor responses to exhaled breath of; (a) healthy, (b) low TB, and (c) high TB

Table 3. Dataset distribution across classes

Category	Total subjects
Healthy	80
Low TB	46
High TB	34

3.2. Data preprocessing analysis

Feature extraction was carried out by computing the average value of each sensor data. The time interval from 40 to 100 seconds was divided into 12 window segments, with each segment containing five averaged data points, as shown in Figure 8. Each window was evaluated independently, and the optimal window was selected according to the highest F1-score. Table 4 summarizes the evaluation results of all windows using LDA on the healthy and TB dataset. Window 5 achieved the highest F1-score of 93.74%,

while Windows 1, 2, 3, 6, 7, and 10 showed comparable performance with F1-scores around 93.10%. Lower performance was observed for Windows 4, 8, 9, 11, and 12. As a result, Window 5 represents the most discriminative segment of the sensor response, corresponding to a stable phase with greater class separation. The performance metrics for Window 5 were 93.76% accuracy, 94.37% precision, 93.80% recall, 93.80% specificity, and an F1-score of 93.74%; therefore, it was selected for the next normalization stage.

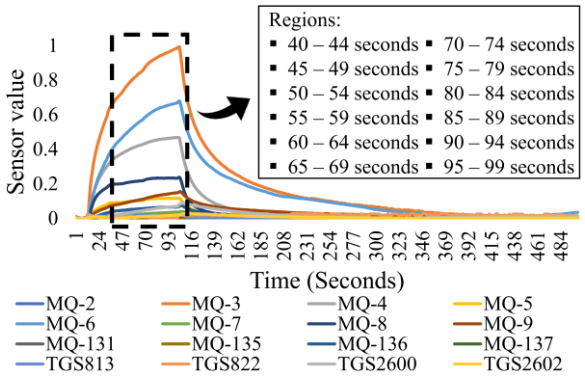


Figure 8. Region selection strategy for feature extraction

Table 4. Classification results for each region using LDA

Region	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)
1	93.13	93.90	93.19	93.19	93.10
2	93.13	93.90	93.19	93.19	93.10
3	93.13	93.90	93.19	93.19	93.10
4	91.87	92.87	91.93	91.93	91.82
5	93.76	94.37	93.80	93.80	93.74
6	93.13	93.77	93.16	93.16	93.10
7	93.14	93.71	93.16	93.16	93.12
8	92.51	93.25	92.55	92.55	92.48
9	92.50	92.97	92.52	92.52	92.48
10	93.13	93.53	93.16	93.16	93.11
11	91.88	92.29	91.90	91.90	91.86
12	91.89	92.29	91.93	91.93	91.88

Subsequently, normalization was applied to the dataset with the selected window, as depicted in Figure 9, and the results were evaluated using LDA. All performance metrics demonstrated improvement, with accuracy reaching 94.39%, precision 94.79%, recall 94.42%, specificity 94.42%, and an F1-score of 94.37%. These findings show that normalization successfully reduced feature-scaling imbalance, allowing the model to better capture the underlying class structure and improve overall classification performance. Accordingly, the normalized dataset was retained for further analysis.

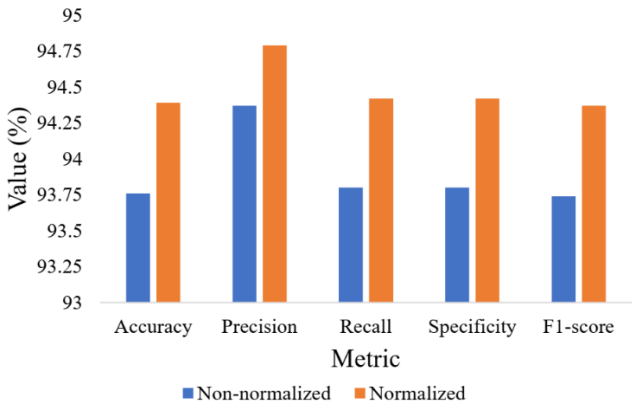


Figure 9. Comparison of LDA performance on non-normalized and normalized datasets

3.3. Evaluation of all datasets using LDA, SVM, RF, XGBoost, and AdaBoost models

Classification was performed on all datasets, including the healthy and TB dataset, the TB severity dataset, and the combined healthy and TB severity dataset, using LDA, SVM, RF, XGBoost, and AdaBoost. Table 5 shows the classification results for each model across all datasets. For the healthy and TB dataset, RF slightly exceeded LDA, yielding an F1-score of 94.40%. In comparison, XGBoost and AdaBoost obtained F1-scores of 93.77% and 92.48%, respectively, while SVM showed the lowest performance, with an F1-score below 69%. These results indicate that LDA and RF can accurately discriminate between healthy and TB breath profiles, likely due to LDA's linear decision boundaries and RF's ensemble-based structure, which enable effective modeling of class differences. In contrast, the lower performance of SVM is likely due to feature-space characteristics that are not well suited to the model.

Table 5. Classification results for each dataset using LDA, SVM, RF, XGBoost, and AdaBoost models

Model	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)
Healthy and TB					
LDA	94.39	94.79	94.42	94.42	94.37
SVM	71.26	80.98	71.34	71.34	68.67
RF	94.41	94.62	94.37	94.37	94.40
XGBoost	93.78	93.89	93.78	93.78	93.77
AdaBoost	92.49	92.54	92.47	92.47	92.48
TB severity (low TB and high TB)					
LDA	63.72	63.69	63.98	63.98	63.43
SVM	56.27	35.39	49.91	49.91	39.67
RF	62.54	62.70	61.42	61.42	59.92
XGBoost	65.00	64.64	64.61	64.61	64.30
AdaBoost	62.54	62.46	62.20	62.20	61.94
Healthy and TB severity (low TB and high TB)					
LDA	73.11	66.87	67.51	87.41	66.76
SVM	57.55	27.97	43.21	73.81	32.85
RF	76.24	70.93	69.16	87.73	69.61
XGBoost	73.77	68.31	67.31	86.95	66.77
AdaBoost	73.10	68.42	67.80	86.96	67.55

For the TB severity dataset, LDA achieved an F1-score of 63.43%, RF 59.92%, XGBoost 64.30%, AdaBoost 61.94%, and SVM 39.67%, with SVM again showing the lowest performance. LDA, XGBoost, and AdaBoost were still able to distinguish TB severity levels because they can capture small differences in sensor response patterns through either linear discriminant projection or ensemble-based feature weighting. In contrast, the lower performance of SVM in this scenario is influenced by overlapping feature distributions between low TB and high TB classes, which can reduce the ability of the hyperplane to separate the classes.

Similarly, in the combined dataset of healthy subjects and TB severity levels, RF yielded the highest F1-score of 69.61%, followed by AdaBoost at 67.55%, XGBoost at 66.77%, LDA at 66.76%, and SVM at 32.85%, with SVM again showing the lowest performance. The comparatively lower performance of SVM in this setting is related to the increased dimensionality and complexity of the feature space, which can reduce its ability to achieve optimal class separation. Overall, the healthy and TB dataset can be distinguished reliably, while differentiating TB severity levels remains challenging. In addition, the combined dataset shows partial separability. These findings motivated the application of more advanced deep-learning models in the next stage to further evaluate classification performance.

3.4. Evaluation of all datasets with VGG-1D, ResNet-1D, Inception-1D, and EfficientNet-1D

This stage evaluated advanced deep-learning models across all datasets. The evaluation began with the healthy and TB dataset, followed by the elimination of models with lower classification performance. Figure 10 depicts the classification results for these models on the healthy and TB dataset. VGG-1D and EfficientNet-1D were eliminated, yielding F1-scores below 93.80% and performing worse than traditional models. In contrast, Inception-1D and ResNet-1D achieved high performance, each reaching an F1-score of 94.99%, exceeding the results of the traditional models as well as VGG-1D and EfficientNet-1D. These results indicate that the deeper and more optimized structures of Inception-1D and ResNet-1D enable more accurate extraction of discriminative patterns from the sensor data.

Subsequently, Inception-1D and ResNet-1D were evaluated using the TB severity dataset and the combined healthy and TB severity dataset. Figure 11 depicts the classification performance of both models on these datasets. In Figure 11(a), ResNet-1D accurately distinguished between low TB and high TB, achieving an F1-score of 68.50%, outperforming Inception-1D with 64.13% as well as all traditional models. This improvement is likely due to ResNet's residual connections, which preserve important features while avoiding gradient degradation in deeper layers. The ability to distinguish low TB and high TB severity levels

can support treatment prioritization and closer clinical monitoring for patients with higher bacterial burden. In addition, accurate severity classification can assist clinicians in identifying patients with potentially higher transmission risk, particularly in resource-limited healthcare settings. For the combined healthy and TB severity dataset, as shown in Figure 11(b), both Inception-1D and ResNet-1D achieved F1-scores above 71%, indicating better discrimination of healthy individuals and TB severity levels compared with traditional models. However, Inception-1D outperformed ResNet-1D, achieving an F1-score of 74.07%. This advantage likely stems from Inception's multi-scale feature extraction, which enables more effective capture of diverse temporal patterns in sensor signals. Since ResNet-1D and Inception-1D yielded the highest performance on the healthy and TB dataset, both models were further evaluated using PCA-based dimensionality reduction.

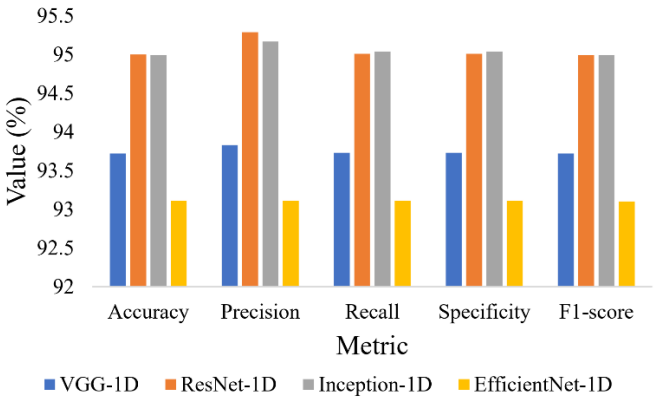


Figure 10. Performance comparison of 1D deep learning models on the healthy and TB dataset

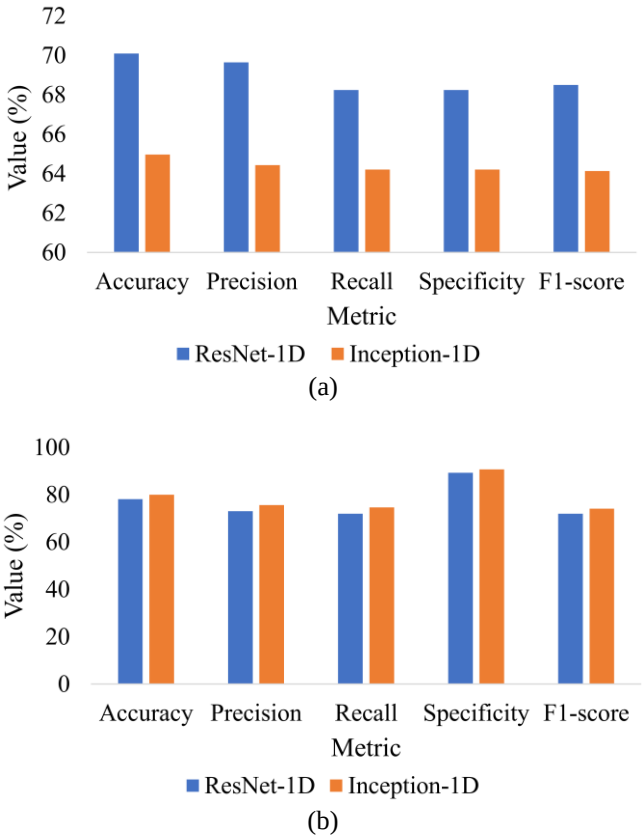


Figure 11. Comparison of ResNet-1D and Inception-1D performance on; (a) the TB severity dataset and (b) the healthy and TB severity dataset

3.5. Evaluation of ResNet-1D and Inception-1D on the PCA-reduced healthy and TB dataset

Since the electronic nose system consists of sixteen gas sensors, the dataset was reduced to between 15 and 2 PCs. Table 6 presents the evaluation results of ResNet-1D and Inception-1D on the PCA-reduced datasets. Both models achieved F1-scores above 91% when using 7 PCs. In contrast, reducing the dataset to 6–2 PCs resulted in F1-scores below 90% for all models, indicating that excessive dimensionality reduction removes essential variance needed for effective pattern discrimination, which makes the classification more difficult. Inception-1D demonstrated stable performance above 92.40% across 7–14 PCs, with the highest F1-score of 95.64% achieved using 10 PCs. This shows that its multi-scale convolutional structure effectively captures complementary information across multiple PCs. In contrast, ResNet-1D exhibited less stability across PCs; for example, using 13 PCs resulted in a drop below 90%, and its highest F1-score was 93.74% using 10 PCs. Nevertheless, with 15 PCs, ResNet-1D slightly outperformed Inception-1D, yielding 91.21% compared with 89.32%. Based on these results, Inception-1D was selected as the optimal model due to its consistently strong and stable performance across 7–14 PCs. To achieve an optimal balance between model accuracy and computational efficiency, 10 PCs were selected for Inception-1D. Table 7 summarizes the best-performing models across all classification scenarios. Overall, advanced one-dimensional deep learning models performed better than traditional machine-learning models, while PCA further improved the performance of Inception-1D by reducing feature redundancy and preserving discriminative information.

Table 6. ResNet-1D and Inception-1D performance on the PCA-reduced healthy and TB dataset

	ResNet-1D					Inception-1D				
PC	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)
15	91.23	91.69	91.26	91.26	91.21	89.35	89.51	89.32	89.32	89.32
14	93.11	93.58	93.04	93.04	93.08	92.48	92.96	92.40	92.40	92.44
13	90.02	91.25	89.98	89.98	89.91	94.36	94.55	94.35	94.35	94.35
12	92.48	92.86	92.55	92.55	92.47	94.37	94.57	94.42	94.42	94.37
11	92.51	93.04	92.52	92.52	92.49	93.14	93.33	93.19	93.19	93.14
10	93.76	94.14	93.73	93.73	93.74	95.64	95.75	95.63	95.63	95.64
9	92.50	93.05	92.45	92.45	92.47	94.36	94.75	94.30	94.30	94.34
8	93.71	94.04	93.66	93.66	93.68	94.34	94.35	94.35	94.35	94.34
7	91.25	91.30	91.24	91.24	91.25	92.49	92.60	92.50	92.50	92.48
6	90.00	90.19	89.98	89.98	89.98	88.76	89.00	88.75	88.75	88.73
5	88.17	88.52	88.11	88.11	88.13	83.79	84.16	83.74	83.74	83.74
4	83.74	84.93	83.88	83.88	83.63	85.62	87.57	85.75	85.75	85.44
3	66.96	67.21	66.98	66.98	66.85	69.47	69.64	69.49	69.49	69.32
2	72.49	74.81	72.48	72.48	71.83	69.39	70.14	69.49	69.49	69.17

Table 7. Summary of best-performing models across all datasets

Classification scenario	Model category	Best model	F1-score (%)
Healthy and TB	Traditional ML	RF	94.40
Healthy and TB	Deep learning	Inception-1D/ResNet-1D	94.99
Healthy and TB (PCA)	PCA+deep learning	Inception-1D+PCA	95.64
TB severity	Traditional ML	XGBoost	64.30
TB severity	Deep learning	ResNet-1D	68.50
Healthy and TB severity	Traditional ML	RF	69.61
Healthy and TB severity	Deep learning	Inception-1D	74.07

The final stage visualizes the performance of Inception-1D across all datasets. For the healthy and TB dataset reduced to 10 PCs, performance was assessed using ROC–AUC curves, as displayed in Figure 12. The ROC–AUC values for both classes reached 0.99, indicating high class separability after PCA reduction. The corresponding confusion matrix is provided in Figure 13, where Inception-1D achieved an accuracy of 95.64%, precision of 95.75%, recall of 95.63%, specificity of 95.63%, and an F1-score of 95.64±3.17%. For the healthy and TB dataset without PCA reduction, the confusion matrix is shown in Figure 14, achieving an accuracy of 94.99%, precision of 95.17%, recall of 95.04%, specificity of 95.04%, and an F1-score of 94.99±0.92%. Figure 15 shows the confusion matrix of Inception-1D on the TB severity dataset, yielding an accuracy of 64.96%, precision of 64.43%, recall of 64.20%, specificity of 64.20%, and an F1-score of 64.13±6.20%. Because ResNet-1D achieved high performance on this dataset, its confusion matrix is shown in Figure 16, with an accuracy of 70.09%, precision of 69.64%, recall of 68.24%, specificity of 68.24%, and an F1-score of 68.50±5.71%. Figure 17 presents the confusion matrix for Inception-1D applied to the healthy and TB severity dataset, with an accuracy of 79.97%, precision of 75.60%, recall of 74.64%, specificity of 90.64%, and an F1-score of 74.07±5.22%.

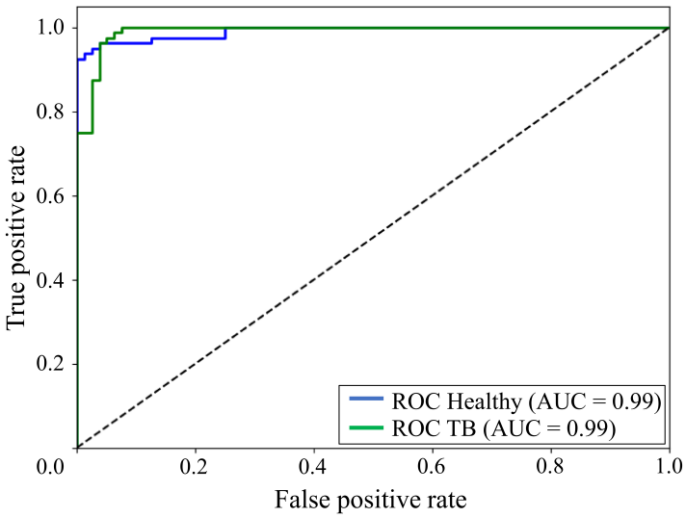


Figure 12. ROC–AUC curve of Inception-1D on the healthy and TB dataset (10 PCs)

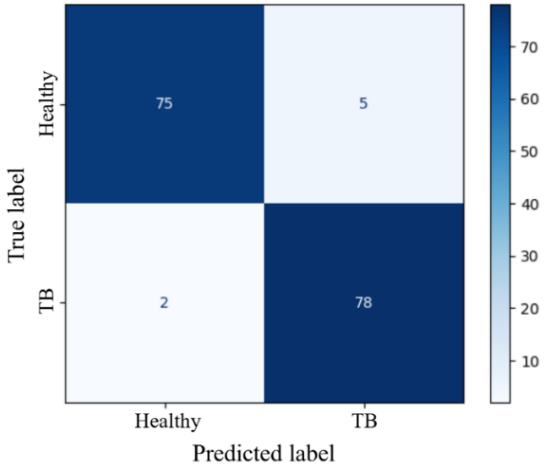


Figure 13. Confusion matrix of Inception-1D on the healthy and TB dataset (10 PCs)

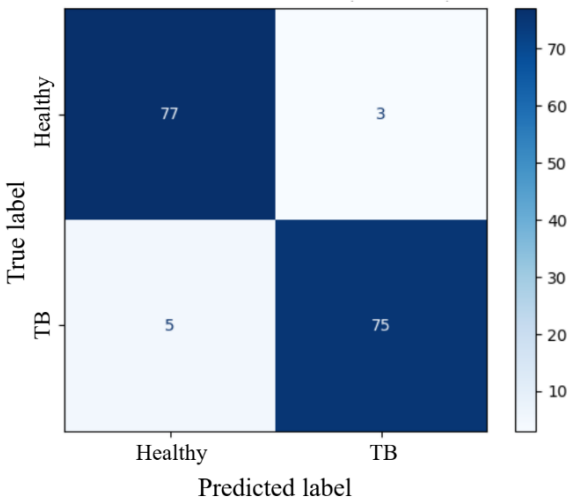


Figure 14. Inception-1D confusion matrix on the healthy and TB dataset (original features)

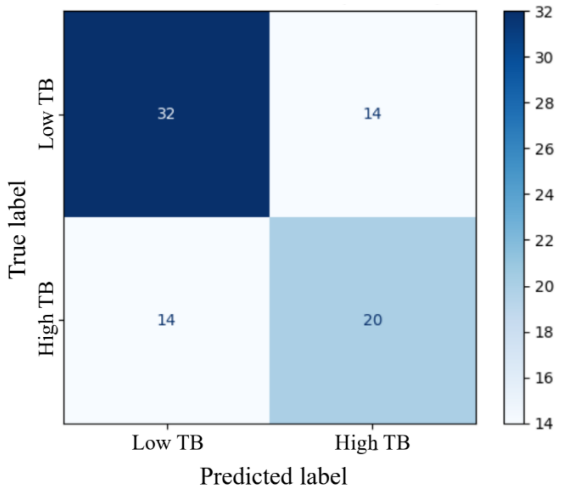


Figure 15. Confusion matrix of Inception-1D on the TB severity dataset (original features)

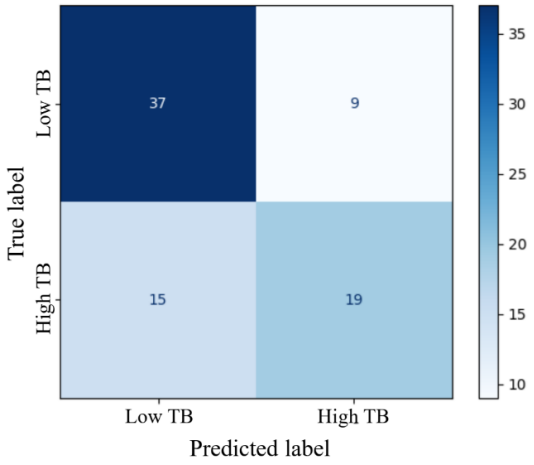


Figure 16. ResNet-1D confusion matrix on the TB severity dataset (original features)

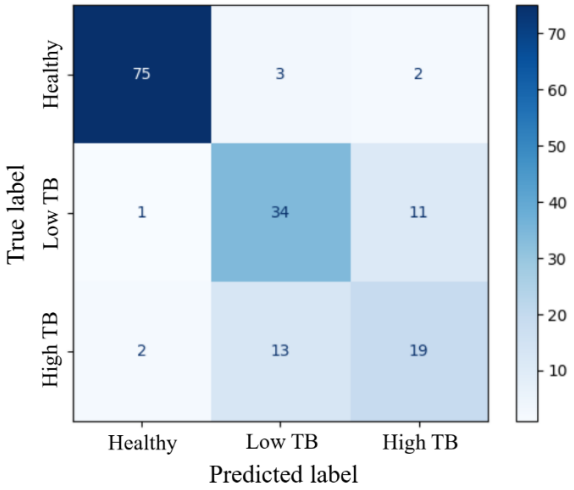


Figure 17. Inception-1D confusion matrix on the healthy and TB severity dataset (original features)

3.6. Limitations and technical discussion

Despite the promising results, several limitations should be acknowledged. First, the dataset size is relatively small, consisting of only 160 subjects, which can limit the generalization ability of the proposed

models. Second, the TB severity dataset is imbalanced, with fewer high TB samples than low TB samples, which may affect classification performance and bias the models toward the majority class. The lower performance in TB severity classification indicates that low and high TB exhibit highly similar gas sensor response patterns, resulting in overlapping feature distributions. This overlap also explains the poor performance of SVM, which depends on clear margin separation. In contrast, ResNet-1D can better capture subtle nonlinear patterns, although it requires higher computational resources. Further optimization of both classical machine-learning and deep learning models may improve overall performance.

4. CONCLUSION

This study employed an electronic nose system with advanced deep-learning models for classifying healthy subjects and TB severity levels using exhaled-breath analysis. The system consisted of sixteen MOS gas sensors, with preprocessing involving window selection, feature extraction, and normalization. Experimental results demonstrated that traditional models—including LDA, RF, XGBoost, and AdaBoost—achieved F1-scores above 92% for healthy and TB classification, whereas TB severity classification remained challenging due to highly similar feature patterns across classes. ResNet-1D and Inception-1D performed best for healthy and TB classification, each obtaining an F1-score of 94.99%. For TB severity classification, ResNet-1D achieved the highest F1-score of 68.50%, while Inception-1D achieved 74.07% for the healthy and TB severity dataset. These results indicate that deep-learning architectures can better capture complex temporal patterns and nonlinear representations from gas sensor signals than traditional models. PCA-based dimensionality reduction was applied to the healthy and TB dataset to reduce redundancy and improve generalization, with Inception-1D achieving the highest performance with ten principal components and an F1-score of 95.64%. However, this study has limitations, including the relatively small dataset size and class imbalance in the TB severity dataset, which may affect model generalization.

Future work should focus on expanding the dataset, addressing class imbalance, improving system deployment through algorithm optimization, real-time processing, and hardware integration, as well as conducting clinical-scale validation, deploying embedded hardware, integrating mobile or cloud environments, and extending the framework for multi-disease detection using breath biomarkers.

ACKNOWLEDGMENTS

The authors sincerely thank Airlangga University Hospital, Surabaya, for facilitating access to TB patients. The authors also acknowledge the support of Institut Teknologi Sepuluh Nopember (ITS).

FUNDING INFORMATION

This research was funded by Institut Teknologi Sepuluh Nopember (ITS) and managed under the Strategic Research Grant (SRG) Type D Scheme (Contract No. 1670/PKS/ITS/2026).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dava Aulia	✓	✓	✓		✓				✓		✓			
Riyanarto Sarno				✓	✓			✓		✓		✓		✓
Muhammad Rivai		✓		✓			✓		✓			✓		✓
Muhammad Amin				✓		✓	✓			✓		✓		
Alfian Nur Rosyid						✓	✓	✓		✓				
Kelly Rossa Sungkono						✓				✓			✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

This study involving human participants was conducted in accordance with the Helsinki Declaration and approved by the authors' institutional review board or equivalent committee.

DATA AVAILABILITY

The data associated with this study are available from [initials: MR] upon reasonable request.

REFERENCES

[1] D. Tiwari and A. R. Martineau, "Inflammation-mediated tissue damage in pulmonary tuberculosis and host-directed therapeutic strategies," *Seminars in Immunology*, vol. 65, pp. 1–16, Jan. 2023, doi: 10.1016/j.smim.2022.101672.

[2] L. Luies and I. du Preez, "The echo of pulmonary tuberculosis: Mechanisms of clinical symptoms and other disease-induced systemic complications," *Clinical Microbiology Reviews*, vol. 33, no. 4, pp. 1–19, Sep. 2020, doi: 10.1128/CMR.00036-20.

[3] D. P. Maison, "Tuberculosis pathophysiology and anti-VEGF intervention," *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, vol. 27, pp. 1–6, May 2022, doi: 10.1016/j.jctube.2022.100300.

[4] B. Xu, M. Yuan, L. Yang, L. Huang, J. Li, and Z. Tan, "Recent Advances in Clinical Research of Prophylactic Vaccines Against Tuberculosis," *Vaccines*, vol. 13, no. 9, pp. 1–14, Sep. 2025, doi: 10.3390/vaccines13090959.

[5] T. A. Nguyen *et al.*, "Population-wide active case finding as a strategy to end TB," *The Lancet Regional Health - Western Pacific*, vol. 46, pp. 1–10, May 2024, doi: 10.1016/j.lanwpc.2024.101047.

[6] World Health Organization (WHO), "Global tuberculosis report 2024," 2025. https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2024. (Accessed Oct. 28, 2025).

[7] B. K. Moore, S. M. Graham, S. Nandakumar, J. Doyle, and S. A. Maloney, "Pediatric Tuberculosis: A Review of Evidence-Based Best Practices for Clinicians and Health Care Providers," *Pathogens*, vol. 13, no. 6, pp. 1–38, Jun. 2024, doi: 10.3390/pathogens13060467.

[8] L. Peng *et al.*, "Diagnostic Accuracy of Mycobacterium tuberculosis Antigen-Based Skin Tests (TBSTs) for Tuberculosis Infection Compared with TST and IGRA: A Network Meta-Analysis," *Pathogens*, vol. 13, no. 12, pp. 1–20, Nov. 2024, doi: 10.3390/pathogens13121050.

[9] D. L. Markussen *et al.*, "The diagnostic utility of microscopic quality assessment of sputum samples in the era of rapid syndromic PCR testing," *Microbiology Spectrum*, vol. 11, no. 5, pp. 1–14, Oct. 2023, doi: 10.1128/spectrum.03002-23.

[10] A. M. Elbrolosy, R. H. El Helbawy, O. M. Mansour, and R. A. Latif, "Diagnostic utility of GeneXpert MTB/RIF assay versus conventional methods for diagnosis of pulmonary and extra-pulmonary tuberculosis," *BMC Microbiology*, vol. 21, no. 1, pp. 1–10, Dec. 2021, doi: 10.1186/s12866-021-02210-5.

[11] E. M. Bijker *et al.*, "Exhaled breath analysis: A promising triage test for tuberculosis in young children," *Tuberculosis*, vol. 149, pp. 1–4, Dec. 2024, doi: 10.1016/j.tube.2024.102566.

[12] T. Jaya *et al.*, "MicroRNAs in exhaled breath Condensate: Novel non-invasive biomarkers for tuberculosis diagnosis," *Tuberculosis*, vol. 154, pp. 1–7, Sep. 2025, doi: 10.1016/j.tube.2025.102670.

[13] L. Fu *et al.*, "A cross-sectional study: a breathomics based pulmonary tuberculosis detection method," *BMC Infectious Diseases*, vol. 23, no. 1, pp. 1–11, Mar. 2023, doi: 10.1186/s12879-023-08112-3.

[14] C. A. Bobak *et al.*, "Breath can discriminate tuberculosis from other lower respiratory illness in children," *Scientific Reports*, vol. 11, no. 1, pp. 1–9, Feb. 2021, doi: 10.1038/s41598-021-80970-w.

[15] T. R. Mellors *et al.*, "Identification of Mycobacterium tuberculosis using volatile biomarkers in culture and exhaled breath," *Journal of Breath Research*, vol. 13, no. 1, pp. 1–13, Jan. 2019, doi: 10.1088/1752-7163/aacd18.

[16] M. Phillips *et al.*, "Breath biomarkers of active pulmonary tuberculosis," *Tuberculosis*, vol. 90, no. 2, pp. 145–151, Mar. 2010, doi: 10.1016/j.tube.2010.01.003.

[17] A. G. Mpolokang, T. C. Setlhare, S. Bhattacharyya, and G. Chimowa, "New volatile organic compounds from the exhaled breath of active tuberculosis patients," *Scientific Reports*, vol. 15, no. 1, pp. 1–9, Feb. 2025, doi: 10.1038/s41598-025-89178-8.

[18] H. Hendrick, R. Hidayat, G. J. Horng, and Z. H. Wang, "Non-Invasive Method for Tuberculosis Exhaled Breath Classification Using Electronic Nose," *IEEE Sensors Journal*, vol. 21, no. 9, pp. 11184–11191, May 2021, doi: 10.1109/JSEN.2021.3061616.

[19] Y. C. K. Mougang *et al.*, "On-Field Test of Tuberculosis Diagnosis through Exhaled Breath Analysis with a Gas Sensor Array," *Biosensors*, vol. 13, no. 5, pp. 1–13, May 2023, doi: 10.3390/bios13050570.

[20] N. M. Zetola *et al.*, "Diagnosis of pulmonary tuberculosis and assessment of treatment response through analyses of volatile compound patterns in exhaled breath samples," *Journal of Infection*, vol. 74, no. 4, pp. 367–376, Apr. 2017, doi: 10.1016/j.jinf.2016.12.006.

[21] M. K. Nakhleh *et al.*, "Detecting active pulmonary tuberculosis with a breath test using nanomaterial-based sensors," *European Respiratory Journal*, vol. 43, no. 5, pp. 1522–1525, May 2014, doi: 10.1183/09031936.00019114.

[22] D. Iyer, K. Gobinath, K. Kowkuntla, V. V. Wanjari, and G. Banumurthy, "Breath-based lung cancer detection using an ML-driven low-cost sensor array," *Scientific Reports*, vol. 15, no. 1, pp. 1–11, Nov. 2025, doi: 10.1038/s41598-025-26416-z.

[23] D. Aulia, R. Sarno, S. C. Hidayati, A. N. Rosyid, and M. Rivai, "Identification of chronic obstructive pulmonary disease using graph convolutional network in electronic nose," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 1, pp. 264–275, Apr. 2024, doi: 10.11591/ijeecs.v34.i1.pp264-275.

[24] A. M. I. Saktiawati *et al.*, "eNose-TB: A trial study protocol of electronic nose for tuberculosis screening in Indonesia," *PLoS ONE*, vol. 16, pp. 1–4, Apr. 2021, doi: 10.1371/journal.pone.0249689.

[25] R. Duwairi and A. Melhem, "A deep learning-based framework for automatic detection of drug resistance in tuberculosis patients," *Egyptian Informatics Journal*, vol. 24, no. 1, pp. 139–148, Mar. 2023, doi: 10.1016/j.eij.2023.01.002.
[26] C. R. Dhivyaa, K. Nithya, C. S. Kumar, and R. Sudhakar, "Explainable Model of Fusion Network with Enhanced Optimization Approach for Tuberculosis Diagnosis," *IEEE Access*, vol. 12, pp. 176920–176937, 2024, doi: 10.1109/ACCESS.2024.3505608.
[27] A. Tripathi, T. Singh, R. R. Nair, and P. Duraisamy, "Improving Early Detection and Classification of Lung Diseases With Innovative MobileNetV2 Framework," *IEEE Access*, vol. 12, pp. 116202–116217, 2024, doi: 10.1109/ACCESS.2024.3440577.
[28] M. Xu *et al.*, "Detection of CWA Simulants by Electronic Nose Based on Low-Powered MEMS Gas Sensor Array," *IEEE Sensors Journal*, vol. 25, no. 16, pp. 30361–30369, Aug. 2025, doi: 10.1109/JSEN.2025.3587138.
[29] B. V. A., M. Subramoniam, and L. Mathew, "Detection of COPD and Lung Cancer with electronic nose using ensemble learning methods," *Clinica Chimica Acta*, vol. 523, pp. 231–238, Dec. 2021, doi: 10.1016/j.cca.2021.10.005.
[30] V. A. Binson, M. Subramoniam, Y. Sunny, and L. Mathew, "Prediction of Pulmonary Diseases with Electronic Nose Using SVM and XGBoost," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20886–20895, Sep. 2021, doi: 10.1109/JSEN.2021.3100390.
[31] J. Cheng, Q. Song, H. Peng, J. Huang, H. Wu, and B. Jia, "Optimization of VGG16 Algorithm Pattern Recognition for Signals of Michelson–Sagnac Interference Vibration Sensing System," *Photonics*, vol. 9, no. 8, pp. 1–9, Jul. 2022, doi: 10.3390/photonics9080535.
[32] B. G. Ram, K. Howatt, J. Mettler, and X. Sun, "Addressing computation resource exhaustion associated with deep learning training of three-dimensional hyperspectral images using multiclass weed classification," *Artificial Intelligence in Agriculture*, vol. 15, no. 2, pp. 131–146, Jun. 2025, doi: 10.1016/j.aiia.2025.02.005.
[33] J. M. Bueno-Barrachina, Y. Ye-Lin, F. Nieto-del-Amor, and V. Fuster-Roig, "Inception 1D-convolutional neural network for accurate prediction of electrical insulator leakage current from environmental data during its normal operation using long-term recording," *Engineering Applications of Artificial Intelligence*, vol. 119, pp. 1–12, Mar. 2023, doi: 10.1016/j.engappai.2022.105799.
[34] M. H. Yazdani, M. M. Azad, S. Khalid, and H. S. Kim, "A Hybrid Deep Transfer Learning Framework for Delamination Identification in Composite Laminates," *Sensors*, vol. 25, no. 3, pp. 1–15, Jan. 2025, doi: 10.3390/s25030826.
[35] F. Naeimiasl, H. Vahidi, and N. Soheili, "Leveraging Principal Component Analysis for Data-Driven and Objective Weight Assignment in Spatial Decision-Making Framework for Qanat-Induced Subsidence Susceptibility Assessment in Railway Networks," *ISPRS International Journal of Geo-Information*, vol. 14, no. 5, pp. 1–21, May 2025, doi: 10.3390/ijgi14050195.
[36] Y. J. Lee, N. Y. Shin, and I. G. Lee, "Privacy-Preserving Data Sharing via PCA-Based Dimensionality Reduction in Non-IID Environments," *Electronics*, vol. 14, no. 13, pp. 1–14, Jul. 2025, doi: 10.3390/electronics14132711.

BIOGRAPHIES OF AUTHORS

Dava Aulia is currently pursuing a doctorate degree in Computer Science at Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. His research interests encompass artificial intelligence, deep learning, the internet of things (IoT), embedded systems, and gas sensor technologies. He can be contacted at email: 7025241016@student.its.ac.id.

Riyanarto Sarno received the Ph.D. degree, in 1992. He is currently a Professor with the Department of Informatics, Institut Teknologi Sepuluh Nopember (ITS). He has researched process mining for five years. He is interested in research projects on machine learning, the internet of things, knowledge engineering, enterprise computing, and information management. He is the author of more than five books and over 300 scientific articles, making him the world top 2% scientist in 2020 by Standford University. He can be contacted at email: riyanarto@its.ac.id.

Muhammad Rivai received the Ph.D. degree, in 2006. He is currently a professor of the Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. His research interests include sensors, chemical sensors, gas sensors, electronics, robotics, and artificial intelligence applications. He can be contacted at email: muhammad_rivai@its.ac.id.

Muhammad Amin is a lecturer, researcher, and respiratory specialist actively involved in education, research, and healthcare at the Faculty of Medicine, Airlangga University, and Airlangga University Hospital. He has been serving as a staff member in the Department of Pulmonology and Respiratory Medicine at the Faculty of Medicine, Airlangga University, since 1977. Various research projects led by him have been published in various national and international journals. His research encompasses topics such as asthma, COPD, lung infections, including COVID-19, and others. He has received several awards, including first place in an oral presentation competition organized by the Indonesian Society of Respirologists. He has authored numerous books, including those on COVID-19, tuberculosis, HIV, and medical emergencies. He can be contacted at email: amin.fkua@gmail.com.

Alfian Nur Rosyid is a lecturer, researcher, and respiratory specialist actively involved in education, research, and healthcare at the Faculty of Medicine, Airlangga University, and Airlangga University Hospital. He has been serving as a staff member in the Department of Pulmonology and Respiratory Medicine at the Faculty of Medicine, Airlangga University, since 2015. Various research projects led by him have been published in various national and international journals. His research encompasses topics such as asthma, COPD, lung infections, including COVID-19, and others. He has received several awards, including first place in an oral presentation competition organized by the Indonesian Society of Respirologists. He has authored numerous books, including those on tuberculosis, HIV, and medical emergencies. He can be contacted at email: alfian-n-r-10@fk.unair.ac.id.

Kelly Rossa Sungkono was born in Surabaya, in June 1994. She received a master's degree and doctoral's degree in computer engineering in 2016 and 2024, respectively. She is a lecturer at the Institut Teknologi Sepuluh Nopember (ITS). Her research interests include information systems, database management, and machine learning. She can be contacted at email: kelly@its.ac.id.