

<https://doi.org/10.1038/s41746-025-01851-4>

Artificially intelligent nasal perception for rapid sepsis diagnostics



Joonchul Shin^{1,2,10}, Gwang Su Kim^{3,10}, Seongmin Ha^{1,10}, Taehee Yoon^{4,10}, Junwoo Lee⁵, Taehoon Lee¹, Woong Heo¹, Kyungyeon Lee¹, Seong Jun Park⁶, Sunyoung Park^{1,6}, Jaewoo Song⁷, Sunghoon Hur³, Hyun-Cheol Song⁸, Ji-Soo Jang³✉, Jin-Sang Kim³✉, Hyo-Il Jung^{1,6,9}✉ & Chong-Yun Kang³✉

Sepsis, a life-threatening disease caused by infection, presents a major global health challenge due to its high morbidity and mortality rates. A rapid and precise diagnosis of sepsis is essential for better patient outcomes. However, conventional diagnostic methods, such as bacterial cultures, are time-consuming and can delay sepsis diagnosis. Considering these, researchers investigated alternative techniques that detect volatile organic compounds (VOCs) produced by bacteria. In this study, we designed colorimetric gas sensor arrays, which change color upon interaction with biomarkers, offer a direct visual signal, and demonstrate high sensitivity and specificity in detecting sepsis-related VOCs. Furthermore, an artificial intelligence (AI) based algorithm, Rapid Sepsis Boosting (RSBoost), was employed as an analytical technique to enhance diagnostic accuracy (96.2%) in blood sample. This approach significantly improves the speed and accuracy of sepsis diagnostics within 24 h, holding great potential for transforming clinical diagnostics, saving lives, and reducing healthcare costs.

Sepsis is a critical condition induced by infection, representing a major global health issue due to its high morbidity and mortality rates^{1–5}. Sepsis affects 148 per 100,000 people annually worldwide, can rapidly lead to organ failure, shock and death without prompt treatment, often caused by various bacteria, and provokes a systemic response^{1–3,6}. The primary bacterial species responsible for sepsis include gram-negative bacteria such as *Escherichia coli* (*E. coli*), *Klebsiella pneumoniae* (*K. pneumoniae*), and *Pseudomonas aeruginosa* (*P. aeruginosa*), as well as gram-positive bacteria like *Staphylococcus aureus* (*S. aureus*) and *Streptococcus pneumoniae* (*S. pneumoniae*)^{7,8}. Thus, rapid and accurate diagnostics for sepsis is essential for an effective and early treatment⁹. However, conventional culture-based diagnostic methods are labor-intensive and slow, often taking several days to yield results, which delays the administration of optimal antibiotics and increases the risk of overuse and resistance. Additionally, prolonged incubation periods can lead to high contamination rates, reducing the accuracy of bacterial identification^{10,11}. Thus, a diagnostic platform that resolves these limitations and delivers faster, more reliable, and accurate results is urgently needed.

To address these challenges, researchers have investigated alternative strategies that detect volatile organic compounds (VOCs), including

trimethylamine (TMA), ammonia (NH₃), and hydrogen sulfide (H₂S), emitted by bacteria. Among the myriad sensor technologies, colorimetric gas sensor arrays stand out in disease diagnostics due to their visual signal and ability to detect a wide range of gas concentrations, identifying complex mixtures of VOCs associated with infections^{12–14}.

Several studies have highlighted the potential of colorimetric gas sensor arrays in sepsis diagnosis^{15–17}. For instance, Jang et al. introduced a cellulose nanofiber pH indicator infused with red radish extract, which effectively monitored minced pork freshness, showing clear color transitions from red (fresh) to purple (spoiled), demonstrating its suitability for intelligent food packaging¹⁵. Sun et al. described the colorimetric paper-based band-aids for detecting and treating bacterial infections. These band-aids change color to indicate infection (yellow) or drug resistance (red) and use antibiotics or photodynamic therapy accordingly. The detection limit of the sensors is 10⁴ colony-forming units (CFU)/ml for drug-resistant *E. coli*¹⁶. Furthermore, Chen et al. successfully identified four spoilage bacteria strains using a colorimetric sensor array. The sensor detected specific VOCs, achieving a 100% classification accuracy with linear discriminant analysis and confirming genetic relationships through hierarchical cluster analysis¹⁷.

¹School of Mechanical Engineering, Yonsei University, Seoul, Republic of Korea. ²Advanced Bio and Healthcare Materials Research Division, Korea Institute of Materials Science (KIMS), Changwon, Republic of Korea. ³Electronic Materials Research Center, Korea Institute of Science and Technology (KIST), Seoul, Republic of Korea. ⁴George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ⁵Department of Molecular Science and Technology, Ajou University, Suwon, Republic of Korea. ⁶The Daborn Inc, Seoul, Republic of Korea. ⁷Department of Laboratory Medicine, College of Medicine, Yonsei University, Seoul, Republic of Korea. ⁸Materials Science and Engineering, Korea University, Seoul, Republic of Korea. ⁹Department of Integrated Medicine, Yonsei University, Yonsei University, Seoul, Republic of Korea. ¹⁰These authors contributed equally: Joonchul Shin, Gwang Su Kim, Seongmin Ha, Taehee Yoon. ✉e-mail: wkdwt92@kist.re.kr; jskim@kist.re.kr; uridle7@yonsei.ac.kr; cykang@kist.re.kr

However, in the early stages of sepsis, bacterial concentrations are notably low, typically ranging from 1 to 10 CFU/ml, making early diagnostics challenging^{18,19}. In this study, we have synthesized pH indicators and poly ionic liquids (PILs) to enhance the sensitivity and dynamic range of sensors. PILs are a unique class of polymers that integrate the innovative features of ionic liquids with enhanced mechanical robustness and dimensional stability post-polymerization. One of the remarkable attributes of PILs is their superior ion-exchange capability, allowing the synthesis of PILs with diverse counteranions by polymerizing a single ionic liquid monomer and performing anion-exchange reactions²⁰. These results enable the extraction of multi-class characteristic equations for classification learning and assignment of weights to effective indices using machine learning²¹. Thus, machine learning has also been employed to interpret correlations within large-scale datasets, showing promising accuracy.

Artificial intelligence (AI) technology has emerged as a transformative tool in sensor applications, healthcare, food logistics, and environmental monitoring^{22–26}. Machine learning-based hybrid algorithms for sensor signal classification²⁷, particularly for array sensors that generate complex and overlapping signals, have garnered significant attention. These algorithms, such as Convolutional Neural Networks (CNNs) for multi-class classification and ensemble models for multi-substance concentration prediction, are becoming innovative tools for large-scale, multi-class data analysis. CNN-based models excel in extracting morphological features from complex blood cell image data, achieving high classification accuracy with minimal preprocessing²⁸. However, increasing the number of training classes leads to longer training times and potential biases in classification outcomes.

Conversely, ensemble-based models analyze the correlation structure of data from multiple substances, deriving latent characteristic equations to predict concentrations with a 6% error rate through regression learning. Despite their high accuracy rates, these models require large datasets and labor-intensive preprocessing, resulting in extended training times²⁹. Hybrid algorithms that integrate the strengths of existing models are essential for analyzing large, complex, multi-class datasets³⁰. These algorithms are precious for early sepsis diagnosis, where timely and accurate detection is crucial for effective treatment³¹. In machine learning-based regression models combined with colorimetric sensors, the number of sensors and independent variables significantly impacts the prediction performance for multi-class data types. Models with fewer independent variables may exhibit improved prediction performance but can become overly complex. Conversely, using many independent variables increases model complexity and prediction performance but at the cost of increased training time³².

Herein, we introduce a novel approach to sepsis diagnostics by integrating an advanced colorimetric gas sensor array with a sophisticated machine-learning model. We then employed the Rapid Sepsis Boosting (RSBoost) combining an advanced colorimetric gas sensor array with a sophisticated machine learning-based hybrid algorithm. The RSBoost hybrid algorithm combines the multi-class classification performance of Convolutional neural networks-support vector machine (CNN-SVM) with the multi-substance concentration prediction capability of least square boosting (LSBoost), enabling simultaneous high-precision bacterial classification and concentration-based regression. This novel approach demonstrates excellent sensitivity and precision in detecting VOCs emitted by bacteria in septic patients. The RSBoost predicts the proportion of unknown bacterial species with an error rate of less than 3.8%, surpassing existing hybrid algorithms in accuracy and analysis speed. This method enhances the robustness and reliability of sepsis detection, ensuring timely and accurate diagnosis and treatment. Thus, integrating advanced sensor technologies with AI-driven algorithms holds great promise for improving sepsis diagnostics. The contributions of this paper are summarized as follows:

The proposed algorithm (RSBoost, Rapid Sepsis Boost) comprises CNN-SVM and LSBoost layers, enabling simultaneous classification and regression training across different data types. The CNN-SVM layer extracts 3 morphological pattern characteristics of colorimetric sensors to achieve high-accuracy multi-class classification, while the LSBoost

layer introduces a feature equation extraction and automated weight allocation based on indexing for multi-class concentration prediction.

The algorithm's hidden layer employs parallel learning, which can improve learning speed and efficiency by providing a 4-fold training on datasets.

The learning results have been evaluated through Monte Carlo Cross Validation (MCCV = 100), correlation matrix analysis, and Pearson correlation coefficients verification.

Results

Gas sensing performance of sensor array

This study presents a novel diagnostic method for sepsis by analyzing volatile organic compounds (VOCs) produced by bacteria. Figure 1a illustrates the two key features: Nasal perception for bacterial gas and Digital diagnostics for sepsis. The gas sensor array comprises a grid of reactive materials, including PILs and pH indicators (Bromophenol Blue as BPB, Tetraiodophenolsulfonephthalein as TET, and Cu-PAN), which selectively interact with certain VOCs to boost the sensitivity. The PILs chemically interact with VOCs, even at low concentrations, inducing color alterations in the array.

The pH indicators embedded in PILs change color reversibly in response to acidity shifts triggered by VOC binding, providing a dynamic visual signal of bacterial presence. The ground truth data is employed to enhance the identification and concentration prediction accuracy of bacteria critical for sepsis diagnostics. Figure 1b shows the RSBoost hybrid algorithm processes the input data (color intensities) through two specialized layers: the CNN-SVM layer for large-scale multi-class image classification and the LSBoost layer for multi-class substance concentration prediction^{28,29}. Each layer utilizes a latent feature extraction technique for classification and regression learning to identify color and morphological features. The extracted features are then processed by the algorithm's output layer, which uses the flatten and fully connected layers to predict the concentration ratios of 3 bacteria. These predictions can be subsequently provided to clinicians for early sepsis diagnostics. The sensing mechanism, chemical surface, structures and compositions of the PIL, BPB-PIL, and TET-PIL were analyzed using x-ray photoelectron spectroscopy (XPS), Ultraviolet-visible (UV) spectroscopy, and ¹H nuclear magnetic resonance (NMR) spectroscopy (Supplementary Note 1–3, Supplementary Figs. 1–5).

The colorimetric gas sensor array was fabricated using a plastic mold and pressed sugar, then soaked into each sensing solution in Fig. 2a. To ensure batch-to-batch consistency, all sensor arrays were fabricated under identical conditions. We performed cross-batch quality checks on randomly selected sensors: XPS confirmed consistent elemental compositions and ¹H-NMR analysis verified that the chemical structure of the PIL-dye complexes was uniform across batches (Supplementary Note 1–3, Supplementary Figs. 1–5). Subsequently, a smartphone-based analysis system was employed to quantify the sensors' color change in real time. The gas-sensing performance of the colorimetric sensor array was evaluated within a customized chamber, as illustrated in Supplementary Fig. 6. A Mass Flow Controller (MFC) precisely regulated the gas flow rate, maintaining a consistent rate of 1000 sccm. The target gas was mixed and calibrated at room temperature (RT) and relative humidity (RH) of 60%, replicating realistic conditions^{33,34}. The color changes of the sensors were recorded using a Samsung Galaxy S7 smartphone. For real-time analysis, the recorded videos were transferred to a laptop, where the RGB values were converted to brightness (V) values. To prevent interference from ambient contaminants, the gas-exposure chamber was purged with inert gas and sealed before introducing the target analytes, ensuring that sensor color change arises solely from the target VOCs. Additionally, the smartphone camera was fixed at a consistent position with uniform focus, exposure, and white balance under an LED-lit closed imaging box, minimizing variability due to external lighting or angle.

The colorimetric sensor arrays comprise BPB-PIL sensors for NH₃ detection, TET-PIL sensors for TMA detection, and Cu-PAN sensors for

H₂S detection. The sensors were exposed to target gas concentrations ranging from 0.2 to 3 ppm, with color changes captured in real time and converted to brightness values. The sensor response was defined in Eq. (1): V_{air} and V_{gas} represent the brightness values in the absence and presence of the target gas, respectively.

$$\text{Response}(\%) = \left(\frac{\Delta V}{V_{\text{(gas or air)}}} - 1 \right) \times 100 \quad (1)$$

In Fig. 2b–d, the application of PIL significantly enhances the responses of BPB and TET sensors. The BPB-PIL demonstrates a response 5.1 times greater than BPB when exposed to 1 ppm NH₃, while TET-PIL shows a

response 2.5 times higher than TET for 1 ppm TMA. Furthermore, the Cu-PAN sensor exhibits a fast response to H₂S up to 1 ppm, followed by gradual saturation up to 3 ppm (Supplementary Figs. 7, 8). Response linearity is an important factor that provides high reliability in computational diagnostic monitoring. The response linearity was evaluated for the target gas concentration between 0.2 and 1 ppm, which is the low concentration range. The BPB-PIL sensor, TET-PIL sensor, and Cu-PAN sensor exhibited R² values of 0.939, 0.945, and 0.96 for NH₃, TMA, and H₂S, respectively. The theoretical detection limits (signal-to-noise ratio >3) were 49.4 ppb for BPB-PIL, 21.2 ppb for TET-PIL, and 173 ppb for Cu-PAN.

Additionally, the selective response of the colorimetric sensor to target gases was confirmed by introducing other reactive gases (1 ppm

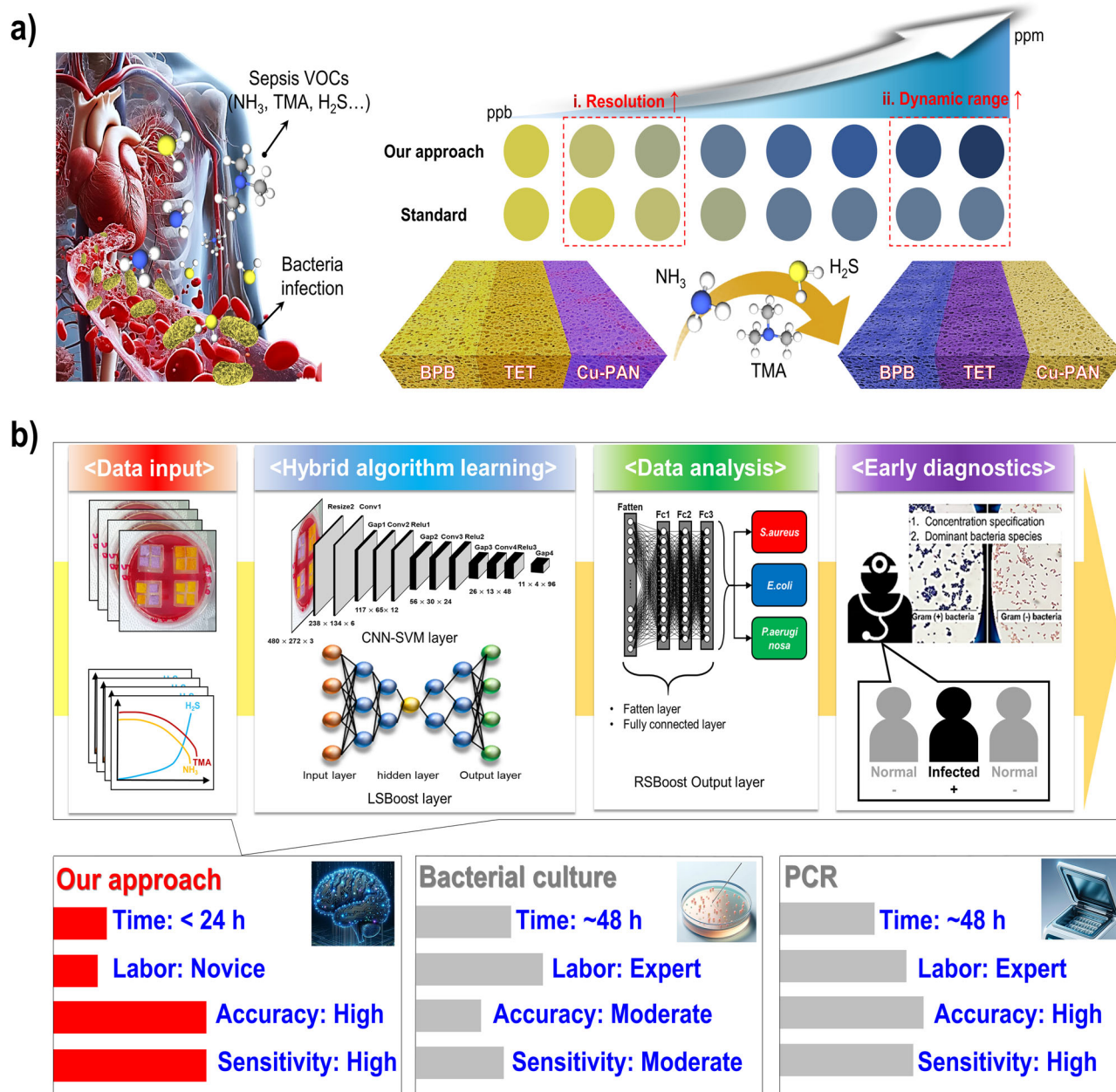


Fig. 1 | The schematic diagram illustrates an advanced approach to sepsis diagnostics through bacterial gas detection and digital analysis. **a** Nasal perception for bacterial gas, where volatile organic compounds (VOCs) associated with sepsis, such as NH₃, TMA, and H₂S, are detected using a gas sensor array (BPB-PIL, TET-PIL, Cu-PAN). This approach surpasses normal pH indicators in resolution and dynamic range, enabling precise detection in parts per billion (ppb). **b** Digital diagnostics for

sepsis employs a hybrid algorithm that integrates CNN-SVM and LSBoost layers for data processing and early detection of bacterial species like *E. coli*, *P. aeruginosa*, and *S. aureus*. The diagnostic platform offers high accuracy and sensitivity within 24 h, requiring only beginners, compared to conventional bacterial culture and PCR methods, which take up to 48 h and need experts.

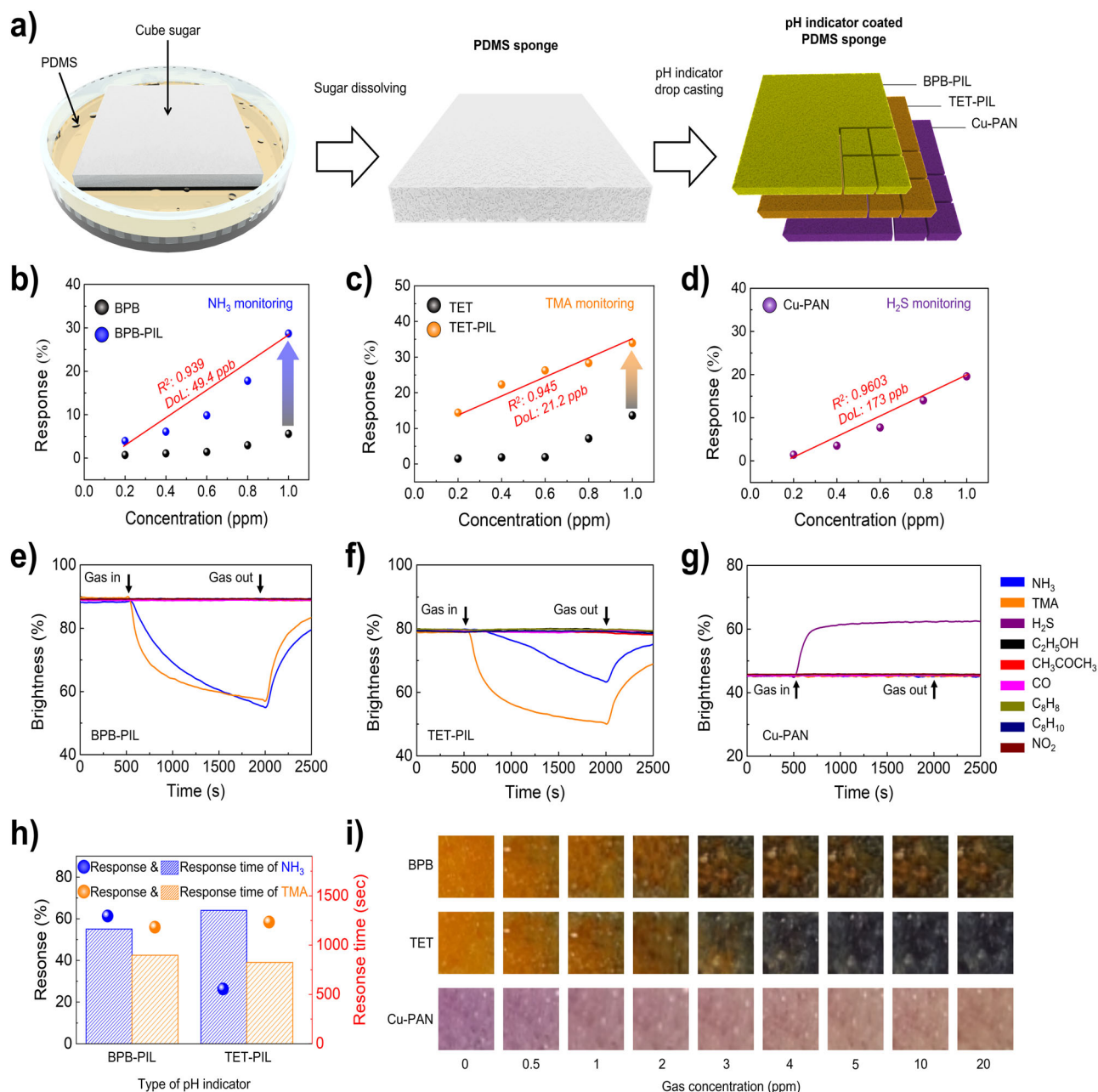


Fig. 2 | The pH-indicator-coated PDMS sponges enable selective, linear, real-time colorimetric detection of NH_3 , TMA, and H_2S . **a** The figure presents the fabrication of pH indicator-coated PDMS sponges designed for selective gas detection. The process begins with sugar dissolution to create porous PDMS sponges, followed by applying three pH indicators (BPB-PIL, TET-PIL, Cu-PAN). **b–d** Response curves demonstrate the sensors' sensitivity to gases such as NH_3 , TMA, and H_2S , with linear correlations (R^2) supporting their detection accuracy. **e–h** Time-dependent

brightness measurements reveal real-time sensor responses during gas introduction and removal. Simultaneously, comparing response times highlights the distinct sensitivities of BPB-PIL to NH_3 and TET-PIL to TMA (The color legend assigns blue to NH_3 , yellow to TMA, purple to H_2S , dark green to $\text{C}_2\text{H}_5\text{OH}$, red to CH_3COCH_3 , pink to CO, turtle to C_8H_8 , dark blue to C_8H_{10} , and brown to NO_2). **i** Color intensities change across various gas concentrations, demonstrating their effectiveness for gas detection.

NH_3 , TMA, H_2S , and 5 ppm $\text{C}_2\text{H}_5\text{OH}$, CH_3COCH_3 , CO, C_8H_8 , C_8H_{10} , NO_2) in the air (Fig. 2e–g). The Cu-PAN sensor exhibited high reactivity solely to H_2S gas and did not react with other gases. BPB-PIL and TET-PIL sensors revealed reactivity to both NH_3 and TMA gases. As shown in Fig. 2h, BPB and TET exhibited similar response of 59 and response time of 800 s for TMA gas. Conversely, for NH_3 gas, the BPB-PIL sensor demonstrated 2.7 times higher response and 1.2 times faster response times compared to the TET-PIL sensor. In this regard, we noted that the pH-indicator sensors (BPB-PIL and TET-PIL) exhibit reversible color changes upon exposure to their target basic gases (NH_3 or TMA). Once the basic gas is removed and the local pH returns to neutral, these sensors

gradually return to their original color. In contrast, the Cu-PAN sensor's color change is chemically irreversible, because H_2S reduces Cu^{2+} to an insoluble copper sulfide (CuS) precipitate that does not revert to the original complex. We emphasize that this makes the Cu-PAN element single-use, whereas the BPB-PIL and TET-PIL sensors can be reused. These performance disparities were evident in the plotted response percentages and reaction times. Subsequent bacterial culture experiments and machine learning analysis confirmed the classification capabilities of these sensors, indicating that their differential cross-reactivity can be effectively utilized for the accurate detection and differentiation of NH_3 and TMA.

Figure 2i illustrates the color change of the actual BPB-PIL, TET-PIL, and Cu-PAN sensors when the main target gases, NH_3 , TMA, and H_2S , are introduced in a broad concentration range of 0–20 ppm. The colorimetric sensor exhibited a detectable change at a low concentration of 1 ppm or less and stabilized after an acceptable change at 5 ppm or more. The exceptional sensor that can observe the reaction within 5 min, even to ppb unit gas, was fabricated based on the limit-of-detection (LOD).

Nasal perception towards VOCs in bacteria species

The gas sensor array, composed of BPB-PIL, TET-PIL, and Cu-PAN sensors, demonstrated its precision in recording the color change, as depicted in Fig. 3a. A comprehensive set of one hundred datasets was meticulously

collected for each bacterial concentration (*E. coli*, *P. aeruginosa*, and *S. aureus*) spanning from 10^1 to 10^3 CFU/mL. Figure 3b vividly represents the gas sensor array's response to *E. coli* emissions. Each sensor showcased distinct reaction rates, onset times for color change, and sensitivity at different bacterial concentrations. In short, “onset time” is now defined as the elapsed time from gas introduction to the first detectable response of the sensor (when the brightness change exceeds noise), whereas “reaction rate” is the rate of change of the sensor response ($\Delta V/\%$ per minute) during the active response period. Notably, the TET-PIL sensor exhibited a faster response than the BPB-PIL sensor, and the gas response was amplified with higher bacterial concentrations. This trend was consistently observed for both BPB-PIL and TET-PIL sensors. The response to *P. aeruginosa*

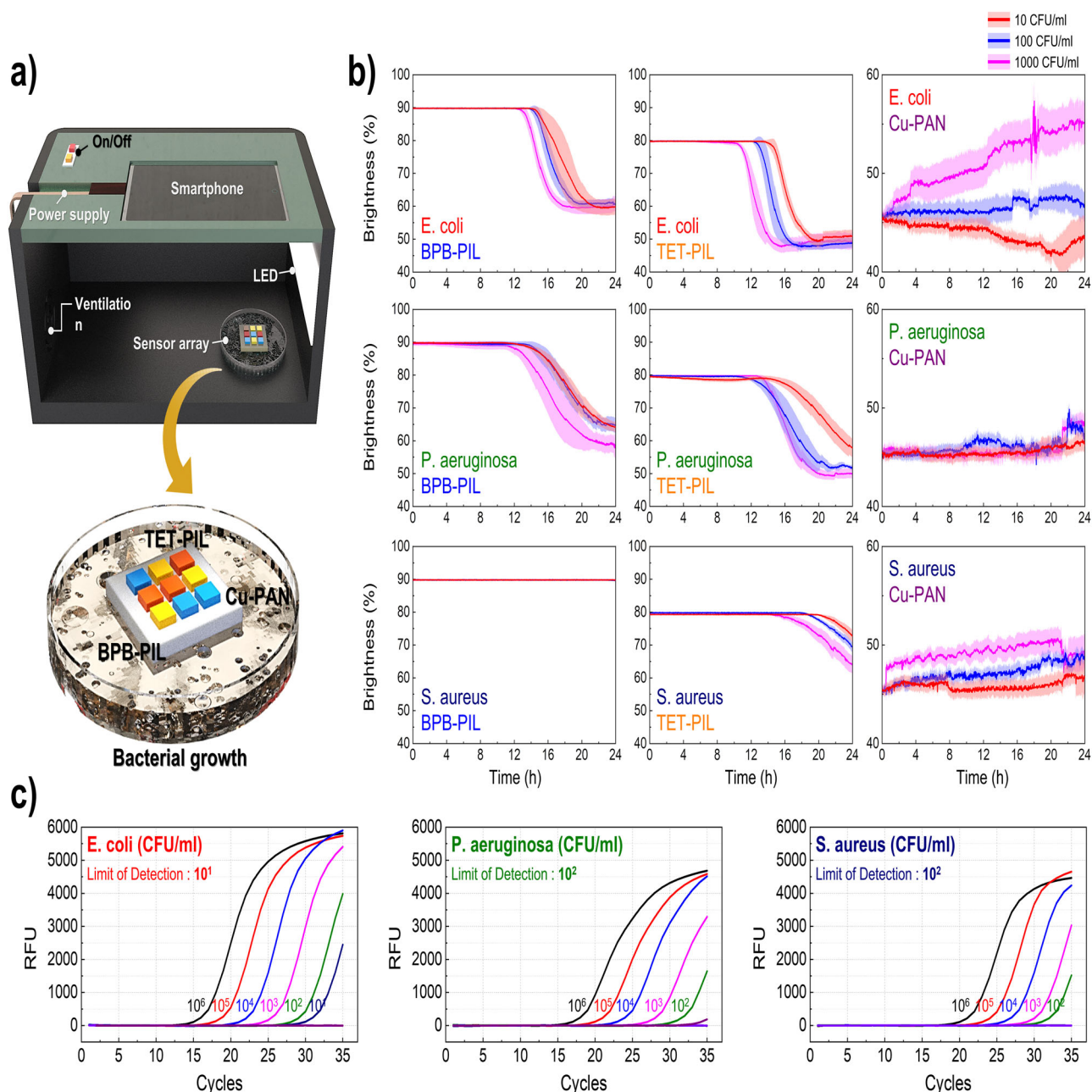


Fig. 3 | A smartphone-based pH-indicator gas sensor array monitors brightness changes to detect *E. coli* at 10^1 CFU/mL and *P. aeruginosa*/*S. aureus* at 10^2 CFU/mL in real time. a The figure demonstrates a smartphone-based gas sensor array system for real-time detection and quantification of bacterial growth. The system features a smartphone cradle, a dedicated power supply, an on/off switch for LED illumination control, an integrated gas sensor array plate, and a built-in ventilation

module to ensure proper airflow. **b** The gas emissions from three bacterial species (*E. coli*, *P. aeruginosa*, and *S. aureus*) cause a time-dependent decrease in brightness across BPB-PIL, TET-PIL, and Cu-PAN pH indicator systems, with 10^1 CFU/mL (Orange), 10^2 CFU/mL (blue), and 10^3 CFU/mL (pink). **c** A real-time PCR evaluates detection limits of 10^1 CFU/mL for *E. coli* (red), 10^2 CFU/mL for *P. aeruginosa* (green), and 10^2 CFU/mL for *S. aureus* (navy blue).

emissions is presented. The gas sensor array's reaction time and response slope varied precisely according to the bacterial concentration. Similar variations were observed with *S. aureus* emissions. Among 3 sensors, the BPB-PIL sensor exhibited a low response to *S. aureus* emissions, while the TET-PIL sensor showed a slightly higher response. The Cu-PAN sensor had a delayed and relatively low response to *S. aureus* emissions. This discrepancy can be attributed to *S. aureus*'s slower growth rate and lower VOC emission concentration.

Based on these results, the LOD for each gas sensor in identifying bacterial species was determined to be 10 CFU/ml (Fig. 3c). A real-time PCR (polymerase chain reaction) was employed to compare the sensitivity of the gas sensor array with the conventional detection method. The evaluations were conducted on each strain in triplicates at varying concentrations ranging from 10^6 to 1 CFU/ml to assess the analytical sensitivity and specificity of bacteria. The PCR-based identification of bacteria was conducted through 30 cycles, focusing on representative gram-negative bacteria (*E. coli* and *P. aeruginosa*) and gram-positive bacteria (*S. aureus*). The standard curves of *E. coli*, *P. aeruginosa*, and *S. aureus* demonstrated the linearity of each dilution, establishing the limit of detection at 10^2 , 10^3 , and 10^3 CFU/ml, respectively. The bacterial concentrations of 1 and 10 CFU/ml were not detectable with PCR. The LOD was 10^2 , 10^3 , and 10^3 CFU/ml for *E. coli*, *P. aeruginosa*, and *S. aureus*. No cross-reactivity was observed in each DNA at a high bacterial concentration of 10^6 CFU/ml. All bacteria types were detected at a minimum concentration of 10^3 CFU/ml. Based on the results, we noted that *E. coli* tends to produce a higher amount of TMA relative to H_2S and NH_3 , thus the TET-PIL sensor (for amines) showed the strongest change for *E. coli* species. *S. aureus*, conversely, emits more NH_3 and few TMA (it lacks the pathways to produce TMA in significant quantity), so the BPB-PIL sensor (for NH_3) dominated its response, with minimal change in TET-PIL. *P. aeruginosa* produces notable H_2S and NH_3 . Therefore, the Cu-PAN sensor (for H_2S) and to some extent BPB-PIL responded for *P. aeruginosa*, while TET-PIL showed little change.

These findings have significant implications for bacterial detection and monitoring. With its excellent sensitivity and real-time results, the gas sensor array offers a practical and efficient alternative to the complex and time-consuming PCR-based identification method. This could revolutionize bacterial detection and monitoring in various applications, including clinical diagnostics and environmental monitoring.

Pattern recognition for sepsis diagnostics

Our research introduces an advanced hybrid algorithm designed to improve the speed and accuracy of sepsis diagnosis. This algorithm leverages the RSBoost algorithm to perform regression and classification tasks, aiming to predict and classify bacterial compositions detected by a gas sensor array. This facilitates prompt and precise sepsis diagnosis and the selection of appropriate antimicrobial treatments. Identifying the dominant bacterial species responsible for sepsis is crucial for enhancing treatment precision and minimizing antimicrobial resistance.

The RSboost algorithm, a novel approach for rapid sepsis diagnosis, stands out for its unique integration of classification and regression learning. It utilizes heterogeneous data (24-h imaging and intensity data) as input, categorizing them based on gas concentration values emitted by three bacterial species. The first dataset contains *S. aureus*; *E. coli*; *P. aeruginosa* at a ratio of (0:0:10) with a concentration of 1000 CFU/ml, measuring TMA, H_2S , and NH_3 gas, with dimensions [10×89959]. The second dataset, with a ratio of (0:10:0) at the same concentration, has dimensions [10×89840]. The third dataset, with a ratio of (10:0:0) at 1000 CFU/ml, measures the same gases with dimensions [21×84305].

These datasets (70% as training and 30% as blind test) train an ensemble-based Least Squares Boosting (LSBoost) layer for regression. For each bacterial concentration class, Root Mean Square Error (RMSE) and Pearson Correlation Coefficient (R^2)³⁵ are calculated to derive indicators defining inter-gas relationships. These indicators are ranked to assign weights, resulting in characteristic equations. The training efficiency is quadrupled by conducting 4 instances of partial parallel training (L_1 to L_4) in

a single process. The correlation between TMA, H_2S , and NH_3 concentration classes emitted by each species is evaluated using a correlation matrix, with indices ranging from -1 to 1 ³⁶. In the classification learning stage, a key component is the use of a CNN-SVM layer. This layer, specialized for multi-class classification, plays a significant role in the algorithm's ability to accurately classify the 3 bacterial species.

Bacterial concentration images recorded over 24 h were split frame-by-frame for classification learning. Morphological features were extracted from each image using three methods (length scale, shape factor, RGB index)²⁸, resulting in distinct feature vectors. The CNN-SVM layer employs a CNN with a 7×7 kernel filter to extract spatial features. Hyperparameters are set as follows: input size (480×272), batch size (32), learning rate (0.01), epoch (1000), and iteration (500).

We set the input resolution to 480×272 pixels (with a superpixel segmentation density of 300) to preserve key spatial features of the sensor's colorimetric response while keeping computational cost low. This resolution was found to capture the relevant morphology in the sensor images without unnecessary detail.

Using mini-batches of 32 stabilized the training convergence and allowed the model to capture spatial variability across sensor images in each update. A smaller batch size risked noisy updates, whereas much larger batches did not improve accuracy.

We chose a moderate learning rate of 0.01 to ensure stable and efficient gradient descent. This value was high enough to allow reasonably fast convergence but low enough to avoid overshooting minimal values during backpropagation.

The algorithm was trained for 1000 epochs to provide sufficient iterations for learning diverse spatial and temporal patterns, consistently resulting in high classification accuracy.

The RSBoost algorithm was configured with 500 boosting iterations to enhance classifier diversity, reduce classification bias, and avoid redundancy or unnecessary computation.

This process extracts latent features, generates feature maps, and transforms them into activation maps. The activation map is classified into multiple classes using an SVM classifier integrated into the CNN model. The three bacterial species are classified based on spiked blood samples using a blind test dataset. RSBoost validates the blind test classification results using a confusion matrix and Area under the curve (AUC) values. The algorithm's regression and classification results are verified through MCCV³⁷, repeated 100 times. These results are visualized in a 3-dimensional scatter plot, enabling trend analysis of accuracy across MCCV iterations (Detailed information for the algorithm learning process is shown in Supplementary Fig. 9).

Data preprocessing and feature extraction

The hybrid algorithm analyzes associations among diverse data collected via enhanced gas sensor, employing classification and regression learning based on gas concentration ratios. In Fig. 4a, image data from a colorimetric sensor over 24 h are preprocessed into numerical datasets, such as intensity values, enhancing analytical accuracy and reducing training time. The RSBoost algorithm's input layer employs 3 image analysis tools for preprocessing classification learning: (1) The superpixel refers to an image segmentation technique used to extract spatial coordinates representing object-shaped regions within the image³⁸. In practice, we applied MATLAB's Superpixel segmentation (SLIC) to segment each 24-h sensor image into ~200–350 regions per sensor spot based on RGB intensity variations, and the averaged RGB values of each region were used to represent the spot's response. This enabled real-time spatial vector extraction that facilitated class discrimination. (2) Region of Interest (ROI) masks identify significant RGB value variations over time, defining RGB index ranges for different categories³⁹. (3) grayscale conversion and binary segmentation extract morphological features for classification learning⁴⁰. For regression learning, preprocessing involves identifying unique indexings from intensity values corresponding to distinct bacterial types. This defines the latent attributes of the data,

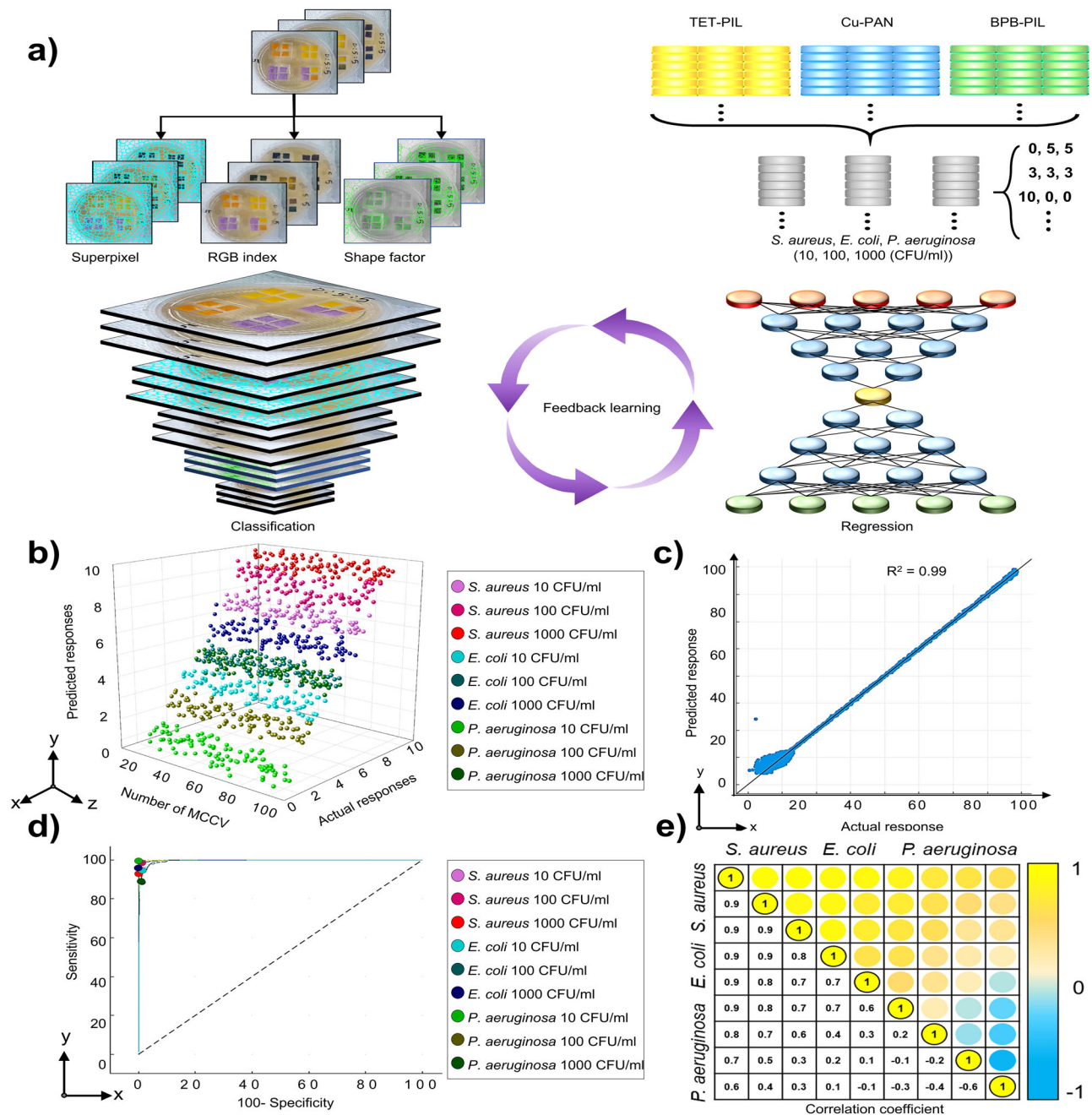


Fig. 4 | The RSBoost hybrid model utilizes CNN-SVM feature extraction with LSBoost regression on heterogeneous datasets to achieve robust classification and regression. a The figure illustrates the dynamic structure of the RSBoost hybrid algorithm, integrating with CNN-SVM and LSBoost layers. The CNN-SVM layer, tailored for classification, computes length scale values to extract 3 key morphological features. It employs a superpixel tool to analyze sensor color distributions, shape factors to track color area changes over time, and the RGB index to identify color variations. The LSBoost layer, designed for regression learning, is trained on RGB intensity datasets from 3 bacterial types based on varying concentration combinations (CFU/ml). This process is enhanced by parallel learning, feature equation extraction, weight allocation, and iterative feedback training, showcasing

the algorithm's adaptability. **b** A 3D scatter plot of nine classes (*S. aureus* at 10^1 CFU/mL in purple; *S. aureus* at 10^2 CFU/mL in pink; *S. aureus* at 10^3 CFU/mL in red; *E. coli* at 10^1 CFU/mL in light blue; *E. coli* at 10^2 CFU/mL in navy; *E. coli* at 10^3 CFU/mL in blue; *P. aeruginosa* at 10^1 CFU/mL in lime; *P. aeruginosa* at 10^2 CFU/mL in turtle; and *P. aeruginosa* at 10^3 CFU/mL in darkest green), evaluated using Monte Carlo cross-validation (MCCV), compares actual versus predicted values. **c** A ROC curve illustrates the classification results. **d** An R-plot compares actual and predicted RGB intensity values for nine classes (Applied the same color legend to all nine classes mentioned in b). **e** In the Pearson correlation matrix, correlation coefficients near +1 (in yellow) indicate minimal latent features, whereas coefficients near -1 (in sky-blue) denote maximal potential for latent feature extraction.

varying with gas combination ratios, and assigns weights to the most valid indexings, enhancing reliability and minimizing prediction error rates. The CNN-SVM hybrid algorithm performs classification learning in the hidden layer, capitalizing on its strengths in large-scale image data analysis and multi-class classification²⁸. Concurrently, the LSBoost layer manages regression learning, utilizing automatic property equation extraction and

weight allocation for concentration categories²⁹. (Detailed hybrid algorithm is shown in Algorithm 1).

The hybrid algorithm employed superpixel and RGB index toolboxes to extract key features related to sensor responsiveness across various bacterial concentrations (10, 100, 1000 CFU/ml) (Supplementary Fig. 10). Three morphological characteristic techniques were employed using a

CNN-SVM layer to enhance classification accuracy. Each technique revealed unique vector value ranges that increased with higher bacterial concentrations. Integrating 3 feature extraction models and applying them to a CNN-SVM layer significantly improves feature extraction, leading to more accurate image classification than using each algorithm individually. Thus, we incorporated these morphological extraction methods into the RSBoost hybrid algorithm.

Specifically, the length scale (L) for each bacterium increased from 10 to 78, while the RGB index ranged from 52 to 220. Depending on the bacterial concentration, the shape factor demonstrated distinct vector values, varying from 14 to 98. Notably, BPB-PIL bacteria exhibited the highest L values, ranging from 16 to 80, with an RGB index between 82 and 158, representing the lowest range. The shape factor values for this class varied from 16 to 96. These findings highlight the crucial role of morphological characteristic techniques in accurately identifying and classifying bacteria based on their concentrations (Supplementary Figs. 11–13).

The output layer monitors predictive learning outcomes through feedback, as shown in Fig. 4b. A 3-dimensional scatterplot illustrates classification results using Monte Carlo cross-validation (MCCV) to validate learning outcomes³⁷. The RSBoost algorithm has demonstrated unbiased and overfitting-free classification accuracy across all categories. Further evaluation in Fig. 4c examines learning performance using AUC values of ROC (Receiver operating characteristic) curves, ranging between 0.97 and 0.99, indicating high classification proficiency. Figure 4d presents Pearson correlation coefficient (R^2) analysis for regression reliability, achieving an R^2 value of 0.99 across all blind test outcomes, indicating high predictive accuracy. And Fig. 4e displays the correlation matrix results for regression-trained multiple-concentration classes, with correlation indices ranging from -1 to 1 , indicating feature correlation strength³⁶. These matrices help identify and weight the most valid values, reducing regression learning error rates (Refer to Supplementary Table 1 for detailed Pearson correlation coefficient analysis metrics and the complete correlation matrix for all concentration classes).

Figure 5 presents an analysis of the classification and regression learning outcomes from training and blind testing using the RSBoost hybrid algorithm. Figure 5a shows the confusion matrix for classification based on the training set. The average accuracy per bacterial type for each gas-specific concentration combination was 97.1%, with classification accuracies of 98.8% for *S. aureus*, 96.9% for *E. coli*, and 95.6% for *P. aeruginosa*. The highest accuracy was observed for *S. aureus* and the lowest for *P. aeruginosa*, attributed to overlapping eigenvector values with other bacteria. Figure 5b depicts a 3D scatter plot of classification results using MCCV, indicating consistent accuracy across all bacterial classes and confirming the RSBoost algorithm's reliability. Figure 5c presents the ROC curve plots for each bacterial class, with AUC values of 0.99 for *S. aureus*, 0.98 for *E. coli*, and 0.97 for *P. aeruginosa*. These high AUC values demonstrate the algorithm's strong classification accuracy and confidence.

Figure 5d–f demonstrate the adaptability of the RSBoost algorithm to real-world scenarios, as seen in the classification results from blind testing using blood sample data. Supplementary Fig. 14 represents a stable color intensity of the sensor array with negligible changes over time in blood sample from healthy donors. It reveals that the RSBoost algorithm exclusively employs the datasets representing color changes caused by bacterial VOCs in blood. Figure 5d presents the confusion matrix, with classification accuracies of 97.4% for *S. aureus*, 96.8% for *E. coli*, and 94.5% for *P. aeruginosa*. These results underscore the algorithm's compatibility and high usability in real blood-based diagnostics. Figure 5e features a 3D scatter plot of blind test results validated with 100 MCCV trials. It shows regions with similar accuracy to the training set, confirming the algorithm's reliable classification performance. Figure 5f reveals ROC curve plots for the blind test, with AUC values of 0.99 for *S. aureus*, 0.98 for *E. coli*, and 0.98 for *P. aeruginosa*. These values affirm the high reliability and accuracy of the RSBoost algorithm under randomized conditions. Figure 5g compares the RSBoost algorithm against other classification models using the same training and blind testing conditions. The RSBoost algorithm outperformed

all models, achieving a classification accuracy of 96.2%, compared to 79.8% for the convolutional neural network (CNN) model. This highlights the RSBoost algorithm's potential for high-accuracy early sepsis diagnostics, a crucial advancement in bacterial classification. Additionally, we compare RSBoost against various machine learning models in terms of accuracy, AUC, sensitivity, specificity, training time, inference time, and memory usage in the Supplementary Table 2. Moreover, Fig. 5h compares the total training times of the RSBoost algorithm and other classification models under identical training and blind testing conditions using a large heterogeneous dataset. The RSBoost algorithm completed training in 96 h, 143 h less than the 239 h required by the CNN model. This significant reduction in training time highlights the potential for early diagnosis by accurately predicting the optimal combination of sepsis-causing bacteria proportions in large, complex datasets. Additionally, an ablation study is conducted to evaluate whether the RSBoost model leads to synergistic improvement or introduces redundancy in the supplementary material (see Supplementary Fig. 15). The results from the ablation study were analyzed using the p-value to maintain a more robust confidence interval. To validate the statistical significance of RSBoost's performance gains, we conducted paired t-tests over 100 bootstrap resampling runs. As a result, p-values were <0.01 , confirming that RSBoost's improvements are statistically significant and reflect a higher confidence level in its performance compared to baseline models.

Discussion

In this study, we incorporate a highly sensitive sensing system and hybrid algorithm to overcome the bottlenecks of the following previous studies. Convolutional neural network (CNN) and Boost model are the most common methods for image segmentation and prediction for accurate diagnostics. For instance, Bukkapatnam et al. employs a convolutional neural network (CNN) to detect sepsis from blood smears. The CNN model is trained on augmented data to classify samples into “sepsis” or “not sepsis.” Despite promising initial accuracy, the model needs to be more balanced, mainly due to the limited dataset size, which impairs overfitting⁴¹. Yuan et al. presented an AI algorithm using XGBoost for early sepsis diagnosis in the ICU, using electronic medical record (EMR) data. The model was trained on 1588 instances with an accuracy of 82%, sensitivity of 65%, and AUROC of 0.89. However, the limitations were found as potential overfitting due to data from a single institution and missing dynamic features⁴². Zhao et al. showed ICU data to apply LightGBM algorithms to predict early sepsis in 6 h. The feature generation method, which used statistical and medical features, outperformed the mean processing method, achieving an AUC of 0.979 with LightGBM, indicating strong prediction capabilities. However, there were challenges, including high data imbalance, missing data, and the difficulty of clinical interpretation⁴³. Random Forest was employed to predict sepsis and ICU admission based solely on complete blood count (CBC) data. The model achieved an AUROC of 0.872 internally and performed well in external validations (AUROC: 0.805–0.845). Procalcitonin (PCT) was added to the model to improve prediction accuracy (AUROC: 0.857). Otherwise, ICD-coded diagnoses underestimated sepsis cases and the challenge of clinical applicability in non-ICU settings⁴⁴.

Moreover, conventional classification and regression methods require different models depending on the data type. Such models commonly employ transfer learning, which limits the extraction of potential features from a myriad of datasets⁴⁵. Considering these, a hybrid learning algorithm using both classification and regression methods has been proposed. This approach can effectively learn from heterogeneous datasets with similar data types, addressing the constraints of traditional methods⁴⁶. For instance, Marek Tatarko et al. introduced a hybrid SVM classification and regression algorithm that employed a multiple-class frequency shift dataset to identify the similar effects of trypsin and plasmin on k-casein. The hybrid algorithm distinguished and quantified these enzymes with over 95% accuracy, outperforming conventional methods by 15–20 min⁴⁷. Still, the model requires adjusting the hyperparameters to avoid overfitting, particularly when using high-harmonic data. Additionally, the hybrid algorithm's non-specific combination reduces classification accuracy when subjected to excessive

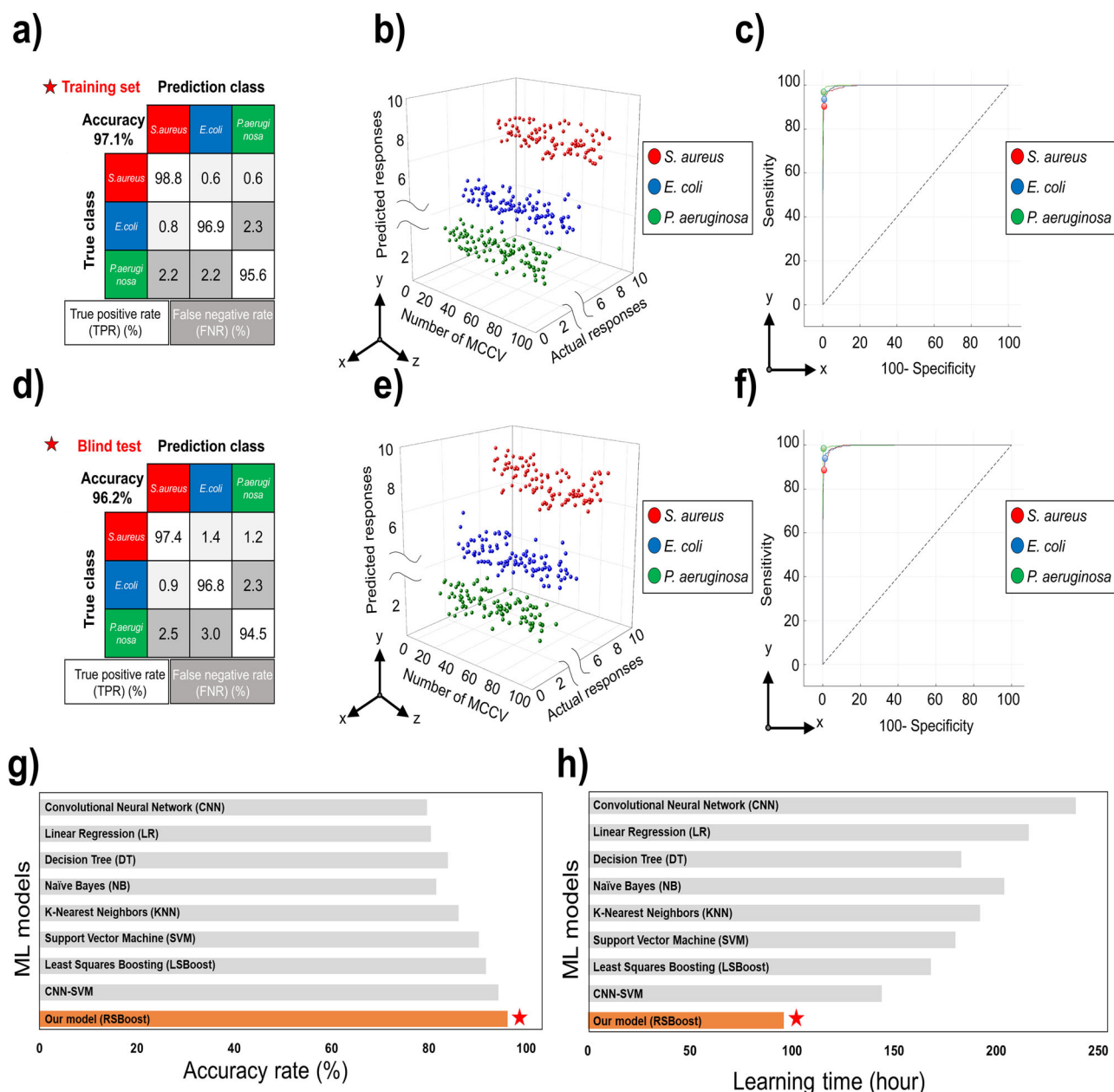


Fig. 5 | Learning results of the RSBoost hybrid algorithm. **a** The confusion matrix of training results for three classes of bacteria (*S. aureus* is red, *E. coli* is blue, and *P. aeruginosa* is green). **b** The 3-dimensional scatter plot of actual and predicted data is validated by the Monte Carlo cross-validation (MCCV) in the training set. **c** The ROC plot and AUC values of each class for the training set. **d** The confusion matrix of the blind test result for three classes of bacteria. **e** The 3-dimensional scatter plot of

actual and predicted data is validated by the Monte Carlo cross-validation (MCCV) in the blind test set. **f** The ROC plot and AUC values of each class for the test set. **g** Comparison of classification accuracy for each model in classification learning for three types of bacteria. **h** Learning time comparison of each classification model for identifying three bacterial species (All bacterial classes are the same as the color legend mentioned above).

iterations during training. Similarly, Adak et al. developed a hybrid artificial bee colony (ABC) algorithm for classifying five types of alcohol, achieving superior performance and a lower error rate compared to the back-propagation (BP) algorithm typically used in such analyses, with an MSE of 1.41×10^{-646} .

Conversely, the model shows a potential overfitting in the artificial neural network (ANN) while tuning specific hyperparameters. Also, the ABC algorithm for training can struggle with parameter instability. Asadollah et al. also proposed a hybrid model, combining Gradient Boosting (GB) and Support Vector Regression (SVR) to predict soil moisture in lake watersheds. This Gradient Boosting-Support Vector Regression (GB-SVR) hybrid model demonstrated improvements of 17%, 10%, and 13% in

correlation coefficient (R^2), RMSE, and MAE, respectively, compared to standalone GB and SVR models⁴⁸. A small number of in-situ soil moisture samples may restrict model generalizability. Moreover, the GB-SVR hybrid model's performance inconsistencies across various soil types and climatic conditions hinder its applicability across diverse environments. Herein, we successfully introduced a hybrid algorithm that can learn from various learning models and large-scale datasets, significantly improving learning performance and accuracy.

In Table 1, ML-based platforms combining colorimetric sensor arrays and machine learning achieve high accuracy rates above 90%. However, including preprocessing and analysis time, the process can take a minimum of 240 h to 180 days. While this approach provides high

Table 1 | Comparison of the performance of an ML-powered sensor platform combining an existing colorimetric sensor array and machine learning to identify three types of bacteria

Analytical instruments	Accuracy (%)	Total analysis time (days)	Number of the classes	Reference
Machine learning-powered colorimetric sensor array	95	>12.5	10	57
	>90	>180	16	58
	91–95	>10	2	59
	97	<25	10	60
	95	<41.6	5	61
	99	>10.8	5	62
	98	<16	3	63–66
	96	<4.1	9	Our study

classification accuracy, it has the drawback of lengthy analysis relative to the number of classes. Additionally, extracting features from the large-scale data generated by conventional colorimetric sensor arrays is highly complicated and time-consuming. This leads to biased learning, resulting in low reliability for specificity and sensitivity values and frequent occurrences of overfitting.

This study presents a comprehensive investigation into sepsis diagnostics, emphasizing the enhanced sensitivity of the colorimetric gas sensor array through poly ionic liquids and pH indicators. We also propose integrating ML-based hybrid algorithms for bacterial classification to provide clinicians with vital information regarding the dominant bacterial species causing sepsis. Additionally, the results demonstrate the model's ability to detect and differentiate bacterial species accurately and estimate bacterial concentrations in controlled laboratory conditions and blood samples. The RSBoost hybrid algorithm, combining the strengths of two existing hybrid models for large-scale multi-class classification on heterogeneous datasets, analyzes complex correlation structures that are visually challenging to discern, performing simultaneous classification and regression learning through latent feature extraction and weight assignment. In blind tests using blood samples, the algorithm achieved an average accuracy of 96.2% and speed of less than 24 h (after training the RSBoost hybrid algorithm). Thus, our approach is expected to improve sepsis diagnosis and patient outcomes by enabling rapid, accurate, targeted interventions to manage this life-threatening condition. Still, we acknowledge that our initial validation was conducted on a tiny and homogeneous donor group (three healthy, non-smoking male donors) to ensure controlled proof-of-concept conditions, which is a study limitation. Therefore, future studies will include larger and more diverse populations to validate and generalize our findings, ensuring that these promising results broadly apply to diverse patient populations.

Methods

Synthesis of PIL and anionic pH dye

Each pH indicator dye was prepared in solution under conditions ensuring its anionic form (for example, 10 mg of BPB dissolved in 14 mL ethanol with a small amount of tributyl phosphate) before combining with the PIL solution. We blended the dye solution with the polymeric ionic liquid (20:1 v/v) to allow anion exchange between the dye and the PIL's original Br⁻ counterions, effectively anchoring the dye anions in the polymer matrix. A homogeneous solution was prepared by stirring and ultrasonically mixing a 1:1 weight ratio of [Bvim][Br] and acrylonitrile with 8 wt% divinylbenzene (based on the monomer's weight) and 1 wt% benzoin isobutyl ether as a photoinitiator. The solution was cast into a glass beaker and photocrosslinked at room temperature under 254 nm UV light. Consecutively, it was immersed in a pH dye solution to undergo an anion-exchange reaction. The solution was cast into a glass beaker and photo-crosslinked under 254 nm UV light (5–10 mW/cm² intensity at 10 cm distance) for 30 min at room temperature, yielding a fully crosslinked polymer gel.

Fabrication of pH indicator-based colorimetric sensor

Optical modeling and the associated A 10:1 weight ratio of PDMS pre-polymer (Sylgard 184A) and curing agent (Sylgard 184B) was mixed to fabricate the porous PDMS sponge. The mixture was consecutively poured over sugar cubes due to complete solubility, macroporosity, and biocompatibility in a petri dish, degassed at 0.08 MPa, and cured at 80 °C. And the sugar templates were dissolved in deionized water, dried at 80 °C, and coated with the sensor solution⁴⁹.

Characterization

The absorbance measurements for the colorimetric sensors were acquired using a UV–Vis spectrophotometer (Infinite® 200 PRO, Tecan Inc.). The results were assessed by recording the UV–Vis absorption spectrum with monochromators, covering the 300 to 800 nm wavelength range.

NMR (Nuclear Magnetic Resonance) spectroscopy were conducted using a Bruker 9.4T wide-bore magnet (400 MHz for the ¹H Larmor frequency), controlled by an AVANCE-III NMR spectrometer. The probe was a Bruker 3.2 mm double-resonance ¹H/X CPMAS probe, and the sample temperature was maintained at 298.1 K. The spectra were acquired using DMSO as the solvent, and the spinning frequency of the sample was set to 10 kHz in both cases. XPS spectra were collected on an X-ray photoelectron spectroscopy (XPS, 5000 VersaProbe) using a monochromatic Al K α (1486.6 eV) radiation. The samples were analyzed under ultra-high vacuum conditions (2 \times 10⁻⁷ Pa). After recording a broad range spectrum (pass energy 187.85 eV), high-resolution spectra were recorded for the B1s, N1s, C1s, and O1s. Spectrum processing was carried out using the Casa XPS software package.

Gas sensing measurement

The gas sensing properties of the pH indicator coated colorimetric sensors (BPB, TFT, and Cu-PAN) were measured in the fabricated chamber. The gas flow was calibrated by mixing dry and humid air with the desired concentration of the target gases using Mass Flow Controllers (MFCs, Phocos, i-300CV-S4) to achieve a constant flow rate of 1000 sccm. The MFC was precisely controlled via the LabVIEW software and used a fixed mixing time of 300 s. All target gases (NH₃, TMA, H₂S, etc.) were high-purity standards (\geq 99.5%) obtained in certified low-concentration cylinders (e.g., 10 ppm NH₃ in N₂), purchased from Shinyang Industrial Gases Co. And before introducing the test gas into the sensor chamber, we diverted all gas streams (the VOC stream plus dry and humid carrier air) to a vent line for a 300-s pre-mixing period.

The color change of the pH indicator coated colorimetric sensors according to the gas flow was recorded using a smartphone (Samsung Galaxy S7), and the smartphone was taken in an automatic calibration state. To ensure that color changes arose solely from target gases, all experiments were conducted in sealed chambers with inert gas purging to eliminate ambient VOC interference; additionally, a fixed smartphone imaging setup (10 cm distance) with uniform internal lighting was used to maintain consistency across tests.

Bacterial strains and growth conditions

The *S. aureus*, *P. aeruginosa*, and *E. coli* were purchased by the Korean Collection for Type Culture. *S. aureus* (KCTC 3881), *P. aeruginosa* (KCTC 1750), *E. coli* (KCTC 2441), Korean Collection for Type Cultures (KCTC, Jeongseup, Republic of Korea), LB medium (Alpha Biosciences, Baltimore, MD, USA). Each bacterium was inoculated on lysogeny broth (LB) medium and grown overnight at 37 °C with shaking at 200 rpm. Then, a 10-fold dilution series was made ranging from 10⁸ CFU/ml down to 10¹ CFU/ml by using LB liquid medium. The bacterial suspension was used for extraction of DNA as shown in Supplementary Fig. 22.

DNA extraction of bacteria and PCR condition

Genomic DNA was extracted using an AccuPrep Genomic DNA Extraction Kit (Bioneer, Korea). The concentration of DNA was measured by Nano-drop 2000™ spectrophotometer. The extracted DNAs from bacteria were

analyzed for identifying the sensitivity and specificity of conventional PCR method as shown in Supplementary Fig. 23. The primers of each bacterium were listed in Supplementary Table 3. The temperature of the PCR reaction had an initial denaturation step of 95 °C for 3 min followed by 35 cycles (3 s at 95 °C, 30 s at 60 °C) in a real time PCR thermal cycler (Biorad, Germany). Bacterial quantification for DNA was performed according to the threshold cycle using the CT method. The values presented in the graphs are mean ± SD values

Preparation of blood sample

Following the acquisition of written informed consent, three healthy, non-smoking male participants (mean age 38.2) each donated 10 ml of blood. Human blood samples were collected using the conventional protocol approved by the Institutional Review Board (IRB) (IRB number of Severance Hospital, Yonsei University: IRB, 4-2023-1279). Each 10 mL donor blood sample was immediately inoculated with target bacteria (*S. aureus*, *E. coli*, or *P. aeruginosa*) to create an infected blood culture, mimicking a sepsis bloodstream infection. A negative-control blood samples (with no bacteria added) were incubated under the same conditions and yielded no sensor response.

Preparation of heterogeneous datasets

Data preprocessing is conducted by dividing the datasets into video data and color value data for each sensor area frame based on the hybrid algorithm's classification and regression learning. First, the video dataset, with a frame rate of 29.97 fps and a frame width of 480 × 272 at 2078 kbps, is split into an image pixel size of 480 × 272 × 3. Only the sensor area is then extracted and rescaled using MATLAB's Vision Toolbox before being sent to the input layer of the CNN-SVM. Secondly, the intensity datasets, with Red, Green, and Blue (R, G, B) values extracted for each frame and arranged by column, are compiled into a datasheet (4 × 89,965) for each combination concentration and sent to the input layer of LSBoost for regression learning of bacterial concentration. Furthermore, all datasets transmitted to each hybrid algorithm are divided into 70% training and 30% testing sets.

Dataset standardization and normalization

The entire dataset (frames 2–10,000) was randomly split into a training set (70%) and a test set (30%) to ensure a robust evaluation. Then, the colorimetric sensor extracted images for each frame from the 24-h video file and converted them into an image dataset of the same resolution. Next, all images were resized to a fixed spatial resolution (e.g., 480 × 272 pixels) and their RGB channels were scaled uniformly, ensuring consistent color representation across samples. These images were then loaded into a MATLAB ImageDataset and shuffled randomly to support unbiased mini-batch sampling during training. Finally, within each sensor's region of interest, pixel intensities are normalized to a mean of 0 and a variance of 1 in both the row and column directions. The number of variables in each row and column is equalized to eliminate class imbalance.

Extraction of color intensities and morphological characteristics from sensor array

To extract the latent characteristics of 3 bacteria species in each frame, the captured images were used to identify areal changes regarding the bacteria ratio. Based on these results, spatial vectors were calculated. These spatial vectors identify the size and number of responsive regions for each gas concentration, with different vector values for each class. This approach allows for determining how the RGB ratios within the contoured sensor area are divided, resulting in distinct RGB values for each bacterial species. Following this, the V values were derived from the RGB values of the scanned images, as V values in the HSV (Hue, Saturation, and Value) model demonstrated high sensitivity and extensive measurement ranges^{49–52}.

Hyperparameter determination for RSBoost hybrid algorithm

The hyperparameters were assigned to set the conditions for training and blind testing of the hybrid algorithm: an input size of 480 × 272, a batch size

of 32, a learning rate of 0.01, 1000 epochs, and 500 iterations for classification and a leaf size of 8, a learning rate of 0.01, 1000 epochs, and 500 iterations for regression.

Classification process using CNN-SVM layer of RSBoost hybrid algorithm

The hybrid algorithm employs a superpixel tool (MATLAB 2023b), specialized in extracting space vector⁵³ values, to uncover latent features in the data. By setting the superpixel scale to 200, the spatial coordinates of the sensor area are extracted, and the length scale is calculated to derive the features. The length scale is computed using Eq. (2). Here, x and y represent the coordinates of the area along the sensor boundary.

$$\text{Length scale}(L) = \sqrt{\frac{\sum_{i=1}^n (x_i^2 + y_i^2)}{n}} \quad (2)$$

The RGB index utilizes the masking region tool (MATLAB 2023b) to collect morphological contour information per frame per second and extract latent features classified by bacterial combination ratios for each class. The RGB index is calculated using Eq. (3). Here, n is the number of color channels in the image.

$$\text{RGB index} = \frac{(R, G, B)_{\max} - (R, G, B)_{\min}}{10^n} \quad (3)$$

The shape factor is employed to detect changes in color boundaries over time within a colorimetric sensor⁵⁴. It extracts consistent contour information correlating with color variations due to different bacterial concentrations. By analyzing the shape of these color changes, the shape factor is calculated using Eq. (4) to estimate the area of the color region, which is then divided into 1 to 5 distinct regions, and the internal structure of the sensor is evaluated. P refers to the perimeter of the sensor, and A represents the area of the sensor.

$$\text{Shape factor (SF)} = \frac{4\pi A}{P^2} \quad (4)$$

Regression process using LSBoost layer of RSBoost hybrid algorithm

For regression learning of gas concentration combinations for three types of bacteria in the hybrid algorithm, automatic feature equations are derived using Eqs. (5) and (6). Weight assignment is then performed using Eq. (7), allocating weights to the index most effective for concentration prediction.

$$F_n(x) = F_{n-1}(x) + y_n W_n(x, E_n) \quad (5)$$

$$y_n = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Predicted concentration}^{(i)} - \text{Actual concentration}^{(i)})^2} - \text{RMSE}_{\min}}{\text{RMSE}_{\max} - \text{RMSE}_{\min}} \quad (6)$$

$$\text{Corr}(X, Y) = \frac{1}{\sigma_X \sigma_Y} \frac{1}{n} \sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y) \quad (7)$$

To quantitatively evaluate the regression learning results, mean absolute error (MAE), mean square error (MSE), and root-mean-square error (RMSE) equations were applied to the output layer of LSBoost (Eqs. (8), (9), and (10), respectively).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{x}_i - x_i| \quad (8)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|^2 \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|^2} \quad (10)$$

Correlation index

The correlation index pattern, ranging from -1 to 1 , was used to analyze the correlation of multi-class regression learning results for the three bacteria at concentration combinations of 10, 100, and 1000 cfs³⁶. Through the RSBoost hybrid algorithm, this pattern evaluates the linear relationship between variables. A value close to 1 indicates minimal latent features among the same class, while a value close to -1 suggests diverse latent features across different bacterial concentration classes. This correlation index pattern is calculated using Eq. (11) [Supplementary Fig. 16].

$$R^2 = \frac{\sum (\hat{x}_i - x_i)(\hat{y}_i - y_i)}{\sqrt{\sum (\hat{x}_i - x_i)^2 \sum (\hat{y}_i - y_i)^2}} \quad (11)$$

Here, R^2 represents the correlation coefficient, and x_i and y_i denote different bacterial classes x and y , respectively. \hat{x}_i and \hat{y}_i represent the mean values of each sample [Supplementary Figs. 17–21]. Residuals are a critical metric for assessing the performance of regression models⁵⁵. In this study, the RSBoost hybrid algorithm was trained on a dataset featuring three bacterial proportions relevant to sepsis diagnosis. The LSBoost layer, designed for regression learning within the hybrid algorithm, was employed to analyze the mean and variance of residuals—the discrepancies between actual and predicted values—to evaluate the consistency of the algorithm's prediction errors. Residuals are computed using Eq. (12), and verifying that they follow a normal distribution is essential. The residual e_i is defined as the difference between the actual value y_i and the predicted value \hat{y}_i .

$$e_i = y_i - \hat{y}_i \quad (12)$$

Monte Carlo Cross Validation (MCCV)

MCCV, suitable for large datasets, was employed to evaluate and verify the reliability of the classification learning results of the RSBoost hybrid algorithm⁵⁶. The number of MCCV iterations was set to 100, as calculated by Eq. (13).

$$\varepsilon_{MCCV}(\%) = \frac{1}{100} \sum_{i=1}^{100} \varepsilon_i \quad (13)$$

Probability value (p-value)

To evaluate the statistical significance of performance differences between the RSBoost algorithm and baseline models, a paired t -test was conducted. The p -value was used as the primary statistical metric to determine whether the observed differences were likely due to random variation under the null hypothesis. Where \bar{d} represents the mean difference performance metrics, s_d is the standard deviation of those differences, and n is the number of paired observations used for statistical comparison. It is calculated by Eq. (14).

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \quad (14)$$

All algorithm training processes were executed on an Intel Core i9-12900KS CPU running at 3.40 GHz, 32GB of RAM, and an NVIDIA GeForce RTX 3090 Ti graphics card, using MATLAB R2023b (The MathWorks, Inc.; Natick, MA, USA).

Algorithm 1. The training process of rapid sepsis boosting (RSBoost)

Input: Data sources collected via colorimetric sensors, P ; Number of bacterias, n ; The input datasets (image & intensity values) conditions are randomly set for training, R_n ($n = 1, 2$); Raw value of each condition, x .

Output: Weight assigned to bacteria class with lowest root mean square error (RMSE), W ; Classified bacteria class, C .

Training Begin {LSBoost layer}: The dataset is fed into the LSBoost layer of the hybrid algorithm.

1. $F_o(x) = \bar{y}$: The raw dataset is loaded into the workspace of the LSBoost layer and transformed into a function ($F(x)$).
2. For all bacteria ($n = 9$), regression training is repeated for each bacteria class per epoch.
3. The optimized hyperparameters are 0.01 learning rate, 1.0 maximum feature map depth, 8 minimum leaf size, 1000 epochs, and 500 iterations.
4. $\bar{y}_n = y_n - F_{n-1}(x_n)$: The LSBoost layer of the algorithm normalizes the regression results for each of the nine classes of bacteria.
5. $(R_{n=1,2}, P_n) = \text{argmin}_p \sum_{n=1}^9 [\bar{y}_i - RW_n(x_n; P)]^2$: Among the normalized variables of multi-classes for each bacteria gas concentration, weight is assigned to the value with the smallest average error.
6. $F_n(x) = F_{n-1}(x) + R_n W_n(x; P_n)$: Every bacteria class assigned a weight is exported from the workspace to the output layer of the algorithm.
7. **EndFor:** The iterative learning for each bacteria class concludes.
8. **Return {CNN-SVM layer}:** The results of regression learning are sent to the CNN-SVM layer of the hybrid algorithm.
9. For The algorithm is trained on nine classes for the concentration range of three bacteria (*S. aureus*, *E. coli*, *P. aeruginosa*): 10 to 1000 CFU/ml.
10. The top nine classes for a concentration of bacteria sorted by descending RMSE trained by the LSBoost layer are used in a 7×7 kernel filter to extract latent features.
11. $(R_{n=1,2}, P_n) = \text{argmin}_p \sum_{n=1}^9 W[\bar{y}_i - RW_n(R_{n=1,2} + W_n(x; P_n); PW_n)]^2$: Among the ranking using the nine class gas concentrations of bacteria, the weight is reapplied to the class with the highest accuracy.
12. The weights are applied to each of the nine classes, ranked by RMSE values, and the eigenvectors are computed using the characteristic equation as input to the CNN-SVM layer.
13. **Function:** A feature map is generated based on the eigenvector range for each class.
14. Save the trained CNN-SVM layer of the RSBoost algorithm.
15. Send the eigenvector values into the test set of the CNN-SVM layer for a blind test.
16. The accuracy of each class is calculated using an SVM classifier of the CNN-SVM layer and exported.
17. End While;
18. End Algorithm.

Data availability

This study uses a heterogeneous dataset authorized by Mr. Ha, Dr. Shin and Dr. Jung at Yonsei University. Users wishing to access information from all datasets must obtain their approval. In comparison, some of the datasets are partially open and accessible below. Detailed information on some of the heterogeneous datasets, including blood samples used for the blind test of the RSBoost algorithm and the training dataset for bacterial gas concentration, is available on Zenodo (<https://doi.org/10.5281/zenodo.14053366>).

Code availability

The hybrid algorithm developed for this study has been made part of the code available on GitHub. The algorithm was developed and tested using the MATLAB 2023b environment, and information about the code scripts and toolboxes used can be found at (<https://github.com/SeongminHA/Artificially-intelligent-nasal-perception-for-rapid-sepsis-diagnostics?tab=readme-overview>).

Received: 15 April 2025; Accepted: 29 June 2025;

Published online: 24 July 2025

References

- Liu, Y. C. et al. Frequency and mortality of sepsis and septic shock in China: a systematic review and meta-analysis. *BMC Infect. Dis.* **22**, 564 (2022).
- Boeddha, N. P. et al. Mortality and morbidity in community-acquired sepsis in European pediatric intensive care units: a prospective cohort study from the European Childhood Life-threatening Infectious Disease Study (EUCLIDS). *Crit. Care* **22**, 1–13 (2018).
- Maneta, E. et al. Endothelial dysfunction and immunothrombosis in sepsis. *Front. Immunol.* **14**, 1144229 (2023).
- Hotchkiss, R. S. et al. Sepsis and septic shock. *Nat. Rev. Dis. Prim.* **2**, 1–21 (2016).
- Cassini, A. et al. Global Report on the epidemiology and burden on sepsis: current evidence, identifying gaps and future directions. Global Report on the epidemiology and burden on sepsis: current evidence, identifying gaps and future directions. <https://www.who.int/publications/i/item/9789240010789> (2020).
- Fleischmann, C. et al. Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. *Am. J. Respir. Crit. Care Med.* **193**, 259–272 (2016).
- Halili, M. A. et al. Small molecule inhibitors of disulfide bond formation by the bacterial DsbA–DsbB dual enzyme system. *ACS Chem. Biol.* **10**, 957–964 (2015).
- Angus, D. C. & Van der Poll, T. Severe sepsis and septic shock. *N. Engl. J. Med.* **369**, 840–851 (2013).
- Di Domenico, M. et al. Diagnostic accuracy of a new antigen test for SARS-CoV-2 detection. *Int. J. Environ. Res. Public Health* **18**, 6310 (2021).
- Yagupsky, P. & Nolte, F. S. Quantitative aspects of septicemia. *Clin. Microbiol. Rev.* **3**, 269–279 (1990).
- Bekeris, L. G., Tworek, J. A., Walsh, M. K. & Valenstein, P. N. Trends in blood culture contamination: a College of American Pathologists Q-Tracks study of 356 institutions. *Arch. Pathol. Lab. Med.* **129**, 1222–1225 (2005).
- Ai, Z. et al. On-demand optimization of colorimetric gas sensors using a knowledge-aware algorithm-driven robotic experimental platform. *ACS Sens.* **9**, 745–752 (2024).
- Asri, M. I. A., Hasan, M. N., Fuaad, M. R. A., Yunus, Y. M. & Ali, M. S. M. MEMS gas sensors: a review. *IEEE Sens. J.* **21**, 18381–18397 (2021).
- CHUN, H. W. et al. Pure-water-soluble colorimetric chemosensors for highly sensitive and rapid detection of hydrogen sulfide: Applications to evaluation of on-site water quality and real-time gas sensors. *Sens. Actuators B: Chem.* **402**, 134989 (2024).
- JANG, J. H. et al. Development of a pH indicator for monitoring the freshness of minced pork using a cellulose nanofiber. *Food Chem.* **403**, 134366 (2023).
- Sun, Y., Zhao, C., Niu, J., Ren, J. & Qu, X. Colorimetric band-aids for point-of-care sensing and treating bacterial infection. *ACS Cent. Sci.* **6**, 207–212 (2020).
- Chen, Q., Li, H., Ouyang, Q. & Zhao, J. Identification of spoilage bacteria using a simple colorimetric sensor array. *Sens. Actuators B Chem.* **205**, 1–8 (2014).
- Puttaswamy, S., Lee, B. D. & Sengupta, S. Novel electrical method for early detection of viable bacteria in blood cultures. *J. Clin. Microbiol.* **49**, 2286–2289 (2011).
- Narayana lyengar, S. et al. Toward rapid detection of viable bacteria in whole blood for early sepsis diagnostics and susceptibility testing. *ACS Sens.* **6**, 3357–3366 (2021).
- Guo, J., Qiu, L., Deng, Z. & Yan, F. Plastic reusable pH indicator strips: preparation via anion-exchange of poly (ionic liquids) with anionic dyes. *Polym. Chem.* **4**, 1309–1312 (2013).
- Dong, G., & Liu, H. *Feature engineering for machine learning and data analytics* (CRC Press, 2018).
- Ganapathy, N., Baumgärtel, D. & Deserno, T. M. Automatic detection of atrial fibrillation in ECG using co-occurrence patterns of dynamic symbol assignment and machine learning. *Sensors* **21**, 3542 (2021).
- Ha, N., Xu, K., Ren, G., Mitchell, A. & Ou, J. Z. Machine learning-enabled smart sensor systems. *Adv. Intell. Syst.* **2**, 2000063 (2020).
- Goh, K. H. et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat. Commun.* **12**, 711 (2021).
- Sharma, S., Gahlawat, V. K., Rahul, K., Mor, R. S. & Malik, M. Sustainable innovations in the food industry through artificial intelligence and big data analytics. *Logistics* **5**, 66 (2021).
- Ye, Z. et al. Tackling environmental challenges in pollution controls using artificial intelligence: a review. *Sci. Total Environ.* **699**, 134279 (2020).
- Xu, C., Solomon, S. A. & Gao, W. Artificial intelligence-powered electronic skin. *Nat. Mach. Intell.* **5**, 1344–1355 (2023).
- Park, J. H. et al. Classification of circulating tumor cell clusters by morphological characteristics using convolutional neural network-support vector machine. *Sens. Actuators B Chem.* **401**, 134896 (2024).
- Lee, K. Y. et al. Machine learning-powered electrochemical aptasensor for simultaneous monitoring of di (2-ethylhexyl) phthalate and bisphenol A in variable pH environments. *J. Hazard. Mater.* **462**, 132775 (2024).
- Lupoiu, R., Chen, M., Shao, Y., Mao, C., & Fan, J. A. Ultra-fast optimization of aperiodic metasurface superpixels using conditional physics-augmented deep learning. In *Flat optics: components to systems*, FTh1B-3 (Optica Publishing Group, 2023).
- Wang, C., Alaya Cheikh, F., Kaaniche, M., Beghdadi, A. & Elle, O. J. Variational based smoke removal in laparoscopic images. *Biomed. Eng. Online* **17**, 1–18 (2018).
- Jia, X., Ma, P., Tarwa, K., Mao, Y. & Wang, Q. Development of a novel colorimetric sensor array based on oxidized chitin nanocrystals and deep learning for monitoring beef freshness. *Sens. Actuators B Chem.* **390**, 133931 (2023).
- Mansour, E. et al. Measurement of temperature and relative humidity in exhaled breath. *Sens. Actuators B Chem.* **304**, 127371 (2020).
- Benabdelhalim, H. & Brutin, D. Influence of relative humidity and temperature on human whole blood drying. *Dry. Technol.* **41**, 434–443 (2023).
- Cohen, I. et al. Pearson correlation coefficient. In *Noise reduction in speech processing*, 1–4, https://doi.org/10.1007/978-3-642-00296-0_5 (2009).
- Ratner, B. The correlation coefficient: its values range between + 1/– 1, or do they? *J. Target. Meas. Anal. Mark.* **17**, 139–142 (2009).
- Xu, Q. S. & Liang, Y. Z. Monte Carlo cross validation. *Chemometrics Intell. Lab. Syst.* **56**, 1–11 (2001).
- Zhao, Q. et al. Spherical superpixel segmentation. *IEEE Trans. Multimed.* **20**, 1406–1417 (2017).
- Gao, J., Yang, Y., Lin, P. & Park, D. S. Computer vision in healthcare applications. *J. Healthc. Eng.* **2018**, 5157020 (2018).
- Dalla Mura, M., Benediktsson, J. A., Waske, B. & Bruzzone, L. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Trans. Geosci. Remote Sens.* **48**, 3747–3762 (2010).
- Bukkapatnam, A., Manjaly, S. & Senthil, R. A novel approach to sepsis diagnosis: Using artificial intelligence to assist clinicians and innovate. *Proc. Comput. Sci.* **231**, 237–242 (2024).
- Yuan, K. C. et al. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *Int. J. Med. Inform.* **141**, 104176 (2020).
- Zhao, X., Shen, W. & Wang, G. Early prediction of sepsis based on machine learning algorithm. *Comput. Intell. Neurosci.* **2021**, 6522633 (2021).

44. Steinbach, D. et al. Applying machine learning to blood count data predicts sepsis with ICU admission. *Clin. Chem.* **70**, 506–515 (2024).
45. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**, 1345–1359 (2009).
46. Adak, M. F., Lieberzeit, P., Jarujamrus, P. & Yumusak, N. Classification of alcohols obtained by QCM sensors with different characteristics using ABC based neural network. *Eng. Sci. Technol. Int. J.* **23**, 463–469 (2020).
47. Tatarko, M. et al. Machine learning enabled acoustic detection of sub-nanomolar concentration of trypsin and plasmin in solution. *Sens. Actuators B Chem.* **272**, 282–288 (2018).
48. Asadollah, S. B. H. S., Sharafati, A., Saeedi, M. & Shahid, S. Estimation of soil moisture from remote sensing products using an ensemble machine learning model: a case study of Lake Urmia Basin, Iran. *Earth Sci. Inform.* **17**, 385–400 (2024).
49. Shin, J. et al. Smart forensic kit: Real-time estimation of postmortem interval using a highly sensitive gas sensor for microbial forensics. *Sens. Actuators B Chem.* **322**, 128612 (2020).
50. Yang, J. S., Shin, J., Choi, S. & Jung, H. I. Smartphone Diagnostics Unit (SDU) for the assessment of human stress and inflammation level assisted by biomarker ink, fountain pen, and origami holder for strip biosensor. *Sens. Actuators B Chem.* **241**, 80–84 (2017).
51. Shin, J. et al. Smart forensic phone: colorimetric analysis of a bloodstain for age estimation using a smartphone. *Sens. Actuators B Chem.* **243**, 221–225 (2017).
52. Choi, W., Shin, J., Hyun, K. A., Song, J. & Jung, H. I. Highly sensitive and accurate estimation of bloodstain age using smartphone. *Biosens. Bioelectron.* **130**, 414–419 (2019).
53. Achanta, R. et al. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 2274–2282 (2012).
54. Phillip, J. M., Han, K. S., Chen, W. C., Wirtz, D. & Wu, P. H. A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. *Nat. Protoc.* **16**, 754–774 (2021).
55. Ling, R. F. Residuals and influence in regression, 413–415. <https://hdl.handle.net/11299/37076> (1984).
56. Gourvénec, S., Pierna, J. F., Massart, D. L. & Rutledge, D. N. An evaluation of the PoLiSh smoothed regression and the Monte Carlo cross-validation for the determination of the complexity of a PLS model. *Chemometrics Intell. Lab. Syst.* **68**, 41–51 (2003).
57. Carey, J. R. et al. Rapid identification of bacteria with a disposable colorimetric sensing array. *J. Am. Chem. Soc.* **133**, 7571–7576 (2011).
58. Zhang, Y. & Xu, X. Modulus of elasticity predictions through LSBoost for concrete of normal and high strength. *Mater. Chem. Phys.* **283**, 126007 (2022).
59. Yang, M. et al. Machine learning-enabled non-destructive paper chromogenic array detection of multiplexed viable pathogens on food. *Nat. Food* **2**, 110–117 (2021).
60. Yang, J. et al. Machine learning-assistant colorimetric sensor arrays for intelligent and rapid diagnosis of urinary tract infection. *ACS Sens.* **9**, 1945–1956 (2024).
61. Chen, X. & Yang, X. Image semantic recognition algorithm of colorimetric sensor array based on deep convolutional neural network. *Adv. Multimed.* **2022**, 4325117 (2022).
62. Lin, Y., Ma, J., Cheng, J. H. & Sun, D. W. Visible detection of chilled beef freshness using a paper-based colourimetric sensor array combining with deep learning algorithms. *Food Chem.* **441**, 138344 (2024).
63. Lin, Y., Ma, J., Sun, D. W., Cheng, J. H. & Wang, Q. A pH-Responsive colourimetric sensor array based on machine learning for real-time monitoring of beef freshness. *Food Control* **150**, 109729 (2023).
64. Bräuer, L. et al. Staphylococcus aureus and Pseudomonas aeruginosa express and secrete human surfactant proteins. *PloS One* **8**, e53705 (2013).
65. Kim, Y. G. et al. Antibiofilm activity of phorbaketals from the marine sponge Phorbasp. against Staphylococcus aureus. *Mar. Drugs* **19**, 301 (2021).
66. Dalmasso, G. et al. Genes mcr improve the intestinal fitness of pathogenic E. coli and balance their lifestyle to commensalism. *Microbiome* **11**, 12 (2023).

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2024-00356330, RS-2024-00333848). This work was also supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A3A14038599) and the Korea government (MSIT) (RS-2024-00333848) and the Technology Innovation Program (00144157, Development of Heterogeneous Multi-Sensor Micro-System Platform) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea), the Korea Institute of Science and Technology (2E33181, 2V10301, and 2V10120), and supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.2020R1A5A1018052). Dr. Hyuk Min Lee from Department of Laboratory Medicine, College of Medicine, Yonsei University offered bacteria strain and gram stain images.

Author contributions

J.C.S., S.M.H., G.S.K., and T.H.Y. wrote the main manuscript, designed the methodology, and conducted the experiments. T.H.L., W.H., K.Y.L., S.J.P., S.Y.P., S.H.H., and J.W.S. performed formal analysis and data curation. J.C.S., J.W.L., H.C.S., J.S.J., J.S.K., H.I.J., and C.Y.K. supervised the project and contributed fund resources. J.C.S., J.S.J., H.I.J., and C.Y.K. conceptualized the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-025-01851-4>.

Correspondence and requests for materials should be addressed to Ji-Soo Jang, Jin-Sang Kim, Hyo-Il Jung or Chong-Yun Kang.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025