Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (361-373)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.54559

JURNAL TEKNIK INFORMATIKA

Homepage : http://journal.uinjkt.ac.id/index.php/ti

Machine Learning for H_2S Exposure Detection with a Low-Cost MQ-136 Sensor: Signal-Stability Screening and Time-Blocked Validation

Mula Agung Barata[1*], Dwi Irnawati[2], Ifnu Wisma Dwi Prastya[3], Shukor Sanim Mohd Fauzi[4], Indra Dharma Wijaya[5]

[1,3]Department of Informatics Engineering, Faculty of Science and Technology, University of Nahdlatul Ulama Sunan Giri, Bojonegoro, Indonesia
[2]Department of Retail Management, Faculty of Economics, University of Bojonegoro, Indonesia
[4]Faculty of Computer Science and Mathematics, Universiti Teknologi MARA, Perlis, Malaysia
[5]Department of Business Information System, Information Technology, State Polytechnic of Malang. Indonesia

Article:
Accepted: June 14, 2025
Revised: August 17, 2025
Issued: October 30, 2025

© Barata et al, (2025).

This is an open-access article under the CC BY-SA license

***Correspondence Address:**
mula.ab@unugiri.ac.id

ABSTRACT

Hydrogen sulfide (H_2S) is a toxic gas in oil and gas operations, and low-cost sensors could support continuous monitoring, but their reported accuracies can be inflated by transient responses and randomly shuffled validation. This study aimed to assess whether H_2S exposure can be distinguished from background air with a single MQ-136 sensor and how screening and validation design affect performance. The method recorded one background and one exposure session (10,000 readings, 100 windows), screened the windows by signal stability (standard deviation ≤ 0.05 ppm), leaving 66, extracted four statistical features, and compared C4.5 with a threshold baseline and five classical classifiers using time-blocked 5-fold cross-validation. The results showed that C4.5 achieved 87.8% accuracy and 96.7% H_2S recall with a single rule on the window maximum, and all classifiers except SVM performed comparably (83.3–89.4%). Without screening, C4.5 accuracy fell from 89.6% with random to 77.0% with time-blocked cross-validation. In conclusion, signal-stability screening and time-aware validation mattered more than classifier choice, although confirmation with repeated sessions is needed.

Keywords: *C4.5; electronic nose; hydrogen sulfide; MQ-136; signal-stability screening; time-blocked cross-validation*

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (361-373)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.54559

1. INTRODUCTION

Occupational health and safety programs are essential to protect workers in companies, government agencies, and private organizations. Oil and gas-producing companies have high occupational health and safety risks [1]. Crude oil exploration, production, and processing release several hazardous gases that threaten the workers involved, one of which is hydrogen sulfide (H_2S) [2]. H_2S is one of the leading causes of fatal gas inhalation in the workplace, and at high concentrations it paralyzes the sense of smell, so workers cannot rely on its odor as a warning [3]. A monitoring system that measures H_2S concentrations from gas leaks is therefore needed to protect workers [4], and an electronic nose can be used to detect H_2S leaks and estimate their hazard level [5].

An electronic nose (e-nose) is a device that detects and recognizes gases or odors by combining gas sensors with data processing algorithms, mimicking the human sense of smell. For H_2S detection, an e-nose relies on sensors that are sensitive to H_2S to measure its concentration in the surrounding air. E-noses have also been applied to environmental monitoring, including the odor emissions of waste and wastewater treatment plants [6]. Xie et al. combined an e-nose with machine learning for the non-destructive determination of volatile compounds and the prediction of amino acid nitrogen during sufu fermentation [7]. Barata et al. used an e-nose to classify pure tea types with the C4.5 and Chi-square algorithms, achieving 94% accuracy [8]. Feature selection has also been used to improve the accuracy of classical classifiers such as Naive Bayes [9], k-nearest neighbors [10], and artificial neural networks [11].

C4.5 is a widely used algorithm for building decision trees. It was developed by Quinlan as an extension of the ID3 (Iterative Dichotomiser 3) decision tree algorithm. It is popular in data mining because of its ability to handle various types of data and produce easily interpretable models. This interpretability also allows a learned rule to be checked against the physical behavior of the sensor, which accuracy alone cannot reveal. C4.5 has been applied successfully in diverse domains, including intrusion detection in wireless sensor networks [12], landslide susceptibility analysis [13], online voltage stability assessment [14], and rheumatoid arthritis prediction [15].

Pre-processing, such as feature selection or feature extraction, is often used to improve classification accuracy; for electronic nose signals, however, its main role is to represent the sensor response compactly, and the suitable method depends on the measurement conditions [16]. Beyond pre-processing, two methodological issues limit the reliability of the accuracy reported for low-cost gas sensors. First, metal-oxide sensors drift over time and respond to exposure events with transient peaks and slow recovery, so the recorded response depends on when it was measured as well as on the gas present [17], [18], [19]. Second, consecutive sensor readings are strongly correlated in time. When they are randomly shuffled into training and test folds, nearly identical readings appear on both sides, which leaks information and inflates accuracy estimates; blocked cross-validation, which keeps temporally contiguous data in the same fold, gives a more realistic estimate of performance on unseen measurement periods [20], [21]. Statistical features extracted from steady-state segments are commonly used to summarize electronic nose responses [16], [22], but whether they improve accuracy over raw readings must be established empirically against appropriate baselines.

Low-cost H_2S detectors and electronic noses have been developed for leak monitoring [5], [23], but, to our knowledge, few studies have evaluated machine learning models for low-cost H_2S sensors under time-aware validation or examined how transient responses affect the reported accuracy. Previous low-cost H_2S studies have mainly focused on gas discrimination and classification accuracy, for example, C4.5 classification with statistical features in our previous study [23] and selective detection of H_2S with a single metal-oxide sensor [24], whereas the effect of temporal correlation and transient responses on the reported accuracy has received less attention. Recent machine learning studies on gas sensing, including the identification of toxic gases such as H_2S with metal-oxide sensor arrays [25] and a virtual electronic nose for volatile organic compound classification [26], still estimate performance with sample-level train/test splits or k-fold cross-validation, which can overestimate accuracy when consecutive

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (361-373)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.54559

readings are correlated [20], [21]. Without time-aware validation and attention to signal stability, transient responses such as exposure peaks and sensor recovery can dominate a small dataset, and random shuffling can hide a model that does not represent the target condition.

This study therefore aims to (a) assess whether H_2S exposure can be distinguished from background air using window-level statistical features from a single MQ-136 sensor, (b) compare C4.5 with a single-threshold baseline and five other classical classifiers, and (c) quantify how signal-stability screening and time-blocked validation affect the reported performance. The contributions of this study are (1) quantitative evidence that randomly shuffled cross-validation can overestimate accuracy and conceal a physically inverted decision rule, (2) a comparison in which a single concentration threshold performed comparably to more complex classical classifiers, with no significant differences detected in this small dataset, and (3) evidence that window-level statistical features do not improve accuracy over raw readings for single-sensor data.

2. METHODS

The research workflow is shown in Figure 1. Readings were collected with a single-sensor electronic nose, segmented into fixed-length windows, screened for signal stability, summarized with four statistical features, and classified with seven classical classifiers. Performance was estimated with time-blocked cross-validation as the primary protocol.

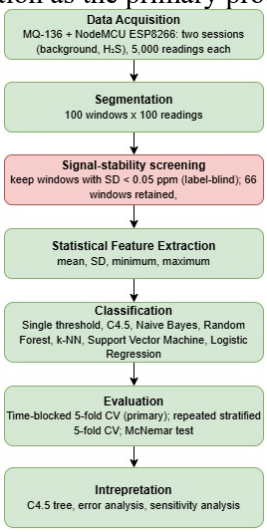


Figure 1. Research workflow

2.1. Sensing Device

The sensing device follows the electronic nose developed in our previous study [23], adapted for H_2S. It consists of an MQ-136 metal-oxide gas sensor, which is sensitive to H_2S, connected to a NodeMCU ESP8266 microcontroller (Figure 2). The sensor was used as a module with a 10 kΩ load resistor (RL) connected in series with the sensing element, which converts the change in sensor resistance into an analog output voltage. Metal-oxide sensors change their resistance in the presence of target gases, and electronic noses combine such sensors with pattern recognition [27], [28], [29]. The microcontroller reads the sensor through its analog-to-digital converter and converts each reading into voltage, the sensor resistance ratio (Rs/R0), and H_2S concentration in parts per million (ppm). The heater and circuit supplies were 5 V, as specified in the manufacturer's datasheet [35], and the sensor was preheated for 24 h before data collection, shorter than the over 48 h recommended for standard test conditions. The ADC reading was converted to voltage as $V = ADC \times 5/1023$, and the resistance ratio was computed from the load-resistor divider as $Rs/R0 = (Vc - V)/V$, where Vc = 5 V and R0 was taken as equal to the load resistance RL = 10 kΩ rather than calibrated in clean air. The concentration was then computed as ppm = a·(Rs/R0)^b, with a = 0.141 and b = −4.38, a power law whose coefficients were estimated by power-law (least-squares) fitting to the Rs/R0 versus concentration sensitivity curve in the MQ-136 datasheet [35], with a rescaled to account for R0 being taken as RL instead of the reference resistance defined in the datasheet, and then used as constants in the concentration calculation. Because a depends on this assumption, it does not correspond to the datasheet curve itself, and the converted concentrations shift if R0 differs from RL. Because no reference H_2S instrument was used, the ppm values are treated as indicative measures of sensor response rather than calibrated concentrations, and thresholds expressed in ppm, such as the screening threshold and the C4.5 split, are specific to this device and conversion. Because the device uses one sensor rather than a sensor array, it is treated in this study as a single-sensor electronic

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (361-373)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.54559

nose prototype. The assembled device is shown in Figure 3.

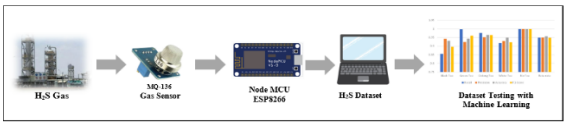


Figure 2. Configuration of the sensing device



Figure 3. Assembled sensing device

2.2. Device Communication System

The NodeMCU transmits each reading over Wi-Fi to a computer, where the readings are stored for analysis. Because H_2S is hazardous, the device was operated and monotored remotely during data collection.

2.3. Data Acquisition

H_2S samples and information on H_2S hazards were provided by Pertamina EP Asset 4 Field Sukowati, Bojonegoro. Data were collected in a closed room in two consecutive sessions. The background session recorded room air without an H_2S sample, and the exposure session recorded the room after the H_2S sample was introduced. Each reading was labeled by its session (background or H_2S exposure); no reference H_2S instrument was used. The room measured 14 m × 8 m × 4 m (≈ 448 m^3) and was not ventilated. The H_2S sample (1,500 mL) was supplied in a 1,500 mL bottle, which was opened 100 cm from the sensor. The bottle was left open for 60 mins before the exposure readings began and remained open throughout the exposure session. Readings were recorded every 0.6 s, so each window of 100 readings spans 1 min and each session of 5,000 readings lasts approximately 50 mins. The exposure session started 60 mins after the background session, without ventilating the room in between. Each session produced 5,000 consecutive readings, stored as 50 files of 100 readings. Each file is treated as one window, giving 50 windows per class and 10,000 readings in total. Each reading contains four attributes: raw sensor value, voltage, ratio, and H_2S concentration (Table 1). Because the first three attributes are transformations of the same analog reading, only the H_2S concentration was used.

Table 1. Sample of raw sensor readings

No.	Sensor Value	Voltage	Ratio	H_2S Concentration (ppm)
1.	607	2.97	0.69	0.74
2.	607	2.97	0.69	0.74
3.	607	2.97	0.69	0.74
4.	606	2.96	0.69	0.72
5.	606	2.96	0.69	0.72
6.	606	2.96	0.69	0.72
7.	606	2.96	0.69	0.72
8.	607	2.97	0.69	0.74
…	…	…	…	…
9999.	636	3.11	0.61	1.24
10000.	636	3.11	0.61	1.24

The two sessions were recorded continuously. In 42 of the 49 file boundaries of the background session and 37 of the 49 boundaries of the exposure session, the last reading of a file equals the first reading of the next file, and the remaining boundaries differ by at most 0.05 ppm. Consecutive windows are therefore strongly correlated in time, which motivates the time-blocked validation described in Section 2.7.

2.4. Signal-Stability Screening

The background session contains an abrupt peak of 12.45 ppm in window 10 followed by a slow recovery, and the first windows of the exposure session contain repeated peaks (Figure 4). These transient responses reflect exposure events and sensor recovery rather than the steady-state condition that each class is meant to represent. Windows were therefore screened by signal stability: a window was retained only if the sample standard deviation of its 100 readings was at most 0.05 ppm. This threshold was chosen as a conservative bound on sensor noise in steady conditions: it corresponds to two to three ADC counts (about 0.02 ppm per count at the operating level, Table 1) and equals the largest difference observed between consecutive files at file boundaries (Section 2.3). The threshold also falls in a natural gap in the distribution of window standard deviations: all retained windows had standard deviations of at most 0.046 ppm, whereas the smallest standard deviation among the excluded windows was 0.053 ppm (Figure 5), so small changes to the

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (361-373)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.54559

threshold within this gap would not change which windows are retained. The threshold was fixed before the classification results were examined and was not tuned to maximize accuracy; for example, C4.5 reached higher time-blocked accuracy at 0.03 ppm than at 0.05 ppm (Table 5). The criterion was applied identically to both classes and did not use class labels. Screening retained 36 of 50 background windows and 30 of 50 exposure windows (Table 2). Screening therefore restricts the classification task to steady-state windows; transient windows, which a deployed monitor must also handle, are not covered by the main evaluation (Section 3.7). To check that the conclusions did not depend on this threshold, the analysis was repeated with thresholds of 0.03, 0.08, and 0.10 ppm and without screening.

Table 2. Dataset before and after signal-stability screening

	Background	**H_2S exposure**
Windows recorded	50	50
Readings recorded	5,000	5,000
Concentration range (ppm)	0.10–12.45	1.16–2.70
Windows retained	36	30
Windows excluded	10–16, 18, 25–30	1–12, 14, 20, 27, 28, 31, 39, 42, 46
Window means, retained (ppm)	0.37–1.68	1.21–1.48
Window SD, retained (ppm)	0.002–0.038	0.011–0.046

2.5. Statistical Feature Extraction

Each retained window was summarized with four statistical features commonly used for electronic nose signals [22], [30]: the mean, sample standard deviation, minimum, and maximum, defined in Equations (1)–(4), where xi is the i-th reading in a window and n = 100. To assess whether feature extraction is needed, C4.5 was also trained on the mean feature alone and on individual raw readings of the retained windows (6,600 samples); under time-blocked validation, each reading was assigned to the fold of its window, and random cross-validation on raw readings was repeated three times.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (1)$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (2)$$

$$x_{min} = \min_{1\le i\le n} x_i \quad (3)$$

$$x_{max} = \max_{1\le i\le n} x_i \quad (4)$$

2.6. Classification Algorithms

C4.5 builds a decision tree by recursively selecting the attribute and split that maximize the gain ratio [31]. For a set of cases S with k classes, where pj is the proportion of class j, entropy is defined in Equation (5). The information gain of splitting S on attribute A into partitions S1, …, Sv is given by Equation (6), the split information by Equation (7), and the gain ratio by Equation (8). For continuous attributes, such as the features used here, C4.5 evaluates binary splits at candidate thresholds.

$$Entropy(S) = -\sum_{j=1}^{k} p_j \log_2 p_j \quad (5)$$

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{v} \frac{|S_i|}{|S|} Entropy(S_i) \quad (6)$$

$$SplitInfo(S,A) = -\sum_{i=1}^{v} \frac{|S_i|}{|S|} \log_2\left(\frac{|S_i|}{|S|}\right) \quad (7)$$

$$GainRatio(S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \quad (8)$$

C4.5 was run with the J48 implementation in Weka 3.8.6 [32] using the default pruning confidence (0.25) and a minimum of two instances per leaf. It was compared with six classifiers implemented in scikit-learn 1.4.2 [33]: (i) a single-threshold baseline, implemented as a depth-one decision tree on the mean feature; (ii) Gaussian Naive Bayes (NB); (iii) Random Forest (RF) with 200 trees and random_state = 42; (iv) k-nearest neighbors (kNN) with k = 5; (v) a support vector machine (SVM) with a radial basis function kernel; and (vi) logistic regression (LR). Features were standardized within each training fold for kNN, SVM, and LR. All other hyperparameters were left at their library defaults, and no hyperparameters were tuned, so test folds were never used for model selection. The time-blocked folds are deterministic. For stratified random cross-validation, the shuffling was repeated with ten random seeds (0–9), and results are reported as the mean over the seeds.

2.7. Evaluation

Because consecutive windows are correlated in time, the primary evaluation used time-blocked 5-fold cross-validation [20]. Fold k used windows 10k−9 to 10k of both sessions (k = 1, …, 5) as test data and the remaining windows as training data, so most training windows were temporally distant from the test windows, although windows at block boundaries remained adjacent. After screening, the five test blocks contained 9, 9, 12, 18, and 18 windows; the first block contained only

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (361-373)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.54559

background windows because exposure windows 1–10 were excluded. For comparison with common practice, stratified 5-fold cross-validation with random shuffling was also run with ten different random seeds.

Accuracy is reported as the mean and standard deviation over folds. Precision, recall, and F1-score were computed from the confusion matrix pooled over all test folds (Equations (9)–(12)), where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, with H_2S exposure as the positive class. Paired differences between C4.5 and each other classifier under time-blocked validation were tested with the exact McNemar test [34].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$F1 = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (12)$$

3. RESULTS AND DISCUSSION

3.1. Recorded Signals and Screening

Figure 4 shows both sessions on the same concentration axis. The background session started at 0.74 ppm and decreased to about 0.4 ppm in windows 4–9. Window 10 contained an abrupt peak of 12.45 ppm, after which the signal recovered slowly, with window means decreasing from 3.59 ppm in window 11 to 1.04 ppm in window 24. A second disturbance occurred in windows 25–30, and the signal then settled between 1.00 and 1.53 ppm. The exposure session ranged from 1.16 to 2.70 ppm; windows 1–12 contained repeated peaks, and windows 13–50 were largely stable, although windows 14, 20, 27, 28, 31, 39, 42, and 46 still exceeded the screening threshold; the retained exposure windows had means of 1.21–1.48 ppm.

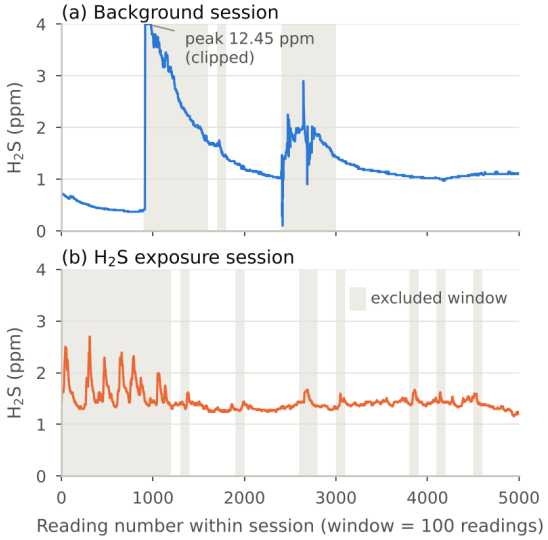


Figure 4. H_2S concentration in (a) the background session and (b) the H_2S exposure session. Shaded areas mark windows excluded by screening; values above 4 ppm are clipped

Before screening, 45% of the background readings were at or above the lowest exposure reading, and all 49 readings at or above 5 ppm occurred in the background session. The background session therefore contains at least one exposure-like event, although no H_2S sample was intentionally present. The background session was recorded before the exposure session in the same room, so the peak cannot be explained by residual gas from the H_2S sample, and its source could not be identified. The cause of the peak in window 10 and of the elevated windows 25–30 was not logged during data collection. The baseline of the session also rose from about 0.4 ppm at the start to about 1.1 ppm later, which indicates sensor drift or an unidentified H_2S source in the room rather than residual gas from the exposure sample, which had not yet been introduced. Figure 5 shows the window standard deviations used for screening. Excluded windows had standard deviations of 0.053–2.91 ppm, whereas all retained windows were at or below 0.046 ppm.

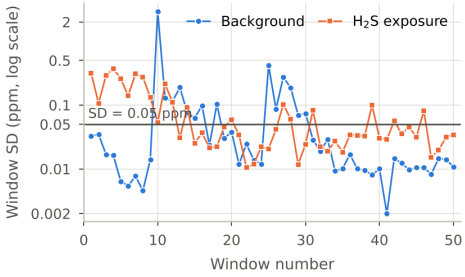


Figure 5. Standard deviation of each window and the screening threshold

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (361-373)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.54559

3.2. Extracted Features

Table 3 lists examples of the extracted features. After screening, background windows had means of 0.37–1.68 ppm and exposure windows 1.21–1.48 ppm. The classes overlap mainly in background windows 17, 19, 20, and 31–33, which belong to the recovery periods of the background session.

Table 3. Examples of extracted features (ppm)

Session	Window	Mean	SD	Min	Max
Background	1	0.672	0.033	0.62	0.74
Background	9	0.382	0.014	0.37	0.41
Background	17	1.679	0.023	1.64	1.73
Background	31	1.379	0.029	1.33	1.44
Background	50	1.105	0.011	1.10	1.14
H_2S	13	1.373	0.031	1.30	1.42
H_2S	25	1.337	0.021	1.30	1.37
H_2S	40	1.400	0.030	1.35	1.47
H_2S	50	1.212	0.034	1.16	1.28

3.3. Classification Performance

Table 4 summarizes the performance of the seven classifiers after screening. Under time-blocked validation, C4.5 achieved 87.8 ± 14.9% accuracy, with an H_2S recall of 0.967 and an F1-score of 0.879. The single-threshold baseline reached 88.9 ± 14.2%, NB 89.4 ± 8.9%, RF 87.2 ± 13.0%, kNN 85.6 ± 12.8%, and LR 83.3 ± 16.7%. These differences, at most 6.1 percentage points, are smaller than the fold-to-fold standard deviations, and the McNemar test found no significant difference between C4.5 and any of these classifiers ($p \geq 0.69$), although with 66 windows the test has limited power to detect small differences. The fold accuracies are also unstable because the test blocks contain only 9–18 windows and the first block contains only background windows, so its accuracy reflects specificity alone; the standard deviations of 9–17 percentage points reflect this instability rather than precise estimates. The similar performance of the classifiers should therefore be read as an absence of evidence for a difference, not as evidence that they perform equally.

Table 4. Classification performance after screening

	Time-blocked CV				Random CV
Classifier	Acc. (%)	P	R	F1	Acc. (%)
Threshold	88.9 ± 14.2	0.829	0.967	0.892	88.8 ± 8.9
C4.5	87.8 ± 14.9	0.806	0.967	0.879	87.9 ± 8.7
NB	89.4 ± 8.9	0.848	0.933	0.889	91.1 ± 7.5
RF	87.2 ± 13.0	0.824	0.933	0.875	89.2 ± 8.3
kNN	85.6 ± 12.8	0.818	0.900	0.857	87.3 ± 9.2
SVM	66.7 ± 39.1	0.651	0.933	0.767	90.7 ± 8.1
LR	83.3 ± 16.7	0.812	0.867	0.839	87.7 ± 8.5

Acc. = mean ± SD of fold accuracy. P, R, and F1 refer to the H_2S class and are computed from the pooled confusion matrix. Random CV = stratified 5-fold, 10 repeats.

SVM was the only exception (66.7 ± 39.1%, $p = 0.022$), although this difference would not remain significant after Bonferroni correction for the six comparisons ($\alpha = 0.0083$). It misclassified all nine windows of the first block, whose low background concentrations (0.37–0.67 ppm) lay outside the range of the background windows available for training. A flexible model can therefore extrapolate poorly when the test period differs from the training period, a failure that random cross-validation (90.7% for SVM) does not reveal.

Most errors were shared by all classifiers. Background windows 19, 20, and 31–33 were misclassified as H_2S by all seven classifiers, and window 17 by all except NB. These windows belong to the recovery periods after the exposure events in the background session, and their concentrations fall within the range of the exposure session.

3.4. Effect of Screening and Validation Protocol

Figure 6 compares accuracy with and without screening. Without screening, C4.5 reached 89.6% under random cross-validation but only 77.0% under time-blocked validation, with fold accuracies from 50% to 95%. The C4.5 tree trained on all 100 windows used only the minimum feature: it classified windows with a minimum between 1.20 and 1.57 ppm as H_2S and windows above 1.57 ppm as background, a rule that contradicts the physical meaning of exposure. The rule arose because the recovery windows of the background session had higher concentrations than the exposure session.

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (361-373)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.54559

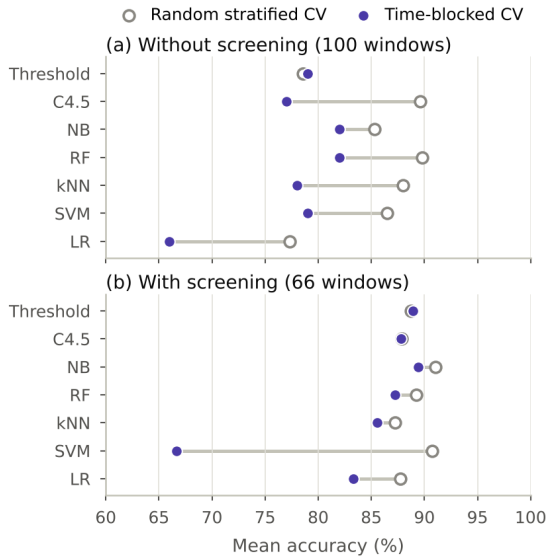


Figure 6. Mean accuracy under random stratified and time-blocked cross-validation (a) without and (b) with signal-stability screening

After screening, the gap between the two protocols nearly disappeared for C4.5 (87.9% under random and 87.8% under time-blocked validation). Table 5 shows that the time-blocked accuracy of C4.5 stayed between 87.7% and 89.3% for thresholds of 0.03–0.10 ppm, compared with 77.0% without screening. RF (87.2–90.6%) and NB (89.4–91.9%) were similarly stable, whereas LR and SVM varied more with the threshold. The large gap without screening is consistent with reports that random cross-validation overestimates performance on temporally structured data [20], [21], and it suggests that gas-sensor accuracies obtained with random k-fold cross-validation, such as the 91% reported for a virtual electronic nose [26], should be compared with time-aware estimates with caution.

The present setting also differs from other H_2S studies with metal-oxide sensors. Shaposhnik et al. [24] achieved selective detection of H_2S with a single metal-oxide sensor by temperature modulation, which increased the response to H_2S by two to four orders of magnitude compared with constant-temperature operation, and evaluated the sensor at known concentrations of 2–50 ppm in clean air. Viet et al. [25] identified H_2S and other noxious gases with an array of eight metal-oxide sensors (four MiCS-4514 chips) operating on the thermal fingerprint principle and measured at known concentrations, with quadratic discriminant analysis performing best among the compared machine learning models. Both studies used known gas concentrations and richer sensor responses, from temperature modulation in [24] and from multiple sensors with different sensitivities in [25], whereas the present study used a commercial MQ-136 module at a constant heater voltage, an uncalibrated concentration scale, and uncontrolled exposure in a room. Their results therefore address gas discrimination under laboratory conditions, whereas the present results concern whether a low-cost, constant-temperature sensor can separate exposure from background air in a setting where temporal correlation and transient responses dominate. Temperature modulation, as in [24], or additional sensors with different sensitivities, as in [25], could provide the additional response features that a single constant-temperature MQ-136 lacks.

Table 5. Time-blocked accuracy (%) for different screening thresholds

SD threshold (ppm)	0.03	0.05	0.08	0.10	None
Windows (bg/H_2S)	32/16	36/30	40/34	42/38	50/50
Threshold	90.7	88.9	84.6	83.7	79.0
C4.5	89.3	87.8	87.7	87.9	77.0
NB	91.7	89.4	89.9	91.9	82.0
RF	88.7	87.2	89.6	90.6	82.0
kNN	86.7	85.6	81.9	87.0	78.0
SVM	86.7	66.7	87.7	90.2	79.0
LR	77.7	83.3	75.0	76.6	66.0

3.5. Interpretation of the C4.5 Model

After screening, the C4.5 tree trained on all 66 windows contained a single split (Figure 7): windows with a maximum concentration of at most 1.20 ppm were classified as background, whereas windows above 1.20 ppm were classified as H_2S exposure. All 30 windows below the split were background windows, while 6 of the 36 windows above the split were background windows originating from the recovery periods (Figure 8a). This finding indicates that the C4.5 model relied primarily on the maximum concentration feature to distinguish background conditions from H_2S exposure. A similar application of C4.5 for H_2S leak detection was reported in our previous study [23], which achieved an average accuracy of 89% using five-fold cross-validation on a dataset of 10,000 records after statistical feature extraction. However, the results of the two studies should be interpreted

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (361-373)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.54559

in the context of their different dataset structures and evaluation settings. In particular, the 89% reported in [23] is close to the 89.6% obtained here for C4.5 on unscreened windows under stratified random cross-validation, whereas the same data yielded only 77.0% under time-blocked validation. Because [23] used five-fold cross-validation on temporally correlated records, its accuracy is likely to be optimistic for the same reason. The previous study employed a larger dataset consisting of balanced normal and hazardous records, whereas the present study used 66 screened windows derived from exposure and recovery sessions. Furthermore, the present analysis reveals that the resulting decision tree is equivalent to a single concentration-threshold rule, which explains why the single-threshold baseline achieved similar classification performance. The six background windows classified as H_2S exposure also indicate that residual sensor responses during recovery periods can influence threshold-based classification. Nevertheless, the resulting rule is computationally simple and suitable for direct implementation on the NodeMCU microcontroller. The findings therefore highlight the importance of considering recovery-related sensor responses when developing lightweight H_2S classification systems.

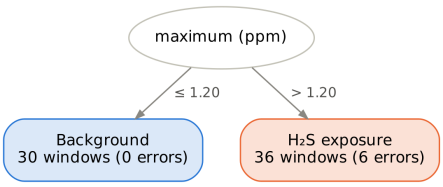


Figure 7. C4.5 decision tree trained on the 66 retained windows

Under time-blocked validation, C4.5 missed one H_2S window and flagged seven background windows as H_2S (Figure 8b). Under the conversion in Section 2.1, the 1.20 ppm split corresponds to a raw ADC count of 634, so windows classified as H_2S had a maximum ADC count of at least 635. Because the ppm values are uncalibrated, the split describes the sensor output of this device rather than a physical H_2S concentration, cannot be transferred to other devices without calibration, and should not be interpreted as an occupational exposure limit.

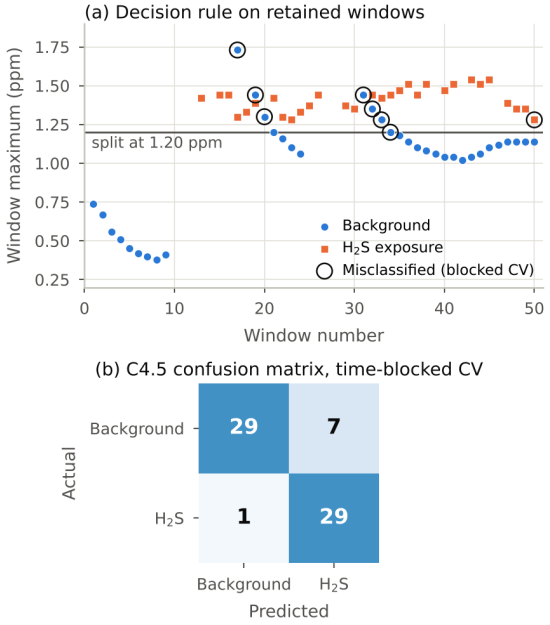


Figure 8. (a) Window maximum of the retained windows with the C4.5 split and the windows misclassified under time-blocked cross-validation; (b) pooled confusion matrix of C4.5

3.6. Contribution of Feature Extraction

Table 6 compares C4.5 trained on the four features, on the mean only, and on individual raw readings without feature extraction. Under time-blocked validation, raw readings gave 88.8% accuracy, compared with 87.8% for both the four features and the mean alone. Statistical feature extraction therefore did not improve accuracy in this dataset; its main benefit is that it summarizes each window into interpretable quantities. The higher accuracy of raw readings under random cross-validation (92.6%) is likely optimistic, because neighboring readings, which are nearly identical, are split between training and test folds.

Table 6. Accuracy (%) of C4.5 with and without feature extraction

Input to C4.5	Samples	Time-blocked CV	Random CV
Raw readings (no feature extraction)	6,600	88.8 ± 14.0	92.6 ± 0.7
Mean only	66	87.8 ± 14.9	88.3 ± 8.2
Mean, SD, min, max	66	87.8 ± 14.9	87.9 ± 8.7

Raw readings: each reading of the 66 retained windows is one sample; random CV for raw readings was repeated three times.

3.7. Limitations

This study has several limitations. First, each class was recorded in a single continuous session, so the difference between classes

cannot be fully separated from sensor drift or changes in room conditions over time; the drift of the background baseline from about 0.4 to 1.1 ppm is comparable to the difference between the classes [19]. Because the background session was always recorded first, session order is fully confounded with class: a classifier that separates the two sessions may respond to drift or to changes in room conditions rather than to the H_2S sample. Time-blocked folds reduce this risk only partly, because they compare the same within-session positions of both sessions but cannot remove the effect of absolute recording time. The reported accuracies therefore describe how well these two recordings can be separated and should not be read as evidence that the classifiers generalize to new sessions. The conditions during the background session, including the cause of the peak in window 10 and of the elevated windows 25–30 and the source of H_2S in a session recorded before the sample was introduced, were not controlled or logged, so this session cannot be considered a clean reference for ambient air. Second, the MQ-136 datasheet specifies a detection range of 1–200 ppm [35]. The exposure concentrations observed here (1.16–2.70 ppm), including the 1.20 ppm classification split, lie near the lower boundary of this range, where the converted concentrations are least reliable; part of the background readings (down to 0.37 ppm) even fall below it. In addition, no reference H_2S instrument was used, so the accuracy of the estimated concentrations was not established independently. Therefore, the results concern the presence of H_2S at low concentrations rather than hazard levels. The ADC reading was scaled with a 5 V full-scale reference although the NodeMCU analog input accepts 0–3.3 V, so the reported voltage, resistance ratio and concentration may be biased. Together with the assumption R0 = RL and the rescaled coefficient a, this means that the ppm values, the 1.20 ppm split, and the 0.05 ppm screening threshold are expressed on a device-specific scale rather than as calibrated concentrations. Because the conversion is monotonic, the ordering of readings and the single-split C4.5 and threshold rules are unaffected, but the window standard deviations used for screening and the feature scales used by distance-based classifiers would change if different conversion constants were used. Third, only one sensor was used, and temperature and humidity were not recorded. Fourth, screening excluded 34% of the windows, including exposure windows 1–12, which contain repeated peaks, and other transient periods that a real monitoring system must handle. The main accuracies therefore apply to steady-state windows only; the unscreened time-blocked accuracies in Table 5 (66.0–82.0%) indicate the performance when transient windows are included. In deployment, a window that fails the stability criterion would need separate handling, for example a provisional alert or deferring the decision to the next stable window. Future work should record multiple independent sessions per condition in alternating order, let the sensor return to its baseline between sessions, assign labels with a calibrated reference instrument, and evaluate models with leave-one-session-out validation.

CONCLUSION

This study evaluated classical machine learning for detecting H_2S exposure with a low-cost single-sensor electronic nose based on the MQ-136 sensor. After windows were screened by signal stability, C4.5 achieved 87.8% accuracy and 96.7% recall for H_2S under time-blocked cross-validation and learned a single interpretable rule based on the maximum concentration in a window. The other classifiers, except SVM, performed similarly, and none differed significantly from C4.5; a single concentration threshold performed comparably to the more complex models, although the small number of windows limits the power to detect differences. Statistical feature extraction did not improve accuracy compared with raw readings. Because each condition was recorded in only one session, these results show that the two recordings can be separated, but not yet that the models generalize to new sessions.

The evaluation protocol and the handling of transient signals strongly affected the reported results. Without screening, randomly shuffled cross-validation overestimated the accuracy of C4.5 by 12.6 percentage points and hid a physically inverted decision rule. Studies of low-cost gas sensors should therefore report time-aware validation, inspect the learned model, and collect multiple independent sessions per condition with reference measurements.

ACKNOWLEDGEMENTS

The authors would like to express sincere gratitude to the Directorate of Research, Technology, and Community Service (DRTPM), Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, through the Beginner Lecturer Research (Penelitian Dosen Pemula) under the 2024 Fiscal Year Funding Agreement, DIPA Number SP DIPA-023.17.1.690523/2024. Special thanks are also extended to Nahdlatul Ulama Sunan Giri University for the facilities and technical support provided during the research process through the Institute for Research and Community Service. A special appreciation goes to Pertamina Field Sukowati Bojonegoro for its valuable insights and knowledge shared during discussions and for its significant contributions throughout the research.

REFERENCES

[1] J. E. S. Pakpahan, W. Hasan, and I. Chahaya, "Analisa Kadar H2S (Hidrogen Sulfida) Dan Keluhan Kesehatan Saluran Pernapasan Serta Keluhan Iritasi Mata Pada Masyarakat Di Kawasan Pt. Allegrindo Nusantara Desa Urung Panei Kecamatan Purba Kabupaten Simalungun Tahun 2013," J. Kesehat. Lingkung. dan Keselam. Kerja, vol. 3, no. 2, pp. 1–9, 2014.

[2] A. W. Bande, "Analisis Penanganan Konsentrasi Hidrogen Sulfida Dalam Minyak Bumi Lapangan Tiaka Kabupaten Morowali Utara Provinsi Sulawesi Tengah," J. Geomine, vol. 4, no. 2, 2016, doi: 10.33536/jg.v4i2.51.

[3] T. L. Guidotti, "Hydrogen sulfide: Advances in understanding human toxicity," Int. J. Toxicol., vol. 29, no. 6, pp. 569–581, 2010, doi: 10.1177/1091581810384882.

[4] P. Migas, "PPSDM Migas Bekali ASN-nya tentang Penanganan Bahaya Gas H2S." Accessed: Mar. 29, 2023. [Online]. Available: https://ppsdmmigas.esdm.go.id/id/Landing/lihat_berita/ZDOCitR0

[5] H. AR and R. A. Palini, "Analisis Alat Pendeteksi Gas Hidrogen Sulfida Menggunakan Hazard and Operability Study Di Perusahaan Minyak Dan Gas," J. Tekno, vol. 19, no. 1, pp. 36–48, 2022, doi: 10.33557/jtekno.v19i1.1661.

[6] L. Capelli, S. Sironi, and R. Del Rosso, "Electronic noses for environmental monitoring applications," Sensors, vol. 14, no. 11, pp. 19979–20007, 2014, doi: 10.3390/s141119979.

[7] Y. Xie, C. Guo, S. Devahastin, L. Jiang, M. Du, and J. Yi, "Non-destructive determination of volatile compounds and prediction of amino acid nitrogen during sufu fermentation via electronic nose in combination with machine learning approaches," LWT, vol. 207, p. 116648, 2024, doi: 10.1016/j.lwt.2024.116648.

[8] M. A. Barata, Edi Noersasongko, Purwanto, and Moch Arief Soeleman, "Improving the Accuracy of C4.5 Algorithm with Chi-Square Method on Pure Tea Classification Using Electronic Nose," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 7, no. 2, pp. 226–235, 2023, doi: 10.29207/resti.v7i2.4687.

[9] A. Purnomo, M. A. Barata, M. A. Soeleman, and F. Alzami, "Adding feature selection on Naïve Bayes to increase accuracy on classification heart attack disease," J. Phys. Conf. Ser., vol. 1511, no. 1, 2020, doi: 10.1088/1742-6596/1511/1/012001.

[10] F. I. Silfana and M. A. Barata, "Using K-NN Algorithm for Evaluating Feature Selection on High Dimensional Datasets," J. Tek. Inform., vol. 17, no. 2, pp. 190–202, 2024, doi: 10.15408/jti.v17i2.40866.

[11] M. Ilmiyah, M. A. Barata, and P. E. Yuwita, "Implementation of ANN Optimization with SMOTE and Backward Elimination for PCOS Prediction," Sci. J. Informatics, vol. 12, no. 1, pp. 133–144, 2025, doi: 10.15294/sji.v12i1.22886.

[12] P. Gite, K. Chouhan, K. Murali Krishna, C. Kumar Nayak, M. Soni, and A. Shrivastava, "ML Based Intrusion Detection Scheme for various types of attacks in a WSN using C4.5 and CART classifiers," Mater. Today Proc., 2021, doi: 10.1016/j.matpr.2021.07.378.

[13] B. F. Tanyu, A. Abbaspour, Y. Alimohammadlou, and G. Tecuci, "Landslide susceptibility analyses using Random Forest, C4.5, and C5.0 with balanced and unbalanced datasets," Catena, vol. 203, Aug. 2021, doi: 10.1016/j.catena.2021.105355.

[14] X. Meng, P. Zhang, Y. Xu, and H. Xie, "Construction of decision tree based on C4.5 algorithm for online voltage stability assessment," Int. J. Electr. Power Energy Syst., vol. 118, Jun. 2020, doi: 10.1016/j.ijepes.2019.105793.

[15] S. Sundaramurthy and P. Jayavel, "A hybrid Grey Wolf Optimization and Particle Swarm Optimization with C4.5 approach for prediction of Rheumatoid Arthritis," Appl. Soft Comput. J., vol. 94, 2020, doi: 10.1016/j.asoc.2020.106500.

[16] J. Yan et al., "Electronic nose feature extraction methods: A review," Sensors, vol. 15, no. 11, pp. 27804–27831, 2015, doi: 10.3390/s151127804.

[17] Z. Ye, Y. Liu, and Q. Li, "Recent progress in smart electronic nose technologies enabled with machine learning methods," Sensors, vol. 21, no. 22, p. 7620, 2021, doi: 10.3390/s21227620.

[18] L. Feng, H. Dai, X. Song, J. Liu, and X. Mei, "Gas identification with drift counteraction for electronic noses using augmented convolutional neural network," Sensors Actuators B Chem., vol. 351, p. 130986, 2022, doi: 10.1016/j.snb.2021.130986.

[19] A. Vergara, S. Vembu, T. Ayhan, M. A. Ryan, M. L. Homer, and R. Huerta, "Chemical gas sensor drift compensation using classifier ensembles," Sensors Actuators B Chem., vol. 166–167, pp. 320–329, 2012, doi: 10.1016/j.snb.2012.01.074.

[20] D. R. Roberts et al., "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," Ecography (Cop.)., vol. 40, no. 8, pp. 913–929, 2017, doi: 10.1111/ecog.02881.

[21] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," Inf. Sci., vol. 191, pp. 192–213, 2012, doi: 10.1016/j.ins.2011.12.028.

[22] S. Wakhid, R. Sarno, S. I. Sabilla, and D. B. Maghfira, "Detection and classification of indonesian civet and non-civet coffee based on statistical analysis comparison using E-Nose," Int. J. Intell. Eng. Syst., vol. 13, no. 4, pp. 56–65, 2020, doi: 10.22266/IJIES2020.0831.06.

[23] M. A. Barata et al., "Hydrogen Sulfide Leak Detection Using the C4.5 Algorithm: Optimizing Feature Extraction For Enhanced Accuracy," AlGhazali Int. Conf., vol. 1, pp. 348–358, 2024.

[24] A. Shaposhnik, P. Moskalev, E. Sizask, S. Ryabtsev, and A. Vasiliev, "Selective detection of hydrogen sulfide and methane by a single MOX-sensor," Sensors, vol. 19, no. 5, p. 1135, 2019, doi: 10.3390/s19051135.

[25] N. N. Viet, P. H. Phuoc, L. V. Thong, N. V. Chien, and N. V. Hieu, "A comparative study of machine learning models for identifying noxious gases through thermal fingerprint measurements and MOS sensors," Sens. Actuators A Phys., vol. 375, p. 115510, 2024, doi: 10.1016/j.sna.2024.115510.

[26] G. Domènech-Gil and D. Puglisi, "A virtual electronic nose for the efficient classification and quantification of volatile organic compounds," Sensors, vol. 22, no. 19, p. 7340, 2022, doi: 10.3390/s22197340.

[27] D. R. Wijaya, R. Sarno, and E. Zulaika, "DWTLSTM for electronic nose signal processing in beef quality monitoring," Sensors Actuators, B Chem., vol. 326, no. September 2020, p. 128931, 2021, doi: 10.1016/j.snb.2020.128931.

[28] D. B. Magfira and R. Sarno, "Classification of Arabica and Robusta coffee using electronic nose," 2018 Int. Conf. Inf. Commun. Technol. ICOIACT 2018, pp. 645–650, 2018, doi: 10.1109/ICOIACT.2018.8350725.

[29] A. I. F. Al Isyrofie et al., "Odor clustering using a gas sensor array system of chicken meat based on temperature variations and storage time," Sens. Bio-Sensing Res., vol. 37, no. July, p. 100508, 2022, doi: 10.1016/j.sbsr.2022.100508.

[30] W. Harsono, R. Sarno, and S. I. Sabilla, "Recognition of original arabica civet coffee based on odor using electronic nose and machine learning," Proc. - 2020 Int. Semin. Appl. Technol. Inf. Commun. IT Challenges Sustain. Scalability, Secur. Age Digit. Disruption, iSemantic 2020, pp. 333–339, 2020, doi: 10.1109/iSemantic50169.2020.9234234.

[31] J. R. Quinlan, C4.5: Programs for Machine Learning. San Mateo, CA, USA: Morgan Kaufmann, 1993.

[32] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," ACM SIGKDD Explor. Newsl., vol. 11, no. 1, pp. 10–18, 2009.

[33] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[34] T. G. Dietterich, "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms," Neural Comput., vol. 10, no. 7, pp. 1895–1923, Oct. 1998, doi: 10.1162/089976698300017197.

[35] Zhengzhou Winsen Electronics Technology Co., Ltd., "Hydrogen Sulfide Gas Sensor (Model: MQ136) Manual," Version 1.6, 2021.