

Article

Assessment of Cooked Meatballs' Edibility Using Calibrated MOS Sensors and Microbiological Validation

Luigi Masi ^{1,2,*}, Revathy Gurusamy ^{2,3} , Daniel Garcia-Romeo ², Andreas Schütze ¹ , Rafael Pagán ³  and Christian Bur ¹ 

¹ Lab for Measurement Technology, Saarland University, Campus A5 1, 66123 Saarbrücken, Germany; schuetze@lmt.uni-saarland.de (A.S.); c.bur@lmt.uni-saarland.de (C.B.)

² BSH Electrodomésticos S.A., 50016 Zaragoza, Spain; rgurusamy@unizar.es (R.G.); daniel.garciad@bshg.com (D.G.-R.)

³ Departamento de Producción Animal y Ciencia de los Alimentos, Facultad de Veterinaria, Instituto Agroalimentario de Aragón (IA2), Universidad de Zaragoza, 50013 Zaragoza, Spain; pagan@unizar.es

* Correspondence: l.masi@lmt.uni-saarland.de

Abstract

Food waste is often driven by consumer uncertainty about the spoilage of stored food, especially for cooked meal leftovers where microbial growth is the main concern. We analyzed whether metal oxide semiconductor (MOS) gas sensors placed inside ordinary food containers can monitor the edibility of leftovers, specifically cooked meatballs. Sensors were operated using temperature cycling to enhance selectivity, and cycle-aligned features were extracted. A prior calibration campaign produced information used to map cycle-aligned features into estimated gas concentrations for relevant VOCs. Total viable counts, which represent the growth of total number of spoilage microorganisms, were analyzed on days 0, 5 and 7 to determine the food's freshness. Both the raw sensor features and the calibration-derived gas concentration estimates were analyzed with principal component analysis (PCA) and evaluated with a leave-one-sensor-out (LOSO) binary classifier for multiple food containers. PCA on the calibrated gas estimates revealed a dominant axis that consistently tracks food degradation over time across various containers. LOSO classification accuracy improved from 81.7% using raw sensor features to 87.8% using calibrated gas concentration estimates. These findings represent a proof of principle that calibrated MOS sensor systems can robustly support in situ edibility assessment for cooked food.

Keywords: gas sensors; MOS sensors; cooked food; edibility; temperature-cycled operation; food freshness monitoring; microbiological validation; sensor calibration



Received: 20 April 2026

Revised: 19 June 2026

Accepted: 23 June 2026

Published: 30 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Approximately one-third of the world's food is lost or wasted each year [1], with a sizable share discarded at the end of the supply chain, particularly in households [2]. At the end of the food supply chain, most of this waste comes from consumer uncertainty about the true condition of stored food and conservative disposal practices driven by “best before” or “use by” labels rather than objective freshness information [2]. Providing reliable, on-site measures of food freshness can therefore help consumers to make informed decisions, potentially leading to a decrease in unnecessary waste, and therefore enabling reuse strategies consistent with circular-economy principles [1].

Metal oxide semiconductor (MOS) gas sensors represent a promising technology for food freshness monitoring due to their low cost, compactness, and ability to provide continuous, real-time and in situ measurement [3–7].

The performance and selectivity of MOS sensors can be enhanced by temperature-cycled operation (TCO), which modulates the temperature of the sensitive layer, to reach greater selectivity in detecting various types of gases and expanding the quantitative ranges of detectable substances [8–10]. Prior studies have demonstrated that arrays of commercially available MOS sensors can detect gases emitted during spoilage of various foods (e.g., fruit) and can track temporal patterns that correlate with stages of edibility [11–17].

However, few studies [18] have examined cooked meals to assess whether low-cost gas sensors can estimate their edibility. Indeed, most studies on cooked food focus on formulation strategies, packaging methods, and cooking methods, all of which are aimed at controlling the food spoilage by microbial growth [19,20]. Hence, most physicochemical property changes in food matrices are indirectly correlated with microbial growth and spoilage.

Motivated by these challenges, in this work, we investigate whether MOS sensors deployed inside ordinary commercial food-storage containers can reliably identify the edibility state of cooked meatballs. This type of food was selected as a representative cooked-meal matrix because it is an international dish; it combines key components found in many prepared meals (minced meat, fat, moisture and spices) and thus encompasses the chemical and microbiological processes that drive spoilage in cooked foods.

From an application perspective, meatballs are widely consumed in many regions and commonly sold as ready-to-eat or precooked products, so methods that can reliably assess their edibility have high relevance for consumers.

The experimental protocol consisted of three integrated elements: (i) deployment of the MOS sensor, in controlled temperature-cycled operating mode, within sealed, commercial food containers filled with meatballs; (ii) parallel microbiological analysis to quantify total viable counts (TVCs) growth; and (iii) a comprehensive data processing pipeline that, starting from raw sensor signals, extracts cycle-aligned features and culminates in a projection of field measurements into a calibration-informed gas space that reconstructs the temporal evolution of the food and a supervised classifier allowing binary edibility decisions (edible/not edible).

By explicitly focusing on a cooked meat application and by validating sensor outputs against microbiological ground truth in consumer-relevant containers, this work extends the scope of MOS-based freshness sensing toward products and use cases that are of direct practical concern to consumers. The results therefore not only contribute to the sensor and data analysis methodology but also provide concrete proof of principle for translating sensor research into household freshness aids that could reduce uncertainty-driven food waste.

2. Materials and Methods

Over the course of a single controlled campaign, we instrumented consumer-grade food storage containers with MOS sensors to follow the evolution of cooked meatballs as they age under refrigerated storage. In this section, the experimental setup and sample preparation, the microbiological analysis, the sensor hardware and operating conditions, the pre-deployment calibration procedure, the data pipeline of the calibration procedure, the field data processing and, finally, the supervised and unsupervised analysis are described.

2.1. Experimental Setup and Sample Preparation

The cooked meatballs were purchased from a local supermarket at 11:00 on the first day of the experiment. Meatballs from a local supermarket were chosen to increase the experimental reproducibility and external relevance, since they are prepared according to controlled, documented recipes and cooking conditions and therefore exhibit much lower batch-to-batch variability than ad hoc self-made kitchen preparations. The tested meatballs were composed of pork and beef meat that had been fried and cooked in an almond sauce. A total of approx. 5.25 kg of meatballs was obtained. Immediately after purchase, the supermarket trays were emptied into a single mixing pot and manually homogenized to promote even distribution of the product. Homogenized food samples were then portioned into identical experimental containers: each container received 350 g of meatballs and sauce with an estimated filling precision of ± 5 g. The moment that containers were closed and sensors started to measure is defined as the start of day 0; this occurred approximately 4 h after the purchase.

The experimental samples were stored in commercially available containers (IKEA, Älmhult, Sweden), which represent a good equilibrium between the accuracy that laboratory glass containers may provide and what people may have at home: the heat-resistant glass containers present an internal volume of 1.8 L and polypropylene lids with silicone seals. Sensor cables and sensors were sterilized with UV light before insertion to avoid contamination. Nine out of fifteen containers were instrumented, with a single MOS sensor installed in each. Sensors were introduced through the lid (via their cable) and positioned in the headspace beneath the lid, adjacent to an upper corner; sensors were therefore exposed to container air only to sample volatiles released from the food but were not in physical contact with the food to avoid contamination. The cable diameter and arrangement allowed the lid to close while maintaining an overall hermetic seal of the container headspace.

All containers were placed within a commercial refrigerator, as can be seen in Figure 1. The refrigerator temperature was set to 6 °C via the appliance controls. During the experiment, the fridge was opened on a daily basis for visual inspection, and, during the biological sampling day, designated food containers were extracted from the fridge and replaced after the biological analysis. Thus, the experimental procedure closely mimicked the food storage practices of consumers.



Figure 1. Fridge with the food containers.

2.2. Microbiological Analysis

The experimental design consists of two setups: a discrete analysis with 12 food containers and a continuous analysis with 3 food containers. These setups were designed

to mimic common household food storage practices; specifically, whether consumers repeatedly take portions from the same container (continuous use) or consume food from separate, single-use containers (discrete use).

Quantitative microbiological analysis was performed for both setups on days 0, 5 and 7. At each sampling point, three containers from the discrete setup and three containers from the continuous setup were removed from the refrigerator for the microbiological analysis.

Discrete food containers, except those discarded after the microbiological analysis at day 0, were retained with their respective sensors for sensor data collection. For both samples, TVC growth [21,22] was quantified at each sampling point to estimate food quality.

As no specific microbiological criteria exist for household-prepared refrigerated foods, microbiological criteria for ready-to-eat food regulations were considered [23–26]. Since the microbiological thresholds vary slightly across the food categories, more stringent limits were applied. Hence, a TVC below 6 log (CFU/g) was considered the upper limit of marginal acceptable microbiological quality of the food. The experiment was conducted until the microbiological spoilage threshold was exceeded.

2.3. Sensor Hardware and Operating Conditions

A total of 9 sensors were used for this analysis and have been deployed in 9 different containers: one continuous food container, two containers related to the discrete analysis of day 5, three containers related to the discrete analysis of day 7 and three containers retained just for sensor data collection. The last three were prepared using food samples from the same batch to ensure consistency. The MOS sensor used in this study was the commercially available SGP40 [27] which contains four individual MOS sensing layers. The first three layers of the MOS sensors were run in temperature-cycled operation [28] which comprises twelve temperature jumps from high to low temperature, in the range 100 °C to 400 °C, as represented in Figure 2. In the high-temperature phase, the sensor was heated at a maximum temperature of 400 °C for 5 s; after the high-temperature phase, the sensor was operated at various low temperatures for 7 s. The temperature of the low-temperature phase increased by 25 °C (starting from 100 °C) after each high-temperature phase. The total length of each cycle was 144 s and data were recorded at a sample rate of 10 Hz. An example of the logarithmic conductance of one sensor element that responds to this temperature cycle is shown in Figure 3.

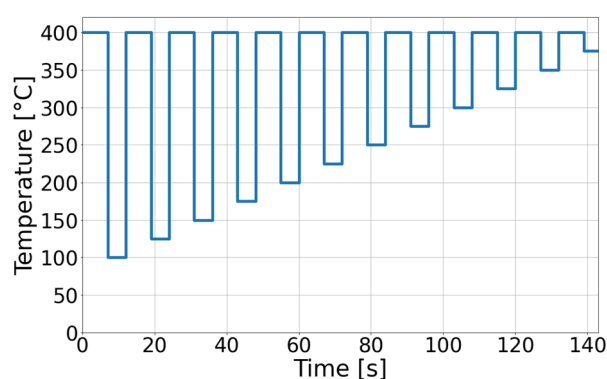


Figure 2. Temperature cycle profile of the first three sensitive layers.

The last layer of the SGP40 sensors was operated with a different temperature profile, described in [28]: here, the temperature cycle repeats the same high and low temperature levels. The high temperature is always set to 300 °C for 5 s, and the low temperature is always set to 250 °C for 7 s.

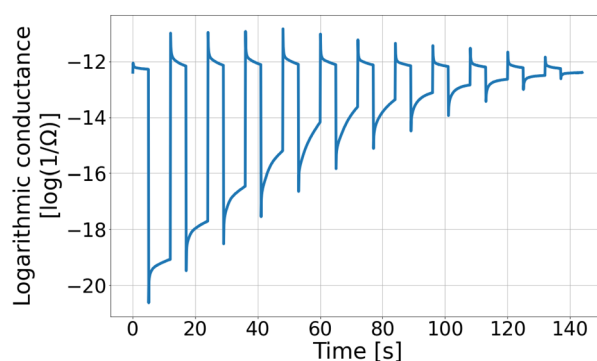


Figure 3. Example of the logarithmic conductance of the first sensing layer of the SGP40.

2.4. Calibration Procedure and Data Pipeline

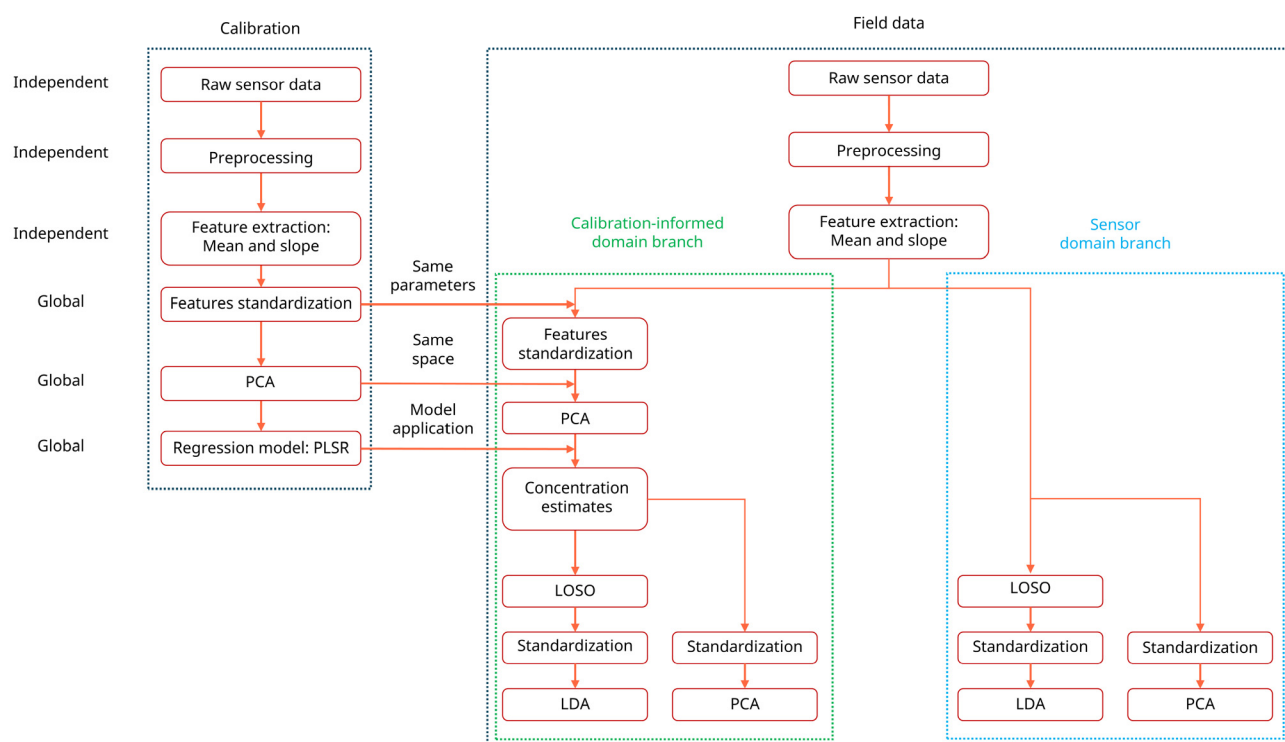
A calibration was performed prior to the food experiments with the aim of calibrating the sensors to reduce sensor-to-sensor variability and improve the comparability of headspace measurements in a predefined and controlled environment before using the sensors in a real-world scenario. This procedure was not intended to provide chemically rigorous quantification of every spoilage compound in complex food matrices; rather, it supports classification by mapping processed sensor data into a calibrated representation built from controlled VOC mixtures. Moreover, this procedure takes into account potential interference from other environmental factors such as humidity fluctuations, which could otherwise affect the readings.

The calibration was performed with a custom-built gas-mixing apparatus, described in detail in [29]. The sensors were exposed to known concentrations of thirteen gases. Based on previous studies regarding typical gases emitted by food and typical interfering gases omnipresent in ambient air such as CO, H₂ and humidity [30], a list of gases and concentration ranges was selected for the calibration as shown in Table 1. Practical constraints, like gas bottles' availability, influenced the selection of calibration compounds, and therefore the calibration set did not include certain meat-spoilage markers. Instead, the resulting gas sensor responses were used to derive a multivariate transformation that maps sensor outputs from the instrument (resistance/processed-features) domain into a representation built around those calibration VOC mixtures, which is therefore referred to as "VOC-space". Note that the VOC mixtures contain representatives of relevant VOC classes, thus allowing an application-independent calibration. For the calibration process, over 230 unique gas mixtures, each 25 min long, were randomly created by using Latin Hyper Cube sampling based on the chosen gases and their defined concentration ranges to achieve a robust machine learning model for quantification of all gases [31].

After the calibration, a specific data workflow, represented in the left branch of Figure 4, was followed: raw resistance signals were converted to conductance ($G = 1/R$) and transformed to a natural logarithm of conductance ($\log G$), which constitutes the pre-processing part. Because the sensors were operated with temperature cycles of 144 s, each cycle was partitioned into 144 equidistant segments of 10 samples which corresponded to segments of one second each. For every segment, we computed two descriptive sensor features: the arithmetic means and the slope (time derivative) across the segment. For each gas-sensitive layer, these two operations produce 288 features per cycle: given the SGP40's four sensing layers, each cycle yields 1152 features. These steps were operated on each sensor independently, as indicated by the text reading "independent" in the scheme of the dataflow.

Table 1. List of substances with their respective concentration ranges included for the random gas mixtures of the calibration procedure.

Compound	Concentration	
	Minimum	Maximum
R.H. at 23 °C	25%	70%
Carbon monoxide	100 ppb	2000 ppb
Hydrogen	25 ppb	1000 ppb
Acetaldehyde	10 ppb	1000 ppb
Acetone	10 ppb	1000 ppb
Ammonia	10 ppb	1000 ppb
Ethanol	10 ppb	1000 ppb
Ethyl acetate	150 ppb	1000 ppb
Formaldehyde	10 ppb	1000 ppb
Isopropanol	10 ppb	1000 ppb
Limonene	50 ppb	800 ppb
Methanol	10 ppb	1000 ppb
n-Hexane	10 ppb	1000 ppb
Toluene	10 ppb	1000 ppb

**Figure 4.** Representation of the complete data workflow.

Feature matrices from all sensors were pooled into a single global calibration dataset, which is indicated by the text “global” in the scheme of the dataflow. Columns of the global dataset were standardized by z-scoring (subtracting the column mean and dividing by the column standard deviation). The computed column means and standard deviations were retained as calibration parameters for later reuse. Principal component analysis (PCA) was performed on the standardized global calibration matrix and a 20-component PCA space was preserved as the calibration PC basis. These principal components were then used as inputs to train separate partial least squares regression (PLSR) models, one for each target gas, using a hold-out split scheme in which 20% of the dataset was reserved for testing.

In all the phases, the samples were divided based on the unique gas mixture for which they were recorded, i.e., group-based, as explained in [32]. This achieves more reliable and robust results as the test data are based on gas mixtures that are not contained in the training data.

All exported calibration parameters (column-wise mean and standard deviation, PCA space, and the trained per-gas PLSR models) were saved and cataloged for direct application in the field data pipeline.

Field acquisition replicated the calibration acquisition settings: raw resistance values were recorded under the same TCO profile and sampled at the same acquisition rate. Field preprocessing followed the calibration pipeline up to feature extraction. After the feature extraction, the field data pipeline, depicted in Figure 4, splits into two complementary branches.

In the calibration-informed domain branch, field sensor features are processed by reusing the parameters obtained during the calibration campaign: each field feature column is first standardized using the column means and standard deviations computed on the calibration dataset (i.e., calibration z-scoring), and the resulting standardized features are projected onto the calibration PCA basis via the stored loadings. The trained global PLSR regressors were applied to the projected scores on a per-segment, per-sensor basis to obtain gas concentration estimates. Stacking these outputs produced a gas space representation in which each cycle is described by one value per gas. As thirteen gases were used in the calibration, each sensor output corresponds to a vector of thirteen features in the gas space.

For exploratory visualization, PCA was performed on the assembled gas space dataset after column-wise standardization (z-scoring of the gas columns across the pooled field samples) to inspect temporal trends and cluster structure. For supervised evaluation we implemented a device-level generalization test using a leave-one-sensor-out (LOSO) protocol across all sensors: for each fold, one sensor was held out as the test device and the remaining sensors provided the training set. In each LOSO fold, the training features were standardized column-wise (means and standard deviations computed on the training set), and the same parameters were applied to the features of the held-out sensor. A linear discriminant analysis (LDA) classifier was then trained on the standardized training features and then used to predict labels for the excluded sensor. In a first step, a simple two-class classification (edible vs. not-edible) was tested, which can be extended to allow a more comprehensive assessment of the food state.

In the second branch of the data workflow designated the “sensor domain branch”, the sensor features (means and slopes of log conductance for each segment of each sensor layer) are used directly. As with the calibration-informed gas space branch, LOSO evaluation is enforced: for each LOSO fold, column-wise means and standard deviations are computed on the training sensors only; those training set parameters are then used to standardize both the training data and the held-out sensor prior to testing. LDA is then trained on the standardized training set and evaluated on the excluded sensor.

PCA was also performed on the sensor domain features for exploratory inspection, with pooled standardization applied for visualization purposes.

In summary, both branches (calibration-informed domain and sensor domain) support unsupervised exploration (PCA visualizations) and supervised evaluation (LOSO-LDA). The leave-one-sensor-out (LOSO) protocol allows us to predict the performance of a generic machine learning model for new sensors.

3. Results and Discussion

3.1. Microbiological Analysis

The results of the microbiological analysis are shown in Table 2 for the discrete samples and in Table 3 for the continuous samples: the tables show the TVC for each sample food (also indicated as a replica) on each sampling day and, as a final row, the average of the values of the replicas on the same sampling day along with the standard error, which is also graphically represented in Figure 5.

Table 2. Results of the discrete samples TVC values over the sampling days.

Replicas	Days		
	0	5	7
TVC Values [log (CFU/g)]			
R1	n.d.	6.45	9.19
R2	n.d.	6.51	9.00
R3	n.d.	6.29	9.16
Average $\pm$ standard error	n.d.	6.42 $\pm$ 0.06	9.12 $\pm$ 0.05

Table 3. Results of the continuous samples TVC values over the sampling days.

Replicas	Days		
	0	5	7
TVC Values [log (CFU/g)]			
R1	n.d.	6.45	9.26
R2	n.d.	6.34	8.68
R3	n.d.	6.20	9.03
Average $\pm$ standard error	n.d.	6.33 $\pm$ 0.06	8.99 $\pm$ 0.14

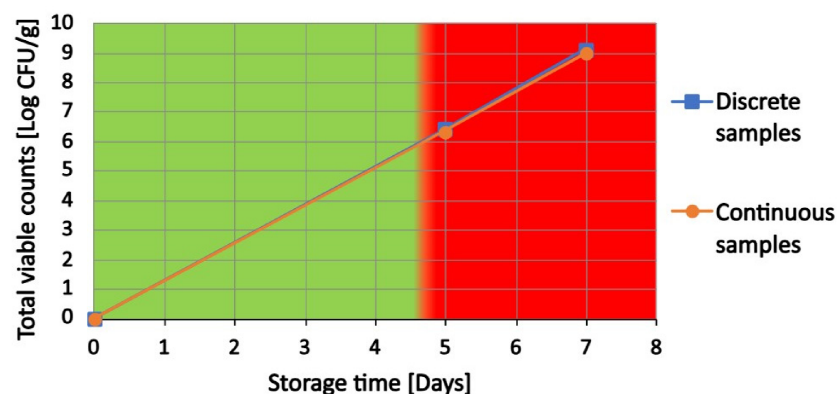


Figure 5. Evolution of TVC in refrigerated cooked meatball samples over the storage trial, with the discrete samples (blue) and the continuous samples (orange). The X- and Y-axis depict storage time (days) and mean TVC values (log CFU/g) with standard error, respectively.

On day 0, the TVC was below the detection limit (not detectable, indicated as n.d. in Tables 2 and 3); by day 5, the average TVC increased to 6.42 $\pm$ 0.06 log (CFU/g) in the discrete samples and 6.33 $\pm$ 0.06 log (CFU/g) in the continuous samples and on day 7, both types of samples reached around 9 log (CFU/g) which is far beyond the selected threshold limit.

In addition, no significant difference in the TVC was observed between continuous and discrete samples, so the reopening of food containers did not influence the development of microbial growth in this experiment. By day 5, the TVC had exceeded the microbiological

threshold of 6 log CFU/g adopted in this study (refer to Section 2.2). Accordingly, the binary decision boundary (green as edible, red as non-edible) was set between days 4 and 5.

Sensor data were labeled according to the first microbiological milestone: sensor data were labeled as “edible” if they were collected before the first biological measurement (TVC) reaching or exceeding a TVC of 6 log (CFU/g), and “non-edible” if collected at or after that timepoint. This constitutes a milestone-based label assignment tied to the measured TVC threshold. Photographic documentation of the product was collected throughout the campaign, as shown in Figure 6, to provide a visual correlation of microbiological data and sensor data. Representative images taken on day 0, day 5 and day 7 show progressive surface deterioration of the food matrix. Notably, a gray viscous film forms on the food surface and the overall color becomes markedly greyer by day 5–7. These images further support the microbiological evidence of food quality deterioration.

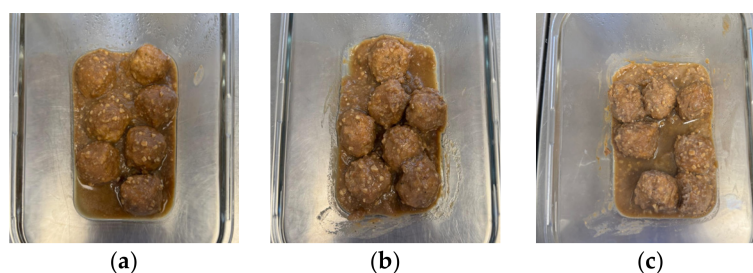


Figure 6. Pictures of the containers on the sampling days: (a): day 0, (b): day 5, (c): day 7.

3.2. Calibration

As described before, the sensors were calibrated prior to their deployment in the food containers. The regressors trained during the calibration campaign were global PLSR models that mapped standardized, PCA-projected sensor features to per-gas concentration estimates. The results of the regressor models built during the calibration are shown in Table 4: for each gas, the table lists the test root mean square error (RMSE) in ppb and the relative accuracy computed as $\text{RMSE}/(\text{max} - \text{min})$, expressed as a percentage, over the calibration range for that gas. These calibration diagnostics set an expectation of achievable accuracy when applying the models to field recordings. While the relative accuracy seems poor, with values between 15 and 30%, these values fall within the broader range of variability reported for analytical chemistry methods, i.e., sampling followed by GC-MS analysis as a gold standard [33]. It is important to note two limitations that have an impact on the interpretation of Table 4 and on field gas concentration estimates in general. First, the calibration mixtures were selected using literature-informed ranges for likely food and ambient-related gases rather than from prior, product-specific quantification of gases resulting from spoilage processes; consequently, the regressors provide a broad, useful mapping but are not chemically anchored to specific spoilage processes for this particular product. Second, despite the calibration mapping, residual sensor-to-sensor variability remains (as explored in the PCA and in LOSO classification), indicating opportunities for improved calibration protocols (e.g., targeted exposures to spoilage volatile organic compounds, transfer learning or domain adaptation methods [34]) to reduce inter-device bias.

Table 4. Performance of the regressors built during the calibration phase. The results are reported in terms of RMSE (ppb) and relative accuracy (%).

Compound	Error	
	RMSE [ppb]	Rel. Acc. [%]
Carbon monoxide	355	19
Hydrogen	290	28
Acetaldehyde	233	23
Acetone	160	16
Ammonia	186	19
Ethanol	177	18
Ethyl acetate	127	15
Formaldehyde	283	30
Isopropanol	278	28
Limonene	114	15
Methanol	168	17
n-Hexane	283	29
Toluene	221	22

3.3. Unsupervised Learning

The TVC values, collected at scheduled sampling days, were used to derive binary labels for supervised experiments: segments occurring after the initial stabilization window (first 12 h after deployment) and up to day 4 were labeled “green” (edible), and segments from day 5 onward were labeled “red” (not edible). To mitigate ambiguity at the microbiological threshold crossing, for supervised analysis, an additional transition window spanning the last 6 h of day 4 and first 6 h of day 5 was excluded. For the exploratory PCA on the sensor features and on the gas concentration estimates, only the first 12 h were excluded; the transition window between day 4 and day 5 was not excluded for the PCA to retain as much temporal information as possible for visualization. Some data gaps occurred during deployment (early interruption and intermittent sensor outages), which fragmented parts of the time series and motivated the conservative exclusion policy described above.

After preprocessing and exclusion windows, the average amount of samples per sensor is approx. 1982 samples for the supervised learning analysis, of which approx. 1144 segments are labeled green (edible) and 838 segments are labeled red (not edible).

For exploratory analysis, we performed principal component analysis, cf. Figure 7, on both the calibration-informed domain based on gas concentration estimates (a) and the sensor domain features (b). The color indicates the sampling day: dark purple, corresponding to the first day, transitions to yellow, corresponding to the last day. Both PCA views reveal a pronounced temporal organization, but they differ in the clarity of the shared spoilage trajectory and in the degree of residual device specific structure: when PCA is performed on the calibration-informed domain based on gas concentration estimates, the first two components explain a large fraction of variance ($PC1 = 60.37\%$, $PC2 = 25.61\%$). The pooled $PC1$ – $PC2$ scatter shows that samples from different containers, despite occupying different positions at day 0, progressively converge and then translate coherently along $PC1$ as storage time increases. This behavior supports the interpretation that the $PC1$ in the calibrated gas space grasps the aging of the food: the dominant mode of variability corresponds to a common temporal trend across food containers, while $PC2$ captures secondary variation, e.g., residual sensor differences, minor compositional heterogeneities or transient environmental effects. The PCA in the sensor domain yields comparable explained variance for the first two components ($PC1 \approx 62.7\%$, $PC2 \approx 23.1\%$) and likewise shows a time-ordered structure, but the pattern is less compact and exhibits more pronounced sensor-specific clustering and vertical spread. In the sensor domain plot, individual sensors often appear

as distinct branches or offset clusters, particularly at early times, and trajectories display greater inter-device dispersion. This increased spread reflects the fact that sensor domain features still contain device-specific information that the calibration procedure is designed to reduce.

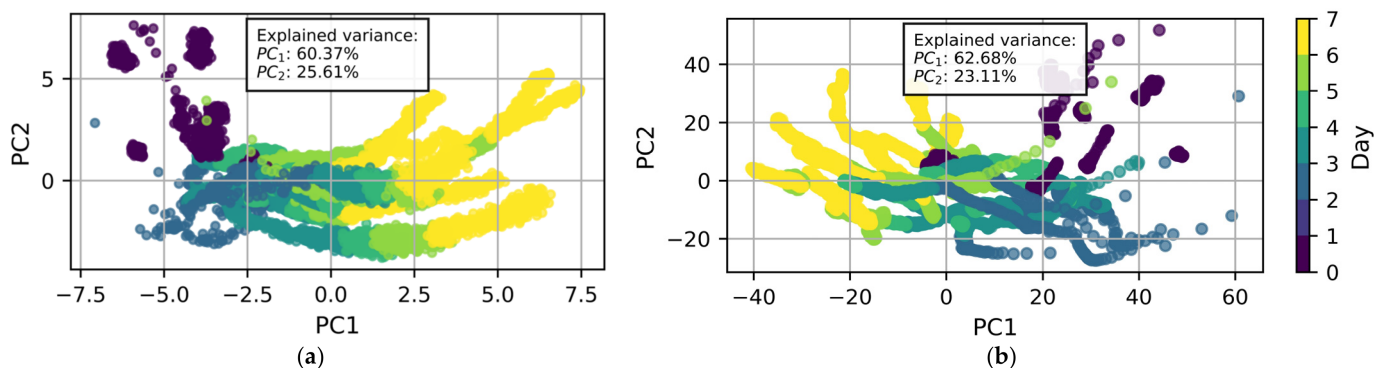


Figure 7. PCA of all sensors in the calibration-informed domain (a) and in the sensor domain (b).

Taken together, these results indicate that the temporal signal associated with spoilage is present in both representations, but that the calibration-informed mapping produces a cleaner, more device-agnostic manifestation of that signal.

Overall, PCA based on the gas concentration estimates provides a compact, low-dimensional visualization of headspace chemical evolution and motivates the subsequent supervised analysis by highlighting a consistent, time-ordered signal across devices.

3.4. Supervised Learning

As described before, to assess the ability of simple classifiers to generalize across devices, we implemented a leave-one-sensor-out evaluation and trained a linear discriminant analysis classifier under two feature representations—the sensor domain one and the calibration-informed one. LOSO was applied across all usable sensors so that, in each fold, the classifier was trained on data from all but one sensor and tested on the held-out device. The confusion matrices in Figure 8 show the sum of all the sample predictions, labelled in green for edible samples and red for non-edible samples, from each LOSO test fold for the LDA classifier applied to the calibration-informed domain (a) and to the sensor domain (b).

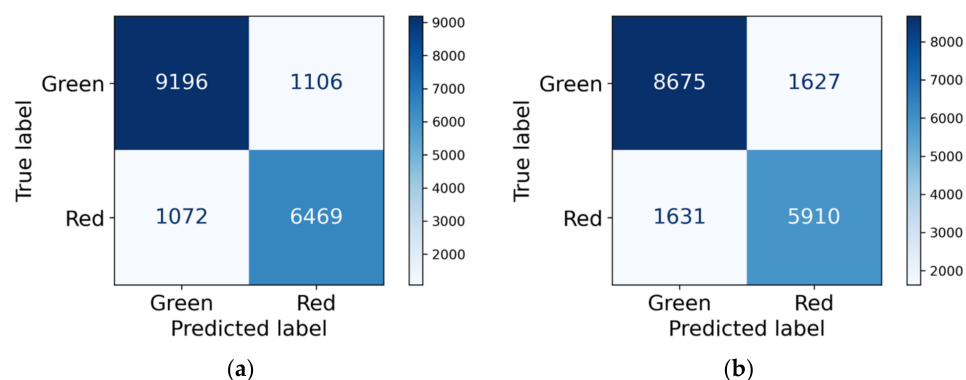


Figure 8. Confusion matrix of the LDA applied to features from the calibration-informed domain (a) and the sensor domain (b) of all sensors.

Under this protocol, LDA trained on sensor domain features achieved an overall accuracy of 81.7% across the LOSO folds, while LDA trained on the calibration-informed gas space features (i.e., the gas concentration estimates) reached 87.8% accuracy. A full set of performance metrics—accuracy, sensitivity, specificity, precision, F1 score and the

false-green (false-negative) rate—for both the calibration-informed domain and the sensor domain are reported in Table 5. The observed improvement indicates that the calibration mapping and gas space representation reduce inter sensor variability relevant to the classification task and consequently improve cross-device generalization. These results support the use of a calibration-informed, global modeling approach when the objective is to deploy a single classifier across multiple, nominally identical MOS devices.

Table 5. LOSO performance of LDA classifiers for the calibration-informed domain and the sensor domain.

Metric	Calibration-Informed Domain [%]	Sensor Domain [%]
Accuracy	87.8	81.7
Sensitivity	85.8	78.4
Specificity	89.3	84.2
Precision	85.4	78.4
F1-score	85.6	78.4
False green rate	14.2	21.6

3.5. Insight Single Device

To illustrate the practical impact of the calibration at the device level and to show the agreement between sensor-derived signals and microbial growth, we present a detailed case study for one specific sensor (sensor A). Figure 9 displays the confusion matrices for this sensor under two LDA configurations: (a) classifier trained and applied in the calibration-informed domain and (b) classifier trained and applied in the sensor domain. The two confusion matrices in Figure 9 summarize all sample predictions obtained for this sensor within its LOSO test fold, i.e., the data recorded for this sensor are not included in the model training.

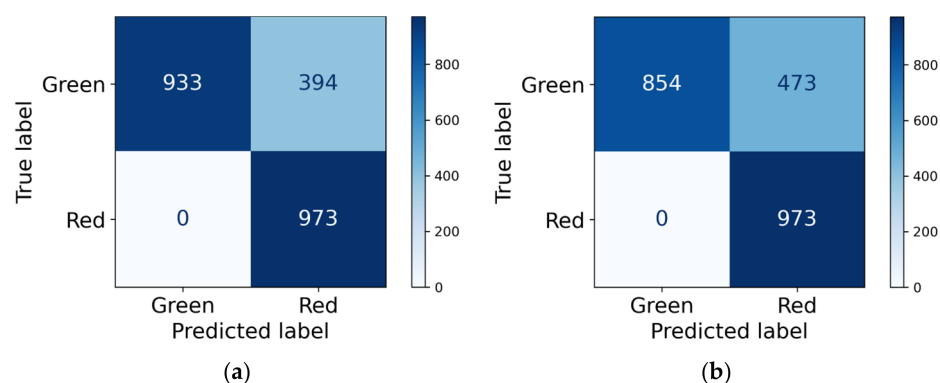


Figure 9. Confusion matrix of the LDA applied to features from the calibration-informed domain (a) and sensor domain (b) for sensor A.

In both representations, the classifier correctly labeled red samples, i.e., non-edible samples were misclassified as edible for this device under either feature representation. Differences between domains are evident for the green class: in the sensor domain, the classifier produced 854 true green predictions and 473 false red predictions, whereas in the calibration-informed domain, the number of true green predictions increased to 933 and false red predictions decreased to 394. For this sensor, the domain shift introduced by calibration therefore improved true green identification by 79 samples while maintaining perfect detection of red samples.

Figure 5 places these classification outcomes in the context of the food container's microbial growth. The measured TVC growth, which is also reported in Table 2 for replica R2, for this food container was not detected on day 0. It reached 6.51 log (CFU/g) on day 5

(exceeding the 6 log (CFU/g) spoilage threshold) and culminated with 9 log (CFU/g) on day 7.

In the single-container view, as shown in Figure 10, both the sensor domain and the calibration-informed representations exhibit a clear, time-ordered trajectory, but they convey complementary information. Projecting this food container's segments into the calibration-informed domain PCA space (a) produces a smooth rightward progression along PC1 from day 1 through days 5 and 7, consistent with the pooled multi-sensor observation that PC1 captures the dominant temporal change associated with the decrease in freshness of the food. When the same food container is examined in the sensor domain PCA (b), a similarly ordered temporal path is visible at the single-device level, indicating that the principal temporal signal is present in the uncalibrated features.

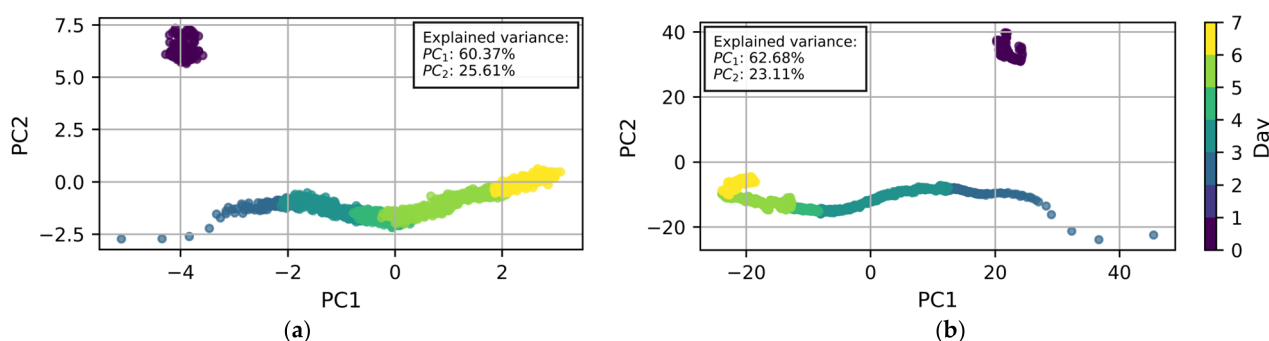


Figure 10. PCA of the features from the calibration-informed domain (a) and sensor domain (b) for sensor A.

However, the two projections differ in one important aspect: the calibration-informed domain PCA places the food container trajectory into a common, device-agnostic space where trajectories from different containers are more directly comparable, whereas the sensor domain PCA retains device-specific offsets and scaling that manifest as larger inter-device dispersion and branching in the pooled plot, cf. Figure 7.

These observations imply that calibration improves cross-device alignment of the same underlying temporal signal but is not strictly required to observe a time-ordered spoilage trend for an individual sensor. At the same time, interpretation must remain cautious: the calibration-informed domain was derived from a calibration campaign that did not include quantified exposures for every specific spoilage volatile organic compound (VOC) emitted by these meatballs. Therefore, while PC1 provides a reproducible, low-dimensional descriptor headspace's dynamic of the food container and correlates well with the observed microbiological deterioration, we cannot unambiguously attribute PC1 variation to one particular chemical species or microbiological processes without targeted chemical identification or product-specific calibration.

4. Conclusions

This study provides a proof of principle that MOS sensor systems deployed inside ordinary consumer containers can detect headspace changes that relate to spoilage progression of a cooked-meat product, as validated by microbiological analysis, and allow edibility classification.

In an unsupervised analysis, PCA on calibration-informed gas concentration estimates revealed a dominant temporal axis (PC1) along which samples from different containers align as spoilage advances.

In a supervised evaluation, an LDA classifier tested under a leave-one-sensor-out protocol achieved 81.7% accuracy on features from the sensor domain which improved to

87.8% accuracy for the calibration-informed domain, proving that the calibration improved cross-device generalization and reduced false red alerts in the examined cases.

Cooked foods present a distinct challenge relative to raw items: VOC emissions reflect overlapping chemical and biological processes. Consumers care foremost about biological safety for cooked meals, not about abstract chemical markers. By validating sensor outputs directly against microbiological measures in consumer-relevant containers, this work moves MOS sensing from laboratory demonstrations toward applications that address real consumer concerns for cooked meals.

Nevertheless, achieving practical robustness, reliability and generalizability will require systematic consideration of many factors: generalization ability under different batches, aging conditions, and environmental conditions, as well as multiple brands and varied cooking methods so that the approach can be validated and applied to other products. This work represents a first step toward the development of a methodology that could be adapted to a broad range of cooked meals using a common sensor platform together with appropriate data processing.

Following this path, future works aim to inform users in time for them to use the food safely before it is spoiled. This objective goes beyond the capabilities of the human nose: while humans can often recognize spoilage, they cannot reliably predict the remaining safe lifetime of a prepared food.

Although we do not report GC–MS compound identification in this campaign, the combination of a calibration pipeline (to reduce device bias) and microbiological validation provides the operational evidence required to develop actionable classifiers for food quality degradation decisions.

Several constraints limit the present study's generality. The calibration campaign did not include product-specific, quantified exposures for relevant spoilage VOCs but relied on gases and concentration ranges found in the literature; the study was limited to one experimental campaign and a modest number of devices, with some data fragmentation; broader campaigns are needed to establish robustness across batches and conditions. These factors restrict the chemical interpretability of individual gas concentration estimates and call for caution when extrapolating the classifier to other products, sensor batches or environments.

To increase robustness, interpretability and applicability, future works should take into account designed calibration exposures and GC-MS profiling that include spoilage-relevant VOCs for a range of cooked-meal recipes to produce chemically anchored regressors. Dedicated controls and stability studies are required: empty container and sensor-only monitoring, as well as long-term continuous recordings, will characterize baseline drift and sensor degradation in realistic storage conditions. Researchers should consider evaluating recent 2D gas-sensing materials for their potential to improve sensitivity and selectivity as well as advanced modeling, domain adaptation (e.g., nonlinear regressors, sensor-specific bias correction, and transfer learning) to further reduce inter-sensor variability while limiting the need for frequent per-device recalibration and include denser microbiological sampling around the transition and a formal sensitivity analysis using alternative cutoff times.

Author Contributions: Conceptualization, C.B., L.M. and R.G.; methodology, C.B., L.M. and R.G.; software, L.M.; validation, L.M., R.G., D.G.-R. and R.P.; formal analysis, L.M. and R.G.; investigation, L.M. and R.G.; resources, D.G.-R., C.B., A.S. and R.P.; data curation, L.M. and R.G.; writing—original draft preparation, L.M. and R.G.; writing—review and editing, D.G.-R., A.S., C.B. and R.P.; visualization, L.M. and R.G.; supervision, D.G.-R., A.S., C.B. and R.P.; project administration, D.G.-R., A.S., C.B. and R.P.; funding acquisition, D.G.-R., A.S., C.B. and R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the EU under the HORIZON-MSCA-DN-2021 program under Grant Agreement number 101072846.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used for this article are not published. However, they can be provided upon request.

Conflicts of Interest: Authors Luigi Masi, Revathy Gurusamy, Daniel Garcia-Romeo are employed by BSH Electrodomésticos. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GC-MS	Gas chromatography–Mass spectrometry
LDA	Linear discriminant analysis
LOSO	Leave one sensor out
MOS	Metal oxide semiconductor
PCA	Principal component analysis
PLSR	Partial least squares regression
RMSE	Root mean square error
TCO	Temperature-cycled operation
TVC	Total viable counts
VOCs	Volatile organic compounds

References

- Corrado, S.; Sala, S. Food waste accounting along global and European food supply chains: State of the art and outlook. *Waste Manag.* **2018**, *79*, 120–131. [CrossRef] [PubMed]
- Valant, J. ‘Best Before’ Date Labels: Protecting Consumers and Limiting Food Waste; PE 548.990; European Parliamentary Research Service: Brussels, Belgium, 2015. Available online: [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2015\)548990](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2015)548990) (accessed on 1 February 2015).
- Jaaniso, R.; Tan, O.K. (Eds.) *Semiconductor Gas Sensors*; Woodhead Publishing: Cambridge, UK, 2013.
- Edita, R.; Darius, G.; Vinauskienė, R.; Eisinaitytė, V.; Balčiūnas, G.; Dobilienė, J.; Tamkutė, L. Rapid evaluation of fresh chicken meat quality by electronic nose. *Czech J. Food Sci.* **2018**, *36*, 420–426. [CrossRef]
- Park, S.J.; Lee, S.M.; Oh, M.H.; Huh, Y.S.; Jang, H.W. Food quality assessment using chemoresistive gas sensors: Achievements and future perspectives. *Sustain. Food Technol.* **2024**, *2*, 266–280. [CrossRef]
- Jayan, H.; Zhou, R.; Sun, C.; Wang, C.; Yin, L.; Zou, X.; Guo, Z. Intelligent gas sensors for food safety and quality monitoring: Advances, applications, and future directions. *Foods* **2025**, *14*, 2706. [CrossRef] [PubMed]
- Sabu, A.; Awasthi, K.; Pandey, K.; Pandey, H. Advances in Food Spoilage Detection Through Gas Sensing: From Material Insights to Prospects for Commercialisation. *J. Agric. Food Res.* **2026**, *26*, 102750. [CrossRef]
- Schütze, A.; Sauerwald, T. Dynamic Operation of Semiconductor Sensors. In *Semiconductor Gas Sensors*; Jaaniso, R., Tan, O.K., Eds.; Woodhead Publishing: Duxford, UK, 2020; pp. 385–412.
- Lee, A.P.; Reedy, B.J. Temperature modulation in semiconductor gas sensing. *Sens. Actuators B Chem.* **1999**, *60*, 35–42. [CrossRef]
- Schütze, A.; Baur, T.; Leidinger, M.; Reimringer, W.; Jung, R.; Conrad, T.; Sauerwald, T. Highly Sensitive and Selective VOC Sensor Systems Based on Semiconductor Gas Sensors: How To? *Environments* **2017**, *4*, 20. [CrossRef]
- Berna, A. Metal oxide sensors for electronic noses and their application to food analysis. *Sensors* **2010**, *10*, 3882–3910. [CrossRef] [PubMed]
- Guohua, H.; Lvye, W.; Yanhong, M.; Lingxia, Z. Study of grass carp (*Ctenopharyngodon idellus*) quality predictive model based on electronic nose. *Sens. Actuators B Chem.* **2012**, *166*, 301–308. [CrossRef]
- Xiong, Y.; Li, Y.; Wang, C.; Shi, H.; Wang, S.; Yong, C.; Gong, Y.; Zhang, W.; Zou, X. Non-destructive detection of chicken freshness based on electronic nose technology and transfer learning. *Agriculture* **2023**, *13*, 496. [CrossRef]

14. Liu, Y.; Peng, N.; Kang, J.; Onodera, T.; Yatabe, R. Identification of beef odors under different storage day and processing temperature conditions using an odor sensing system. *Sensors* **2024**, *24*, 5590. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Joppich, J.; Schütze, A.; Bur, C. Using Human Assessment and GC-MS to Identify Potential Use Cases for Evaluating Food Condition with Gas Sensor Systems. *Chemosensors* **2026**, *14*, 73. [\[CrossRef\]](#)
16. Masi, L.; Arendes, D.; Amann, J.; Schütze, A.; Bur, C. Monitoring of VOC Emissions in Berries During the Spoiling Process. In *Proceedings of the EUROSENSORS XXXVI, Debrecen, Hungary, 1–4 September 2024*; AMA Service GmbH: Wunstorf, Germany, 2024; pp. 131–132.
17. Rahman, K.S.; Islam, M.B.; Joya, T.A.; Rahman, M.H.; Rahman, A.; Rahman, M.A.; Mim, M.J.; Rakib, M.R.I.; Rahman, A. Advancements in quality assessment of fruits and vegetables: A comprehensive review of E-nose technology. *J. Food Meas. Charact.* **2025**, *19*, 9272–9291. [\[CrossRef\]](#)
18. Essiet, I.O.; Dan-Isa, A. Practical discrimination of good and bad cooked food using metal oxide semiconductor odour sensor. *Acta Period. Technol.* **2013**, *44*, 39–47. [\[CrossRef\]](#)
19. Afidah, U.; Wardhani, R.; Dewi, S.; Purwitasari, L. Mechanisms and Applications of Essential Oils as Natural Preservatives in Meat Products: A Review. *J. Appl. Food Technol.* **2025**, *12*, 162–172. [\[CrossRef\]](#)
20. Meng, X.; Wu, D.; Zhang, Z.; Wang, H.; Wu, P.; Xu, Z.; Gao, Z.; Mintah, B.K.; Dabbour, M. An overview of factors affecting the quality of beef meatballs: Processing and preservation. *Food Sci. Nutr.* **2022**, *10*, 1961–1974. [\[CrossRef\]](#) [\[PubMed\]](#)
21. ISO 4833-1:2013(E); Microbiology of the Food Chain: Horizontal Method for the Enumeration of Microorganisms-Part 1: Colony Count at 30 °C by the Pour Plate Technique. International Organization for Standardization: Geneva, Switzerland, 2013.
22. ISO 6887-1:2017(E); Microbiology of the Food Chain—Preparation of Test Samples, Initial Suspension and Decimal Dilutions for Microbiological Examination-Part 1: General Rules for the Preparation of the Initial Suspension and Decimal Dilution. International Organization for Standardization: Geneva, Switzerland, 2017.
23. Food Standards New Zealand. *Compendium of Microbiological Criteria for Food*; Food Standards New Zealand: Wellington, New Zealand, 2022.
24. UK Health Security Agency. Guidelines for assessing the microbiological safety of Ready-to-Eat foods placed on the market. In *Interpretation of Test Results Generated by UKHSA Food Water and Environmental Microbiology Services Laboratories*; UK Health Security Agency: London, UK, 2024.
25. Centre for Food Safety, Food and Environmental Hygiene Department. *Microbiological Guidelines for Food (for Ready-to-Eat Food in General and Specific Food Items)*; Food and Environmental Hygiene Department: Hong Kong, 2014.
26. NSW/FA/CP028/0906; Microbiological Quality Guide for Ready-to-Eat Foods: A Guide to Interpreting Microbiological Results. NSW Food Authority: Newington, Australia, 2009.
27. Datasheet SGP40—Indoor Air Quality Sensor for VOC Measurements, version 1.2; Sensirion AG: Staefa, Switzerland, 2022. Available online: https://sensirion.com/media/documents/296373BB/6203C5DF/Sensirion_Gas_Sensors_Datasheet_SGP40.pdf (accessed on 14 April 2026).
28. Robin, Y.; Amann, J.; Schneider, T.; Schütze, A.; Bur, C. Comparison of transfer learning and established calibration transfer methods for metal oxide semiconductor gas sensors. *Atmosphere* **2023**, *14*, 1123. [\[CrossRef\]](#)
29. Arendes, D.; Amann, J.; Tessier, C.; Brieger, O.; Schütze, A.; Bur, C. Qualification and optimisation of a gas mixing apparatus for complex trace gas mixtures. *TM-Tech. Mess.* **2023**, *90*, 822–834. [\[CrossRef\]](#)
30. Baur, T.; Amann, J.; Schultealbert, C.; Schütze, A. Field study of metal oxide semiconductor gas sensors in temperature cycled operation for selective VOC monitoring in indoor air. *Atmosphere* **2021**, *12*, 647. [\[CrossRef\]](#)
31. Loh, W.L. On Latin hypercube sampling. *Ann. Stat.* **1996**, *24*, 2058–2080. [\[CrossRef\]](#)
32. Allgaier, J.; Rüdiger, P. Cross-validation visualized: A narrative guide to advanced methods. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1378–1388. [\[CrossRef\]](#)
33. Ambient Air Quality Directive, Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. *Off. J. Eur. Union* **2008**, *152*, 1–44.
34. Fonollosa, J.; Fernandez, L.; Gutiérrez-Gálvez, A.; Huerta, R.; Marco, S. Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization. *Sens. Actuators B Chem.* **2016**, *236*, 1044–1053. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.