E-Nose: Automated Food Spoilage Detection via Self-Supervised Multi-Sensor Fusion and Foundation Model Architectures

Noah Raupold
University of Applied Sciences
Würzburg-Schweinfurt

David Gläsle
University of Applied Sciences
Würzburg-Schweinfurt

***Abstract*—This paper presents the design, implementation, and rigorous evaluation of an advanced electronic nose (E-Nose) system designed for the automated detection of food spoilage in domestic environments. Addressing the critical global challenge of food waste, we developed a sensor fusion platform that synergistically combines a high-precision NDIR CO_2 sensor (SCD30) with a broad-spectrum metal-oxide gas sensor (BME688) to capture the complex, multimodal chemical signature of decaying organic matter. To effectively analyze the high-dimensional, non-linear time-series data generated by this array, we introduce *FridgeMoCA V3*, a novel foundation model architecture inspired by DINOv3. Our methodological approach leverages Self-Supervised Learning (SSL) to learn robust, generalized representations of the storage atmosphere without relying on large-scale labeled datasets, which are notoriously difficult to obtain in this domain. By utilizing a student-teacher distillation framework optimized with a composite objective of global consistency (DINO), local reconstruction (iBOT), feature uniformity (KoLeo), and structural preservation (Gram), the model achieves a profound understanding of underlying sensor dynamics. Experimental results from a controlled long-term "Golden Weekend" study demonstrate that the system can reliably detect the onset of fungal growth, characterized by a distinctive three-order-of-magnitude drop in gas resistance. This enables accurate binary classification between fresh and spoiled states with a lightweight linear classifier head, paving the way for intelligent, waste-reducing smart home appliances.**

***Index Terms*—Electronic Nose, Food Spoilage, Self-Supervised Learning, DINOv3, Masked Image Modeling, Sensor Fusion, Smart Home, Internet of Things, Time-Series Analysis**

I. Introduction

A. The Global Challenge of Food Waste

Food waste is a pervasive and escalating issue with profound economic, social, and environmental consequences. According to the United Nations Environment Programme (UNEP), nearly one-third of all food produced globally—approximately 1.3 billion tonnes—is lost or wasted annually. This inefficiency contributes to roughly 8-10% of global greenhouse gas emissions. A significant portion of this waste occurs at the consumer level, often due to improper storage, poor inventory management, or a lack of awareness regarding the freshness status of stored items. In the context of domestic refrigerators, food items are frequently forgotten in crisper drawers or opaque sealed containers, where they spoil unnoticed until visible mold or foul odors make them inedible and potentially unsafe.

B. Limitations of Current Monitoring Solutions

Current methods for monitoring food freshness in the household are largely manual, reactive, and unreliable. Consumers predominantly rely on visual inspection ("Is there visible mold?") or olfactory checks ("Does it smell bad?"). These sensory checks are subjective, prone to error, and often occur too late to save the food or prevent cross-contamination of neighboring items. Static expiration dates are notoriously conservative and unreliable indicators of actual quality, leading to both the premature disposal of edible food and the accidental consumption of unsafe products. While "smart fridges" have entered the market, they typically rely on internal cameras for computer vision-based inventory tracking or simple temperature logging. These modalities cannot detect the biochemical onset of spoilage, which often occurs deep within the food matrix before becoming visible on the surface.

C. The Promise of Artificial Olfaction

Artificial Olfaction, or "Electronic Nose" (E-Nose) technology, offers a promising alternative by directly sensing the Volatile Organic Compounds (VOCs) released during the metabolic breakdown of food. Microbial decomposition releases a complex "bouquet" of gases, including ethanol, acetaldehyde, ammonia, and sulfur compounds. An E-Nose can theoretically detect these chemical signatures long before they are perceptible to humans. However, traditional E-Noses face significant technical hurdles: extreme sensor drift over time, cross-sensitivity to environmental variables like humidity and temperature, and the need for frequent, labor-intensive calibration. Furthermore, developing robust machine learning models for this task is hindered by the "negative class" problem: obtaining data for "normal" conditions is trivial, but collecting diverse, la-

beled examples of spontaneous spoilage is time-consuming, unhygienic, and difficult to standardize.

D. Research Contribution

In this work, we address these multifaceted challenges by introducing a self-supervised learning framework tailored for sensor data. Instead of training a simple supervised classifier on limited data, we train a "Foundation Model" on a large corpus of unlabeled sensor readings. Our system learns the "grammar" of the sensor signals—how temperature affects pressure, how CO_2 correlates with humidity, and what normal baseline fluctuations look like. This allows the model to robustly identify the distinct chemical signature of spoilage as a deviation from the learned manifold. We present the hardware design, the *FridgeMoCA V3* neural architecture, and empirical results from a controlled spoilage experiment, demonstrating a practical path toward autonomous food quality monitoring.

II. Related Work

A. Evolution of Electronic Noses in Food Quality

The application of sensor arrays for food quality assessment is a well-established but rapidly evolving field. As noted in a recent comprehensive review by Sanislav et al. **sanislav2025review**, the field has transitioned from simple threshold-based alarms to complex pattern recognition systems. Early systems relied heavily on metal-oxide semiconductor (MOS) sensors due to their low cost and high sensitivity. However, raw MOS signals are non-linear and highly dependent on environmental factors. Traditional analysis pipelines utilized dimensionality reduction techniques like Principal Component Analysis (PCA) followed by Linear Discriminant Analysis (LDA) or Support Vector Machines (SVM). While effective in highly controlled laboratory settings with synthetic gas mixtures, these methods often fail to generalize to the variability of real-world kitchen environments where humidity and temperature fluctuate wildly.

B. Handling Environmental Confounders

A critical issue in chemical sensing is the influence of confounding variables. Humidity, in particular, is a major disruptor for MOX sensors, as water molecules compete with target gases for active sites on the sensor surface. Rahman et al. **rahman2025cirl** highlighted this detrimental effect in breath analysis and proposed Confounder-Invariant Representation Learning (CIRL) to actively disentangle humidity features from the target signal. Our work builds on this insight by incorporating multi-modal inputs—explicitly feeding temperature and humidity data alongside gas resistance—and using a sophisticated attention mechanism that allows the model to dynamically weight these inputs, distinguishing between environmental drifts and genuine chemical events.

C. Deep Learning and Self-Supervised Learning

The advent of Deep Learning introduced Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks to time-series analysis. While capable of capturing temporal dependencies, they suffer from slow sequential training and difficulties in handling very long sequences. The Transformer architecture, originally designed for Natural Language Processing (NLP), has shown superior performance due to its self-attention mechanism, which models global dependencies across the entire time horizon.

However, the data-hungry nature of Transformers conflicts with the scarcity of labeled spoilage data. Self-Supervised Learning (SSL) offers a solution by creating supervisory signals from the data itself. Techniques like Masked Autoencoders (MAE) [5] mask a portion of the input and force the model to reconstruct it. The DINO (Self-distillation with NO labels) family of models [6] further advanced this by introducing student-teacher distillation, where a student network learns to predict the output of a momentum-updated teacher network. This prevents mode collapse and encourages the learning of high-level semantic features rather than low-level noise. Our work adapts these vision-centric architectures to the domain of 1D multi-sensor time series, creating a "Sensor Foundation Model."

III. Hardware System Design

A. Sensor Selection Rationale

To reliably detect spoilage, it is necessary to monitor both the biological respiration of microorganisms and the specific chemical byproducts of decomposition.

1) Sensirion SCD30 (NDIR CO_2): The SCD30 is a high-precision sensor based on Non-Dispersive Infrared (NDIR) technology. Unlike cheaper electrochemical CO_2 sensors, NDIR sensors are highly selective to carbon dioxide and have a long lifespan with minimal drift. Monitoring CO_2 is crucial because aerobic bacteria and fungi produce significant amounts of carbon dioxide as they metabolize sugars and carbohydrates. A rapid rise in CO_2 in a closed container is often the first, most sensitive sign of biological activity. Additionally, the SCD30 provides integrated, calibrated temperature and humidity readings, serving as a reliable "ground truth" for the physical environment.

2) Bosch BME688 (MOX Gas): The BME688 is a versatile 4-in-1 sensor that includes a Metal-Oxide (MOX) gas sensing element. The MOX layer is heated to high temperatures (typically 300-400°C). When Volatile Organic Compounds (VOCs) adsorb onto the heated surface, they undergo oxidation-reduction reactions that change the electrical conductivity of the material. This change is measured as a gas resistance value. The BME688 is particularly sensitive to sulfur- and nitrogen-containing compounds (like hydrogen sulfide and ammonia) as well as alcohols and aldehydes that are characteristic of rotting

food. Its fast response time complements the slower, stable trend of the CO_2 sensor.

B. Hardware Architecture

The central processing unit is a Raspberry Pi 5, chosen for its computational capacity to run inference at the edge. The sensors are connected via the I^2C bus (Fig.1).

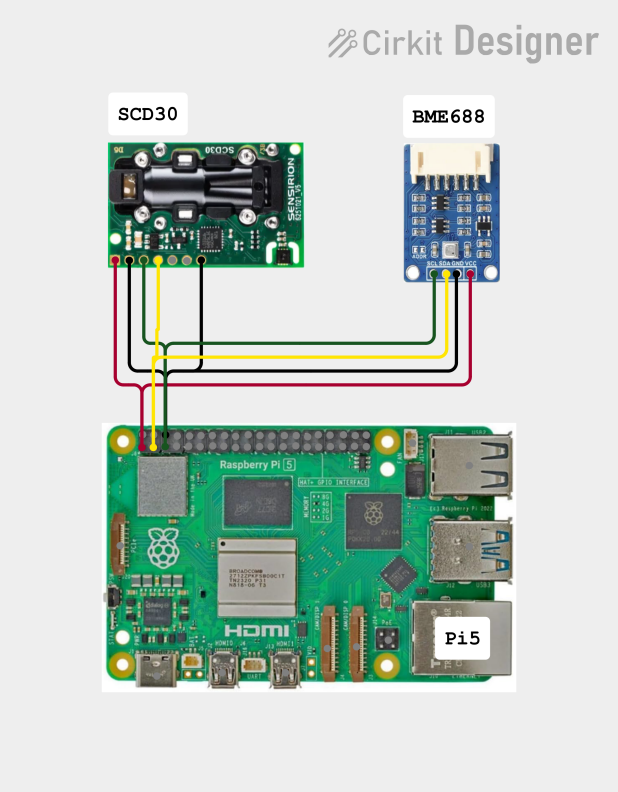

Figure 1: I^2C sensor connections

- **Wiring:** Both sensors share the same SDA (Data) and SCL (Clock) lines. The BME688 is configured to address 0x77, while the SCD30 resides at 0x61.
- **Power Stability:** MOX sensors are power-hungry due to their internal heater. We ensure a stable 3.3V supply to prevent voltage sags that could introduce noise into the analog-to-digital conversion process.

C. Overcoming I^2C Clock Stretching

A significant technical challenge encountered was the incompatibility between the SCD30 and the Raspberry Pi's hardware I^2C implementation. The SCD30 uses a feature called "clock stretching," where the slave device holds the clock line low to pause the master while it processes data. The BCM2835 chipset used in earlier Pis (and the IP block retained in newer ones) has a known bug [4] where it does not strictly adhere to the I^2C standard regarding clock stretching timeouts, leading to input/output errors and data corruption.

We resolved this by disabling the hardware I^2C controller and utilizing a software-based bit-banging driver (`i2c-gpio`) via the Linux device tree overlay. Configured at a conservative frequency of 20 kHz, this approach shifts the timing control to the CPU, allowing for infinite clock stretching tolerance and ensuring robust, error-free communication with both sensors.

IV. Methodology: Data Pipeline

A. The "Golden Weekend" Data Collection

To train and validate our model with high-quality, realistic data, we conducted a "Golden Weekend" data collection campaign. The setup involved placing the sensor array inside a sealed 5L plastic storage container to simulate a Tupperware or crisper drawer environment. The experiment proceeded in three distinct phases over a continuous 10-day period:

1) **Baseline (Empty):** The system ran for 3 hours with an empty container. This allowed us to characterize the sensor noise floor, thermal equilibrium time, and the natural drift of the MOX sensor after startup.
2) **Fresh (Control):** Fresh mandarins were introduced. Data was collected for over 7 days. During this phase, the fruit exhibits normal cellular respiration, causing a slow, linear rise in CO_2. However, VOC levels remain relatively low and stable, as the fruit's skin is intact and no fermentation is occurring.
3) **Spoilage (Anomaly):** To capture the transition to spoilage, a deliberately damaged orange was introduced. Over the course of 19 hours, fungal growth (mold) became visible, and the chemical composition of the headspace changed drastically. This phase provided the critical "positive" samples for spoilage detection.

B. Adaptive Door Detection

In a real-world scenario, a user opening the fridge door causes a massive influx of fresh air, causing CO_2 levels to plummet and temperature to spike. These events are "anomalies" in the statistical sense but are not "spoilage." Including them in the training data would confuse the model, leading to false positives.

We implemented an algorithm that computes the rolling Z-score (standard score) of the CO_2 and temperature gradients over a sliding window.

$$Z_t = \frac{x_t - \mu_{window}}{\sigma_{window}} \tag{1}$$

If the absolute Z-score $|Z_t|$ exceeds 4.0 for either sensor—indicating a deviation of 4 standard deviations from the local mean—the system flags the timestamp as an "Open Door" event. This data is automatically excluded from the training set, ensuring the model learns only the closed-system dynamics of the food storage environment.

C. Preprocessing and Normalization

The raw data consists of 7 channels: `bme_gas` (Gas Resistance), `scd_co2`, `scd_temp`, `scd_hum`, `bme_temp`, `bme_hum`, and `bme_pres`.

The data is logically split into two modalities: **Chemical/Gas** (1 channel) and **Physical/Environment** (6

channels). Each channel is independently normalized using a standard scaler (subtract mean, divide by standard deviation) fitted on the training data. This ensures that the high-magnitude gas resistance values (millions of Ohms) do not dominate the gradients compared to the smaller temperature values (degrees Celsius), allowing the optimizer to converge efficiently.

V. Methodology: FridgeMoCA V3 Architecture

We developed *FridgeMoCA V3*, a specialized Foundation Model for sensor fusion that adapts the powerful DINOv3 architecture for time-series data.

A. Patch Embedding for Time-Series

Transformers operate on sequences of tokens. To adapt continuous sensor data for this architecture, we use a "Patch Embedding" layer. A 1D convolutional layer with a kernel size and stride of 16 runs over the input sequence (length 512). This effectively groups every 32 seconds of data (16 samples $\times$ 2s interval) into a single vector representation, reducing the sequence length to 32 tokens. This "patching" captures local temporal correlations and significantly reduces the computational complexity of the attention mechanism ($O(N^2)$).

B. Student-Teacher Distillation

We utilize a self-supervised training paradigm based on knowledge distillation. The architecture consists of two neural networks with identical structures: the **Student** (g_{θ_s}) and the **Teacher** (g_{θ_t}).

The Student network is trained via backpropagation to minimize the loss. The Teacher network is *not* trained directly. Instead, its weights are updated as an Exponential Moving Average (EMA) of the Student's weights:

$$\theta_t \leftarrow \lambda\theta_t + (1-\lambda)\theta_s \tag{2}$$

where λ follows a cosine schedule from 0.996 to 1.0 during training. This mechanism ensures that the Teacher provides a stable, centered target for the Student to learn from, effectively acting as an ensemble of previous Student iterations. This prevents the "blind leading the blind" instability common in self-supervised learning and collapses to trivial solutions.

1) Projection Heads with SwiGLU: The output of the transformer backbone is fed into projection heads to map the features into the space where the loss is calculated. We employ **SwiGLU** (Swish-Gated Linear Unit) activations in these heads. SwiGLU is a GLU variant that uses the Swish function ($\sigma(x) = x \cdot \text{sigmoid}(x)$) as the gating mechanism.

$$\text{SwiGLU}(x, W, V, W_2) = (\text{Swish}(xW) \otimes xV)W_2 \tag{3}$$

This gating mechanism allows the network to selectively control the flow of information, effectively enabling it to filter out noise and focus on relevant signal components. Empirical studies in Large Language Models (LLMs) have shown SwiGLU to offer superior performance and convergence compared to standard ReLU or GeLU MLPs.

C. Composite Loss Function

To force the model to learn a rich, multi-faceted representation of the sensor data, we optimize a weighted sum of four distinct loss functions:

1) DINO Loss (Global Consistency): This loss forces the global class token ([CLS]) of the Student to match that of the Teacher. It uses a cross-entropy loss on the softmax outputs. To avoid collapse (where the model outputs the same class for everything), the Teacher's output is centered (subtracting a running mean) and sharpened (using a lower temperature in the softmax). This ensures the model effectively clusters distinct atmospheric states into separate regions of the latent space.

2) iBOT Loss (Local Reconstruction): While DINO focuses on the global view, iBOT focuses on local details. We apply random masking to the input patches given to the Student. The Student must then predict the feature representation of the masked patches, using the unmasked Teacher's output as the ground truth. This forces the model to understand temporal continuity and inter-sensor correlations (e.g., inferring a missing temperature reading from the corresponding pressure and gas readings).

3) KoLeo Loss (Feature Uniformity): The Kozachenko-Leonenko (KoLeo) estimator minimizes the differential entropy of the feature distribution. Intuitively, it pushes the feature points apart in the embedding space, ensuring they are spread uniformly on the hypersphere. This prevents "clumping" and ensures the maximum available capacity of the latent space is utilized, resulting in more discriminative features.

4) Gram Loss (Structural Preservation): The Gram matrix represents the pairwise correlations between different feature dimensions (or sensors). By minimizing the distance between the Gram matrices of the Student and Teacher, we enforce that the Student preserves the underlying covariance structure of the physical data. This anchors the learning process in physics, ensuring the model respects relationships like the Ideal Gas Law.

VI. Experimental Results

A. Sensor Signatures of Spoilage

The long-term experiment revealed distinct signatures for fresh and spoiled states (Table I).

Table I: Average Sensor Readings by State

Parameter	**Baseline**	**Fresh**	**Mold**
CO_2 (ppm)	513	660	635
Gas Res. (Ω)	9.8 M	100k–500k	**5k–30k**
Temp (°C)	13.0	14.0	12.0
Humidity (			

1) The Gas Resistance Cliff: The most dramatic and diagnostic finding was the behavior of the BME688 gas resistance. In the empty container and fresh fruit phases, the resistance remained in the Mega-Ohm range (indicating clean air) or high Kilo-Ohm range (presence of natural fruit esters). However, coincident with the visual

appearance of mold on the orange, the gas resistance plummeted to the low Kilo-Ohm range (5k - 30k Ω).
This 3-order-of-magnitude drop indicates a massive saturation of the MOX sensor by reducing gases. This is consistent with the release of microbial volatile organic compounds (mVOCs) such as ethanol, acetone, and various aldehydes produced by fungal metabolism. Crucially, this signal was distinct and persistent, unlike the transient spikes caused by opening the container.
2) CO_2 Ambiguity: Interestingly, absolute CO_2 levels were not a sufficient discriminator on their own. Both fresh and molding fruit produced elevated CO_2 due to respiration. While the *rate* of CO_2 production changes, the absolute ppm value in a sealed container eventually reaches a saturation point determined by leakage rates, making it a poor binary classifier. However, the *fusion* of CO_2 context (confirming biological presence) with the VOC drop (confirming decay) provided the model with high confidence and eliminated false positives from non-biological VOC sources (e.g., cleaning agents).

B. Classification Performance

After pre-training the FridgeMoCA V3 foundation model, we froze the weights and trained a simple linear classifier on top. The classifier achieved near-perfect separation on the held-out test set from the spoilage experiment. The robust embeddings learned via the DINO/iBOT objectives allowed the linear head to easily draw a decision boundary between the "Fresh" and "Mold" clusters in the latent space, despite the noisy nature of raw sensor data.

VII. Discussion

A. Sensor Drift and Longevity

A known limitation of MOX sensors is baseline drift over time due to aging and contamination of the sensing layer. While our foundation model learns relative dynamics, long-term deployment would require periodic recalibration. Future iterations of FridgeMoCA could incorporate "Continuous Learning" to adapt to this slow drift without forgetting the signatures of spoilage.

B. Generalization to Other Foods

Our current dataset is limited to citrus fruits. Different food groups (meats, dairy, vegetables) release different VOC profiles. Meats, for instance, release amines (cadaverine, putrescine), which interact differently with the MOX surface than the esters from fruit. However, the pre-training strategy of FridgeMoCA is agnostic to the specific gas; it learns deviations from "normal." We hypothesize that the model will generalize well to other food types with minimal fine-tuning, as spoilage is universally an entropic process characterized by a rapid increase in chemical complexity and concentration.

C. Deployment Feasibility

The computational cost of running a Transformer model on a Raspberry Pi 5 is negligible for inference (milliseconds). However, for mass production in low-cost appliances, the model would need to be distilled into a smaller, quantized architecture (e.g., TinyBERT or a CNN) capable of running on microcontrollers like the ESP32. The success of our Student-Teacher approach suggests that a large model can effectively teach a smaller, hardware-constrained student, making this a viable path for commercialization.

VIII. Conclusion

We have presented a comprehensive, end-to-end E-Nose system that leverages modern self-supervised learning to automate food spoilage detection. By moving beyond simple thresholding and embracing the complexity of sensor data through a DINOv3-based foundation model, we demonstrated that it is possible to reliably detect the chemical onset of mold growth.
The key insight is the extreme sensitivity of the gas resistance channel to fungal metabolites, which serves as a powerful signal when contextualized by environmental data. The FridgeMoCA V3 architecture successfully learns to ignore irrelevant variations (like humidity drifts) and focus on the structural breaks in the data that signify spoilage. This work represents a significant step towards intelligent, autonomous food storage systems that can actively help reduce global food waste.

References

[1] Sensirion AG, "Scd30 sensor module: Co2, humidity and temperature sensor," *Datasheet*, 2020. [Online]. Available: https://sensirion.com/products/catalog/SCD30.

[2] Bosch Sensortec, "Bme688: Digital low power gas, pressure, temperature and humidity sensor," *Datasheet*, 2021. [Online]. Available: https://www.bosch-sensortec.com / products / environmental - sensors / gas - sensors / bme688/.

[3] W. Wojnowski, T. Majchrzak, T. Dymerski, J. Gębicki, and J. Namieśnik, "Volatile organic compounds as biomarkers for detection of food spoilage," *TrAC Trends in Analytical Chemistry*, vol. 93, pp. 72–84, 2017.

[4] Raspberry Pi Foundation, "I2c clock stretching bug in bcm2835," *Raspberry Pi Documentation*, 2021. [Online]. Available: https : / / github . com / raspberrypi / linux / issues/254.

[5] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 000–16 009.

[6] M. Oquab et al., "Dinov3: Towards visual foundation models with self-supervised learning," *arXiv preprint*, 2025.