Proceedings Paper

Data-Centric Performance Improvement Strategies for Few-Shot Classification of Chemical Sensor Data †

Bhargavi Mahesh *,‡ , **Teresa Scholz** ‡, **Jana Streit, Thorsten Graunke and Sebastian Hettenkofer**

Fraunhofer Institute for Integrated Circuits IIS, 91058 Erlangen, Germany; teresa.scholz@iis.fraunhofer.de (T.S.); jana-streit@web.de (J.S.); thorsten.graunke@iis.fraunhofer.de (T.G.); sebastian.hettenkofer@iis.fraunhofer.de (S.H.)

* Correspondence: bhargavi.mahesh@iis.fraunhofer.de
† Presented at the 8th International Electronic Conference on Sensors and Applications, 1–15 November 2021; Available online: https://ecsa-8.sciforum.net.
‡ These authors contributed equally to this work.

Abstract: Metal oxide (MOX) sensors offer a low-cost solution to detect volatile organic compound (VOC) mixtures. However, their operation involves time-consuming heating cycles, leading to a slower data collection and data classification process. This work introduces a few-shot learning approach that promotes rapid classification. In this approach, a model trained on several base classes is fine-tuned to recognize a novel class using a small number (n = 5, 25, 50 and 75) of randomly selected novel class measurements/shots. The used dataset comprises MOX sensor measurements of four different juices (apple, orange, currant and multivitamin) and air, collected over 10-minute phases using a pulse heater signal. While high average accuracy of 82.46 is obtained for five-class classification using 75 shots, the model's performance depends on the juice type. One-shot validation showed that not all measurements within a phase are representative, necessitating careful shot selection to achieve high classification accuracy. Error analysis revealed contamination of some measurements by the previously measured juice, a characteristic of MOX sensor data that is often overlooked and equivalent to mislabeling. Three strategies are adopted to overcome this: (E1) and (E2) fine-tuning after dropping initial/final measurements and the first half of each phase, respectively, (E3) pretraining with data from the second half of each phase. Results show that each of the strategies performs best for a specific number of shots. E3 results in the highest performance for five-shot learning (accuracy 63.69), whereas E2 yields the best results for 25-/50-shot learning (accuracies 79/87.1) and E1 predicts best for 75-shot learning (accuracy 88.6). Error analysis also showed that, for all strategies, more than 50% of air misclassifications resulted from contamination, but E1 was affected the least. This work demonstrates how strongly data quality can affect prediction performance, especially for few-shot classification methods, and that a data-centric approach can improve the results.

Keywords: metal oxide sensors; few-shot classification; data quality analysis

1. Introduction

Gas detection and classification, as well as the analysis of the composition of gas mixtures, can be performed with analytical tools such as gas chromatography, mass spectrometry or Fourier transform infrared spectroscopy. Unfortunately, these tools are expensive and difficult to operate. Metal oxide (MOX) sensors or arrays of MOX sensors are a promising alternative as they are small and financially competitive [1]. However, these sensors lack the selectivity to target volatile organic compounds (VOCs) and are prone to cross-contamination. Selectivity and stability can be improved with metal oxides such as SnO_2, WO_3, TiO_2, CuO, In_2O_3, ZnO, Fe_2O_3, as well as the addition of noble metals such as Pd or Pt. Moreover, the definition of a heater temperature modulation, which influences the gas-specific reaction with the sensor surface, allows for a more stable classification of results [2]. However, using temperature modulation, MOX sensors consume several

seconds for a single data sample, resulting in a prolonged data collection process. This becomes a hindrance during real-time inferencing as well. For instance, a classification algorithm that learns to detect a particular class must be trained in a supervised manner on several data samples and may require minutes to hours until it learns a new class. Hence, a rapid classification strategy becomes necessary to cope with the inherent delay associated with MOX sensors. In this work, a method to rapidly classify MOX sensor data is presented and strategies to improve the classification performance by obtaining deeper insights into the characteristics of the data are explored.

2. Applications of MOX Sensors in Food Industry

Ideally, data collected using MOX sensors serve as a "fingerprint" of the volatile components emitted by the measured substance. Thus, the data, together with an appropriate algorithm, can serve to detect any deviation from the norm, which, in the food industry, has been applied to control the quality and authenticity of products. A good review of these studies is provided by [3,4]. In the context of food authenticity, MOX sensors paired with pattern recognition algorithms have been used for many applications, such as the identification of adulterated milk, cow ghee [5], olive oil, saffron and cherry tomato juice. Moreover, for various products, such as olive oil, orange juice, meat, milk or honey, the authenticity of the geographical origin could be determined. Moreover, the technique has also served to determine faults in production processes. The "electronic nose" was also able to detect food spoilage, i.e., microbial contamination in soft drinks [6], juices [7,8] and meat products, and assess the freshness of produce such as meat, eggs or fish. In addition, MOX sensors served to assess the age or ripeness of products for which this is a quality-defining parameter, such as fruit or wines. The systems applied in most of these studies consist of an array of MOX sensors combined with a simple pattern recognition algorithm based on principal component analysis, linear discriminant analysis, partial least squares regression or cluster analysis. Recent work [9] has shown that a model based on convolutional neural network results in automatic drift counteraction. Data collection in previous research was usually performed in a laboratory-controlled environment, yielding very clean data and not dealing with the MOX sensor's sensitivity towards temperature, humidity or air composition. This paper distinguishes itself by presenting a fast few-shot learning approach with a convolutional neural network (CNN) trained on the data collected in a uncontrolled regular office environment. It also demonstrates how strongly data collection can impact quality and, in turn, the prediction performance.

3. Data Collection

The data used for this paper were collected using four AS-MLV-P2 [10] sensors with a sensitive layer of SnO_2:Pd. Measurements were conducted with several sensors to build up redundancy in case of sensor failure and also to ensure a robust model that was not overfitted to one sensor's characteristics. As a reference, four more sensors of the same type were placed inside the room to measure the surrounding air composition. All sensors were operated with a temperature modulation of 1 s on 450 °C, 5 s on 200 °C, 1 s on 450 °C and 5 s on 300 °C. The high temperature was selected to generate a surface charge and the intermediate low temperatures to introduce fast temperature changes, with the goal of quantifying specific reducing gases in the juices.

For each measurement, 6 cl of four different types of juices (apple, currant, orange and multivitamin) was poured into a 6 cm high glass that was subsequently covered with a plexiglass cover into which the MOX sticks had been drilled. Apart from the juice headspace, pure air was measured by exposing the sensor to the ambient air. Each sample was continuously measured for 10 minutes (phase), during which the predefined temperature cycle was repeated. The data collection protocol was designed in such a way that each sample was measured subsequently to each other sample, with 4 types of juice and air. This led to a collection protocol of 20 phases, which was repeated 4 times over a timeframe of 8 months. The measurements from different days, despite being influenced

by environmental conditions as well as sensor drift, were shuffled to create the training
and test datasets.

4. Method

Few-shot classification (FSC) is a method to enable rapid classification, i.e., the classifier
learns to identify a new class when trained with a few inputs or shots. In this work, we
used FSC to enhance the capabilities of a baseline model to detect a novel class that was not
significantly different from the base classes. Using the transfer-learning-based approach,
the classifier is initially trained on the base classes (meta-training stage) and a part of the
model is fine-tuned on the novel class data (fine-tuning/meta-testing). The training dataset
in the fine-tuning stage is called the support dataset, whereas the test dataset is known as
the query set [11]. The meta-training stage involves the standard training procedure. In the
fine-tuning phase, a small part of the network is retrained as the support dataset consists of
samples in the order of 10.

In this work, four few-shot classification experiments were conducted, where each
experiment considered one of the 4 juices as novel. Thus, the data for meta-training X_b
consisted of three juice classes and air as the base class, and the data for fine-tuning X_n
contained the novel juice class in addition. Each dataset was further split into balanced
training and test datasets. The few-shot classification model comprises a convolutional
neural network and was divided into two parts. In the meta-training stage, the feature
extractor f_θ, a convolutional neural network parametrized by the network parameters θ,
and the classifier $C(\cdot|W_b)$ parametrized by the weight matrix W_b, are trained by minimizing
the binary cross-entropy classification loss on the train set of X_b. The trained model is
validated on the held-out part of X_b. The feature extractor consists of a Gaussian noise layer
and two convolutional layers, all using the ReLu activation function as well as a dense layer.
The classifier $C(\cdot|W_b)$ consists of a fully connected layer with five output nodes in both
the meta-training and fine-tuning stages. During meta-training, the excess output node
is forced to output zero. In the fine-tuning stage, the parameters θ of the feature extractor
f_θ are frozen and the classifier is fine-tuned to obtain the weights W_n. The support set of
X_n with novel juice class is used to fine-tune the classifier using binary cross-entropy loss
minimization.

In each experiment's fine-tuning stage, four different ways to fine-tune the classifier,
namely 1-shot, 5-shot, 50-shot and 75-shot, varying in the number of shots, were tested. A
special case of zero-shot was tested, where there is no fine-tuning, yet the query set was
classified by the model trained on base classes. An increase in classification performance
from that of the zero-shot regime is likely to depict the information gain from novel classes.
Since iteratively trained algorithms undergo catastrophic forgetting post-fine-tuning, the
validation dataset from the meta-training stage was used to test the extent of forgetting—
the catastrophic forgetting test (CFT). The lower the change in performance before and
after fine-tuning, the more robust is the model.

5. Baseline Few-Shot Classification Results

During the meta-training stage, the feature extractor and classifier were optimized
using teh Adam optimizer, trained for 200 epochs with a batch size of 20 with an initial
learning rate of 0.001, which was increased to 0.01 during fine-tuning. Results are presented
in Table 1. The average validation accuracy of the model over all the experiments during
meta-training was 82.83%. Upon fine-tuning this pretrained model using the five-shot
regime, the average accuracy obtained on the query set was 44.46%, whereas that using
75-shot was 82.47%. The difference in the 5-shot and 75-shot performances reveals that a
pretrained model has difficulty in learning and generalizing to new classes from a small
amount of data. The CFT results show the tendency of the model to overwrite the previous
training information upon training to a new scenario. Some information will be naturally
lost upon fine-tuning with a higher number of shots from a new class. This is reflected
in the reduction in CFT accuracy as the number of shots increased. However, since the
new training scenario also included base classes, the degradation in CFT accuracy was

not as severe as the increase in the test accuracy as shots increased. The pretrained model did not undergo catastrophic forgetting as the CFT validation accuracy was close to the meta-training validation accuracy. Plausible reasons are the shallow network architecture and the low number of fine-tuning iterations.

Table 1. Average validation, test and catastrophic forgetting test accuracies.

Validation	#Shots	Test	CFT
0.82825	0-shot	0.4618	-
	5-shot	0.4446	0.8143
	25-shot	0.6934	0.8118
	50-shot	0.7742	0.8109
	75-shot	0.8247	0.7907

5.1. Sample Screening

The performance of the k-shot learned model relies on the selected k input samples, which should be representative of their class. To verify this, sample screening was carried out: each sample in the novel juice class was used to fine-tune a pretrained model using a one-shot regime. The fine-tuned model was validated on a balanced dataset comprising the rest of the novel juice class and air samples. This one-shot validation was conducted in the same order of data collection. Figure 1 depicts the reduction in test performance in the early minutes of every 10-min measurement phase after fine-tuning the model. The subsequent 10-min phases separated by the 'phase start' line in the figure had time gaps ranging from 10 min to several days. The validation accuracy was significantly lower for samples at the start (first 5–10 temperature cycles, each lasting 12 s) of the phases, indicating contamination from the previous phase. This was likely due to the residual effect of the previously measured class on the sensors.
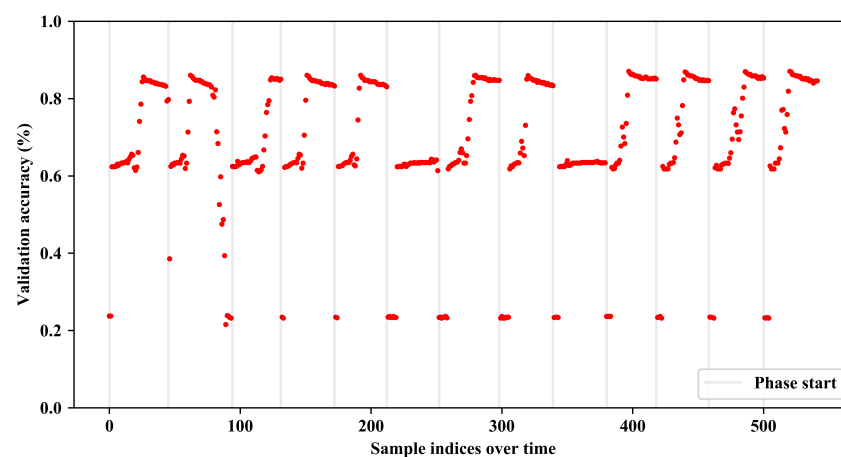
Figure 1. Test accuracies obtained when the classifier was fine-tuned on only one sample of orange juice depending on the selection of this single sample. Samples from the beginning of the phase often resulted in reduced performance.

5.2. Error Analysis

The misclassifications in each experiment were studied based on the number of shots used to fine-tune. Moreover, the percentage of influence of the juice measured in the previous phase on the misclassifications was also calculated. A misclassification qualified for an influence when the predicted juice class coincided with the juice phase prior to the current sample's phase. The metrics were split into air and the juice class in X_n. For all k-shot experiments (except the 25-shot test for the multivitamin), more than 50% of the air misclassifications were related to the previous juice class (refer to Table 2). As the shots increased, the misclassifications for juice decreased and the fine-tuned model

became robust to the previous juice phases' influence as well. This analysis indicates that the contamination effect is reflected in the modeling results.

Table 2. Misclassification (M) out of 3220 per class and previous phases' influences (I) on them.

	Class	#M	#I	% M	% I
5-shot	Air	357	243	11.09	68.07
	Juice	3220	460	100.0	14.29
25-shot	Air	474	336	14.72	70.89
	Juice	1501	233	46.61	15.52
50-shot	Air	443	267	13.76	60.27
	Juice	1011	86	31.40	8.51
75-shot	Air	340	269	10.56	79.12
	Juice	789	22	24.50	2.79

## 6. Data Analysis: Class Separability and Contamination

Section 5 indicated that the first measurements of each phase are not representative of the measured class. To investigate the data quality and separability of the five different classes (air as well as orange, apple, multivitamin and currant juice), the data were transformed using t-Distributed Stochastic Neighbor Embedding (t-SNE), a technique for dimensionality reduction that is particularly well-suited to the visualization of high-dimensional datasets [12]. Figure 2 shows the data projected into the two-dimensional t-SNE plane using a perplexity of 30. It can be seen that all juices formed (sometimes overlapping) clusters, which each could be divided into sub-clusters, indicating the different phases of measurement (data not shown). Each of the sub-clusters was of an oblong form, ending in air measurements. Air overall formed a widespread cluster containing measurements labeled as juice spread throughout it. These patterns can be explained by contamination: after measuring a juice, the air surrounding the glass as well as the sensor still contained volatile components emitted by the juice, distorting the air measurement. Thus, whenever the measurement of an air phase started, the data point was still projected into the area of the 2D plot of the corresponding juice (the 'tips' of the elongated clusters). As the juice aromas disappeared, the voltage signal changed to that of pure air and the corresponding data points were projected into the air cluster. The same phenomenon could be observed when juice was measured after air: the first samples, where the juice VOCs were still strongly diluted by air, were projected into the air cluster. Once the juice aroma concentration was high enough, the data were projected into the space corresponding to the juice. Moreover, as the concentration of the VOCs increased, the samples stretched along the elongated sub-cluster. This is illustrated in Figure 2 on the right, which shows a color-coded plot of orange juice measurements: the first sample taken is dark blue, the last one bright yellow.

These contamination patterns can also be observed directly in the voltage data. Figure 3 shows all measurements taken during a phase of orange juice following a phase of air measurements (left), vice versa (middle) and reference measurements taken outside of the measuring glass (right), with the color bar indicating the sample number.
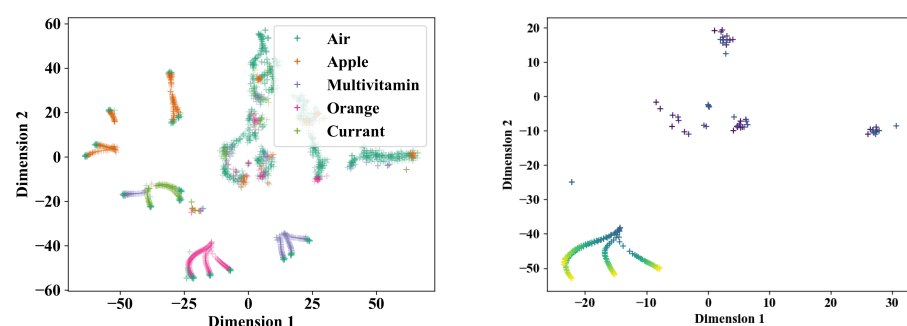
Figure 2. All (**left**) and only orange juice (**right**) measurements projected into t-SNE plane.
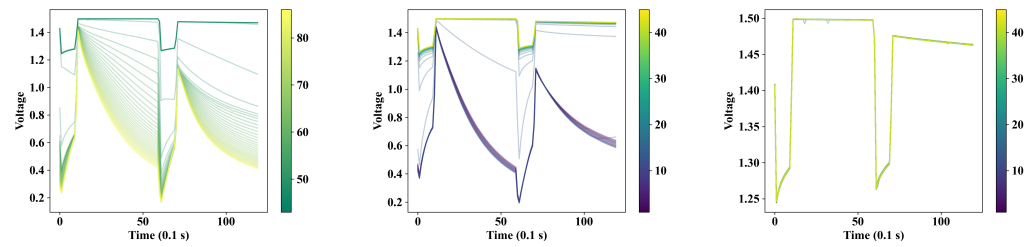
Figure 3. Measurements collected during a phase of orange juice measurements following a phase of air measurement (**left**), air measurements following a phase of orange juice measurement (**middle**) and reference measurements of the room air collected at the same time (**right**).

7. Data-Centric Improvement Strategies and Their Results

Sample screening and error analysis indicated contamination in the data and, therefore, data-centric strategies to improve results were employed. Considering the previously used few-shot classification strategy as the first (E0), three other strategies, involving careful selection of samples for fine-tuning or pretraining, were tested:

E1: Dropping initial and final measurements: The first and last 10 samples of each phase were excluded as they resulted in reduced one-shot validation accuracy (Section 5.1) and could be prone to phase transition errors, respectively. From the remaining samples, the shots for fine-tuning were randomly selected, resulting in significantly improved accuracy (Figure 1).

E2: Dropping first half phase: The samples from the first half of each phase (samples 0–19) were removed as, in the majority of the phases, the measurement cycles stabilized after the twentieth measurement (Figure 1). The shots for fine-tuning were randomly chosen from the remaining data. Contrary to what was expected, the resulting test accuracies for different numbers of shots either decreased or remained the same, except for the 50-shot test, where it increased by 2%.

E3: Dropping first half phase and retraining: The model was retrained with the base classes after removing each first half phase, assuming that the possible contamination affected the model. Shots for fine-tuning were selected from the second half of each phase. With the exception of five shots, all tests resulted in reduced accuracy. This was likely due to overfitting in the model and, hence, a loss of generalizability.

Table 3 shows that strategy E1 improved all k-shot tests' performance, whereas the rest improved for specific shots. Misclassification analysis showed that E1 yielded the fewest air misclassifications and E3 the fewest juice misclassifications. E1 also demonstrated a lesser influence of the previously measured juice on the classification. Retaining a few underperforming samples allowed the model to be robust to contamination.

Table 3. Test accuracies averaged over shots for four strategies.

	E0	E1	E2	E3
5-shot	0.4445	0.4699	0.4710	0.6369
25-shot	0.6933	0.7878	0.7907	0.7725
50-shot	0.7742	0.8528	0.8705	0.7962
75-shot	0.8246	0.8860	0.8601	0.8424

8. Conclusions

This work demonstrates the impact of data quality on prediction performance, especially for few-shot classification methods, and shows that a data-centric approach can improve results. Three strategies were adopted to overcome issues due to the non-representativeness of the samples. Results showed an overall classification improvement in strategy E1. Moreover, each of the strategies performed best for a specific number of shots. Error analysis revealed that, for all strategies, more than 50% of air misclassifications resulted from contamination, but E1 was affected the least.

References

1. Kappler, J. *Characterisation of High-Performance Sno2 Gas Sensors for Co Detection by In Situ Techniques*; Shaker Verlag: Düren, Germany, 2001.
2. Boiger, R.; Defregger, S.; Grbic, M.; Köck, A.; Mücke, M.; Wimmer-Teubenbacher, R.; Travieso, B.Z. Exploring Temperature-Modulated Operation Mode of Metal Oxide Gas Sensors for Robust Signal Processing. *Proceedings* **2019**, *2*, 1058. [CrossRef]
3. Berna, A. Metal Oxide Sensors for Electronic Noses and Their Application to Food Analysis. *Sensors* **2010**, *10*, 3882–3910. [CrossRef] [PubMed]
4. Gliszczyńska-Świgło, A.; Chmielewski, J. Electronic Nose as a Tool for Monitoring the Authenticity of Food. A Review. *Food Anal. Methods* **2017**, *10*. [CrossRef]
5. Ayari, F.; Mirzaee-Ghaleh, E.; Rabbani, H.; Heidarbeigi, K. Using an E-nose machine for detection the adulteration of margarine in cow ghee. *J. Food Process. Eng.* **2018**, *41*, e12806. [CrossRef]
6. Concina, I.; Falasconi, M.; Gobbi, E.; Bianchi, F.; Musci, M.; Mattarozzi, M.; Pardo, M.; Mangia, A.; Careri, M.; Sberveglieri, G. Early detection of microbial contamination in processed tomatoes by electronic nose. *Food Control* **2009**, *20*, 873–880. [CrossRef]
7. Concina, I.; Bornšek, M.; Baccelliere, S.; Falasconi, M.; Sberveglieri, G. Electronic Nose: A Promising Tool For Early Detection Of Alicyclobacillus spp In Soft Drinks. In *AIP Conference Proceedings*; Pardo, M., Sberveglieri, G., Eds.; American Institute of Physics: College Park, MD, USA, 2009; Volume 1137, pp. 535–536. [CrossRef]
8. Sberveglieri, G.; Zambotti, G.; Falasconi, M.; Gobbi, E.; Sberveglieri, V. MOX-NW Electronic Nose for detection of food microbial contamination. In Proceedings of the IEEE SENSORS 2014, Valencia, Spain, 2–5 November 2014; pp. 1376–1379. [CrossRef]
9. Feng, L.; Dai, H.; Song, X.; Liuc, J.; Meia, X. Gas identification with drift counteraction for electronic noses using augmented convolutional neural network. *Sens. Actuators B Chem.* **2022**, *351*, 130986. [CrossRef]
10. Sciosense. Available online: https://www.mdpi.com/1424-8220/10/4/3882 (accessed on 1 November 2021).
11. Finn, C.; Abbeel, P.; Levine, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. JMLR.org. *arXiv* **2017**, arXiv:1703.03400.
12. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.