Data-centric Performance Improvement Strategies for Few-shot Classification of Chemical Sensor Data

Bhargavi Mahesh, Teresa Scholz, Jana Streit, Thorsten Graunke, Sebastian Hettenkofer

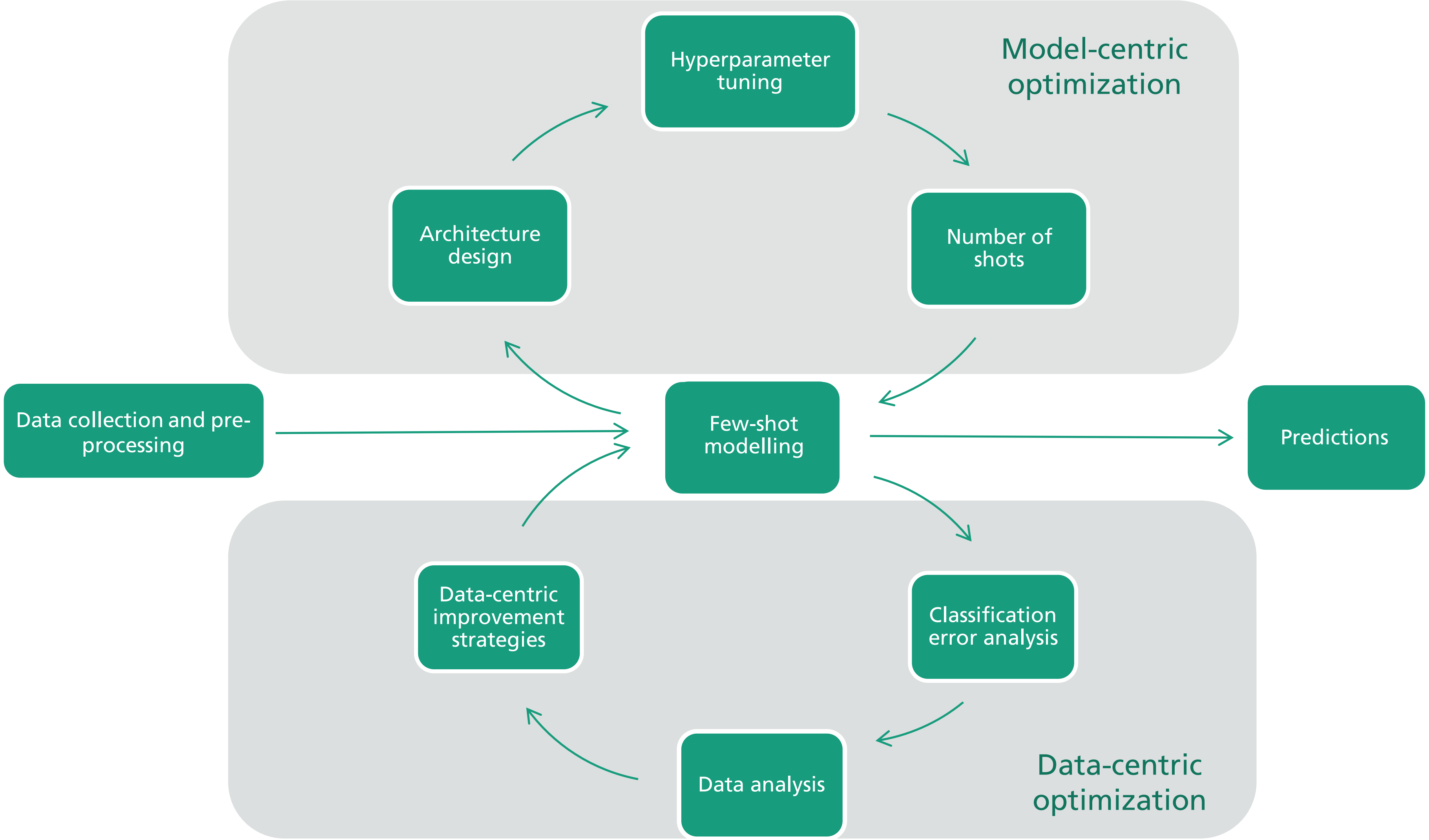
8th International Electronic Conference on Sensors and Applications

Fraunhofer
IIS

Outline

- Motivation/Introduction
- Related work
- Data description
- Few-shot classification: Methodology and illustration
- Few-shot classification: Application and results
- Classification analyses
  - Sample screening
  - Error analysis
- Data analysis
- Data-centric performance improvement strategies
- Conclusions

Overview



Model-centric optimization
Hyperparameter tuning
Architecture design
Number of shots
Data collection and pre-processing
Few-shot modelling
Predictions
Data-centric improvement strategies
Classification error analysis
Data analysis
Data-centric optimization

Motivation

- MOX sensors are an inexpensive alternative to classic methods (GS, MS, FTIR) for gas detection and classification.
- Specific layers and temperature modulation to increase selectivity to target volatile compounds.
- Problem: Temperature modulation yields long data collection process and hinders real-time inferencing.
- Solution: Rapid classification strategy, such as few-shot classification.

Methodology and state of the art

- Data collected using MOX sensors serve as a ``fingerprint'' of the volatile components emitted by the measured substance.
- Detect any deviation of the norm using pattern recognition algorithms!
- Application in food industry[1,2]:
  - Food authenticity: adulterated milk, cow ghee, olive oil, saffron, cherry tomato juice
  - Geographical origin: olive oil, orange juice, meat, milk or honey
  - Food spoilage: microbial contamination in soft drinks and juices
  - Freshness: meat, eggs or fish
  - Ripeness: fruit and wines

1 Gliszczynska-Swigło, A.; Chmielewski, J. Electronic Nose as a Tool for Monitoring the Authenticity of Food. A Review.FoodAnalytical Methods2017,10.

2 Berna, A. Metal Oxide Sensors for Electronic Noses and Their Application to Food Analysis.Sensors2010,10, 3882–3910.

Dataset Description

- Volatile organic compounds (VOCs) in fruit juices/odors
  - Juices: Apple, Orange, Blackcurrant, and Multivitamin
  - Target VOCs: esters, aldehydes, ketones, alcohols, hydrocarbons [1]
- Eight AS-MLV-P2 MOX sensors
  - Four exposed to juice headspace and air (phases)
  - Four exposed to air
- Phase duration = 10 minutes
- Temperature cycle of 12 seconds (10 samples per second)
- Measurement of interest: Sensor voltage over time
- Collected over 4 non-consecutive days

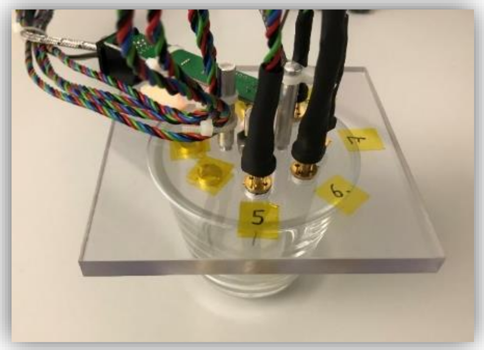


Fig 1: Data collection setup

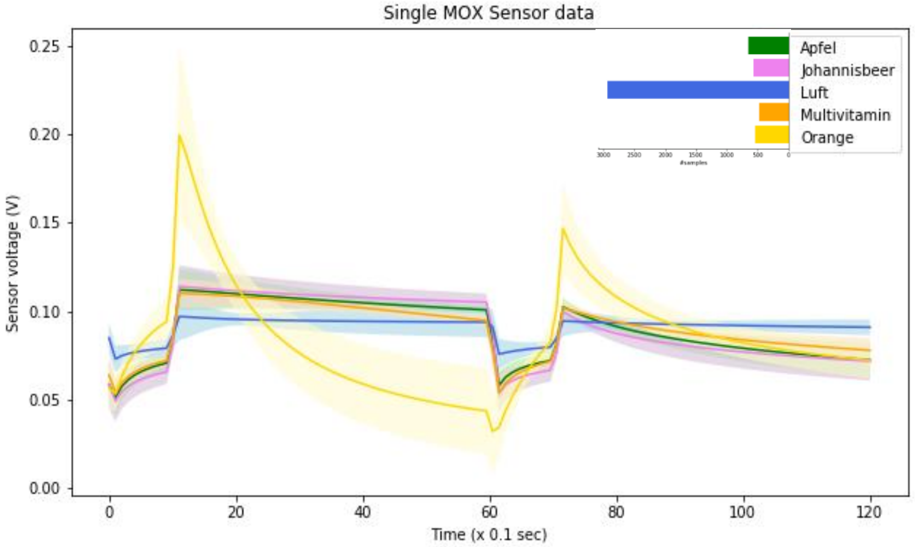


Fig 2: Normalized data from a single MOX sensor

Method: Few-shot Classification

Meta-training stage

- Input: Base classes, i.e. samples not containing novel class
- Prediction: 4 (base) classes
- Learn f_Θ and C(.|w)
- 200 epochs with learning rate 0.01

Meta-testing/ Fine-tuning stage

- Input: Base class samples + 5, 25, 40, 75 shots of novel juice
- Prediction: 5 classes (4 base classes + novel class)
- Freeze f_Θ
- Fine-tune C(.|w)
- 200 epochs with learning rate 0.001

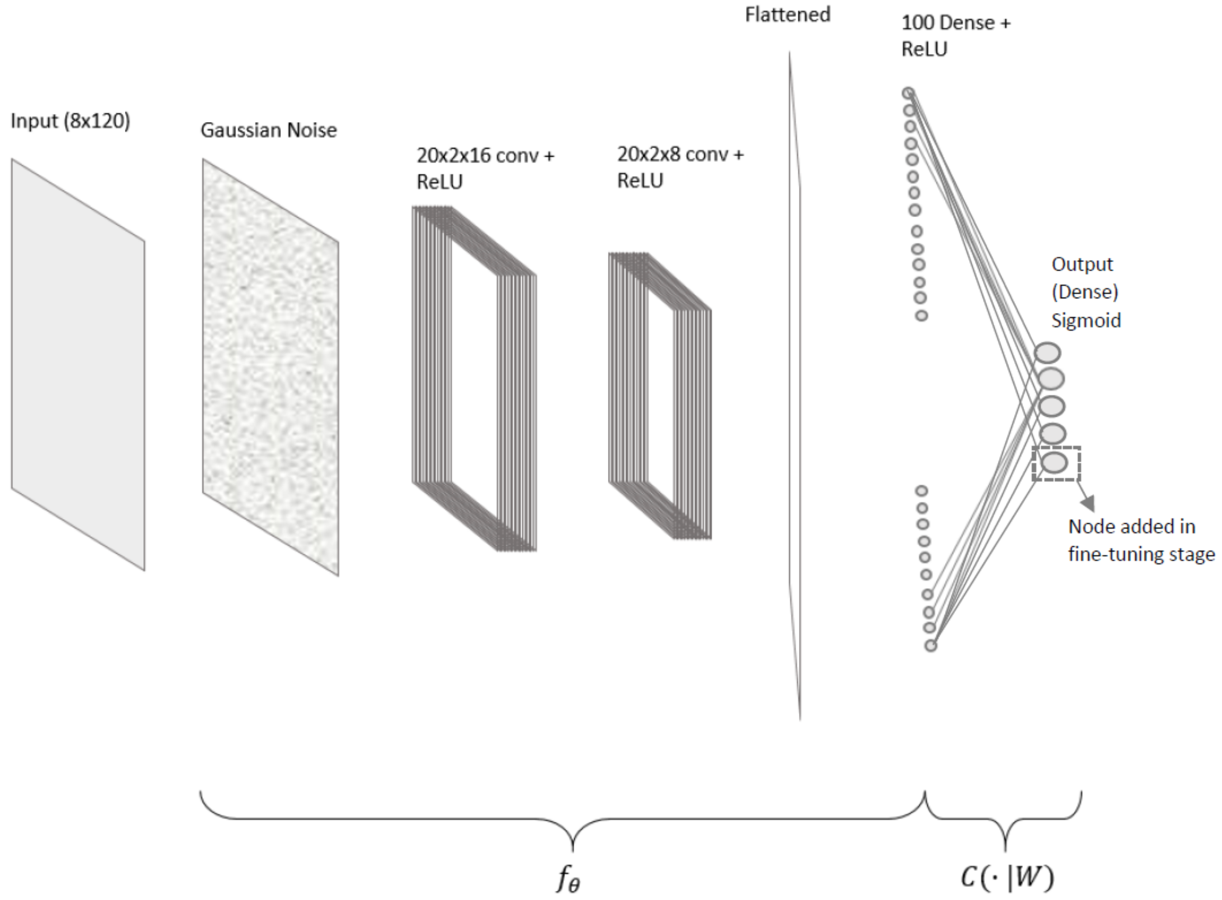
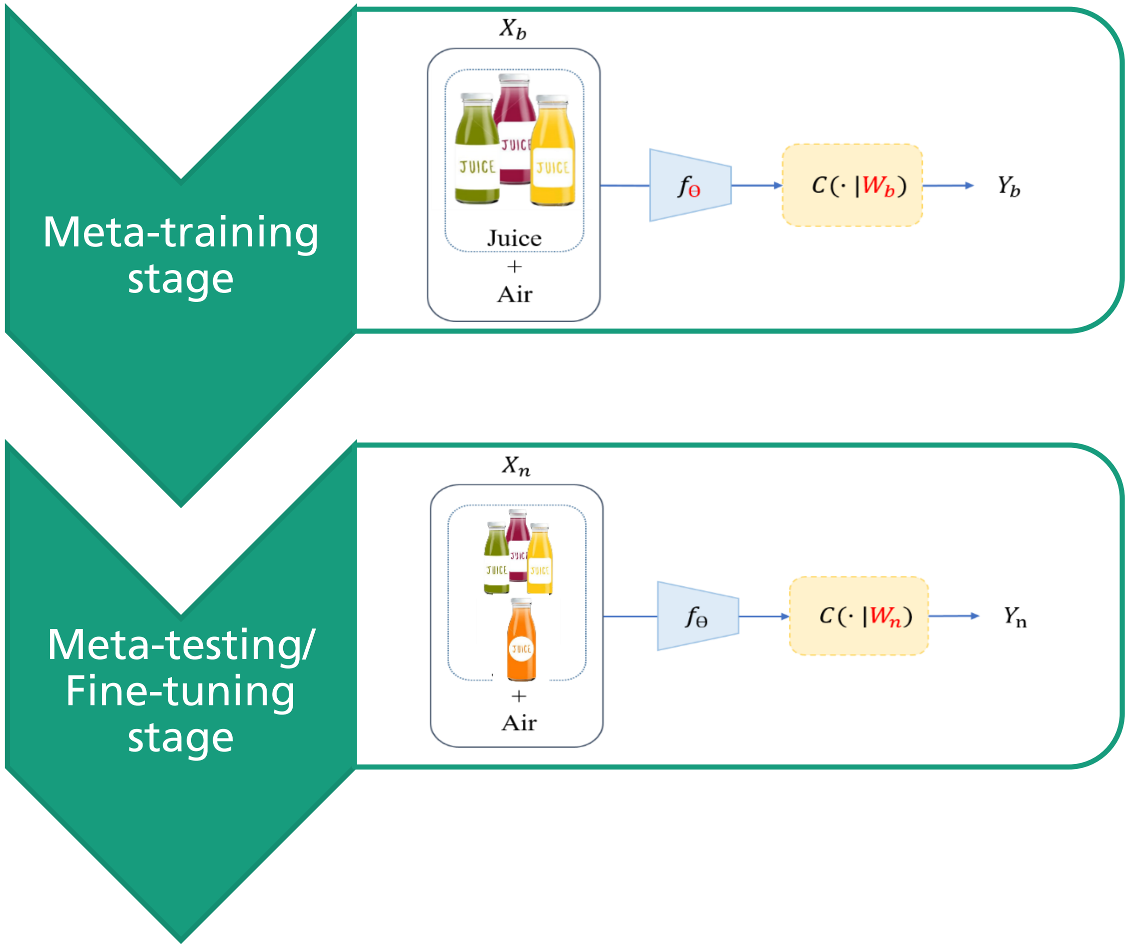


Fig 3. Architecture of the CNN used for multiclass classification.

Illustration: Few-shot Classification



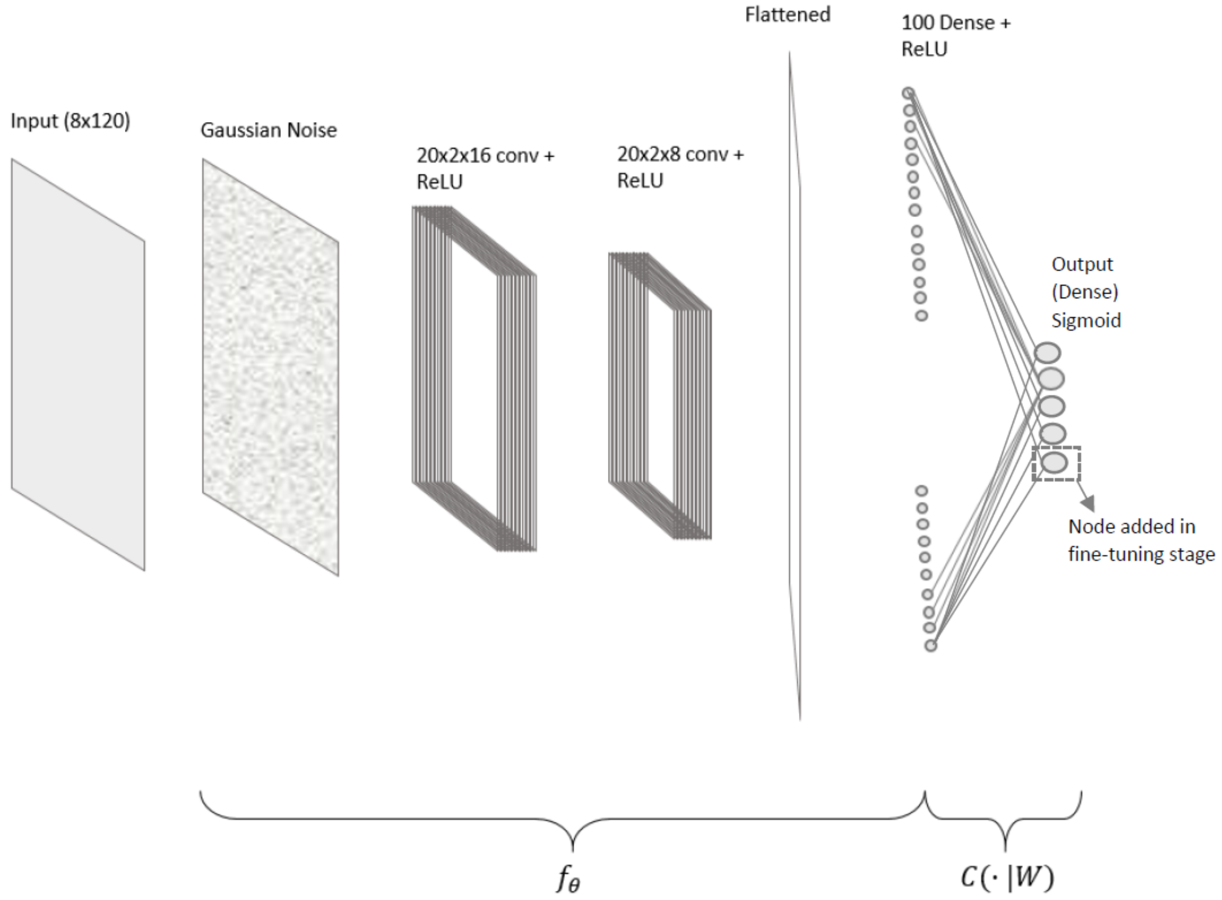


Fig 3. Architecture of the CNN used for multiclass classification.

Experiments

- Four experiments with each of the four juices in new juice class, X_n
  - One experiment = Meta-training stage + Meta-testing (fine-tuning) stage
  - Both stages have training and validation/test data
  - Classification accuracy as metric.
- Test post meta-training
  - Zero-shot test – How different is the new juice odor from base juices?
- Tests post fine-tuning
  - 1-shot/ 25-shot/ 50-shot/ 75-shot test – To what extent **1, 25, 50, 75 inputs** of new juice odor contributes to learning of the new odor?
  - Catastrophic forgetting tests – How well can the model remember the base classes after fine-tuning?

Baseline Few-Shot Classification (E0) Results

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

Tab 1: Average validation, test and catastrophic forgetting test accuracies

Sample Screening

- Design: Each phase = 1 juice measured for ~10 minutes.
- Observation: Not all measurements in a phase are representative of the juice class.
- Hypothesis: Contamination affected initial measurements of each phase.

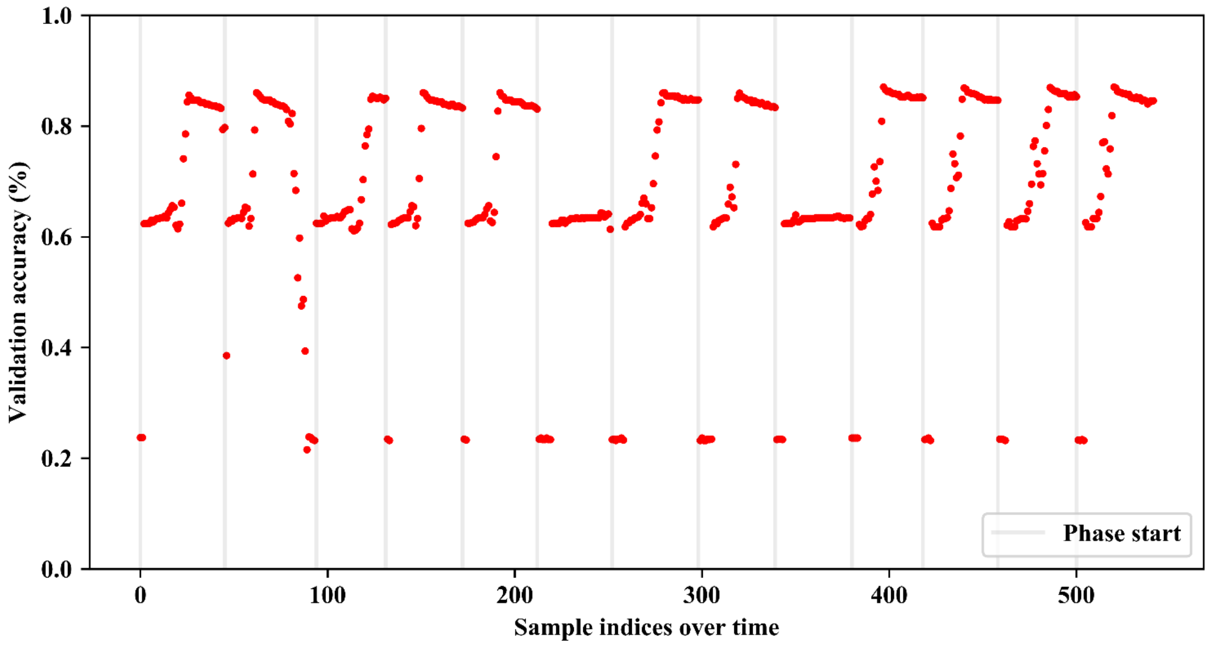


Fig 4. Test accuracies obtained when the classifier is finetuned on every sample of Orange juice. Samples from the beginning of the phase often resulted in reduced performance

Error Analysis

- Observations:
  - Misclassifications for juice decrease as shots increase.
  - More than 60% of air misclassifications are related to the previous juice class

  => Another indication of contamination

	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

Tab 2: Misclassification (M) out of 3220 per class and previous phases' influences (I) on the misclassifications.

Data Analysis - Class Separability and Contamination

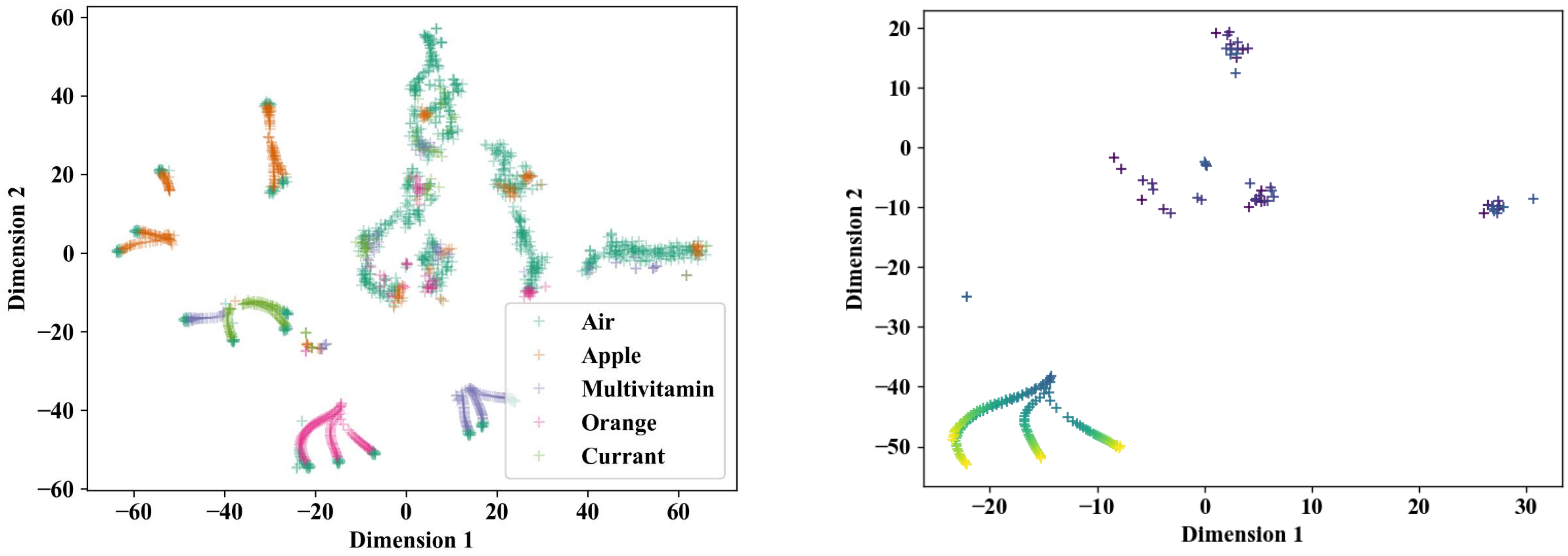


Fig 5: All (left) and just orange juice (right) measurements projected into t-SNE plane.

Contamination through previously measured substance visible in the t-SNE plot.

Data Analysis - Class Separability and Contamination

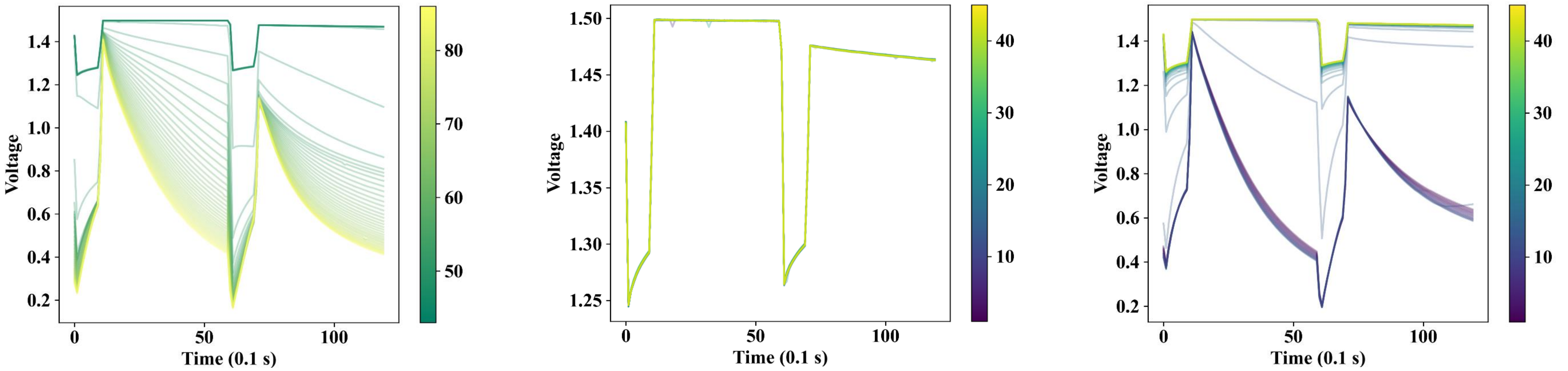


Fig 6: Measurements collected during a phase of orange juice measurements following a phase of air measurement (left), air measurements following a phase of orange juice measurement (right) and reference measurements of the room air collected at the same time (middle).

- Deviation from „air“ signal to „orange“ signal over measurement phase (left).
- Contamination of „air“ signal through previously measured orange juice (right).

Data-centric Performance Improvement Strategies

- Baseline and three data-centric improvement strategies:
  - E0: Baseline
  - E1: Dropping first and last 10 samples for fine-tuning
  - E2: Dropping first half of phase for fine-tuning
  - E3: Dropping first half of phase and retraining
- Results:
  - Baseline (E0): No significant dropping of samples
  - E1 performed the best in terms of average validation accuracy for 75 shots, E2 for 25 and 50 shots, and E3 for 5 shots.
  - E1 had least influence on air misclassification and E3 the least juice misclassification.

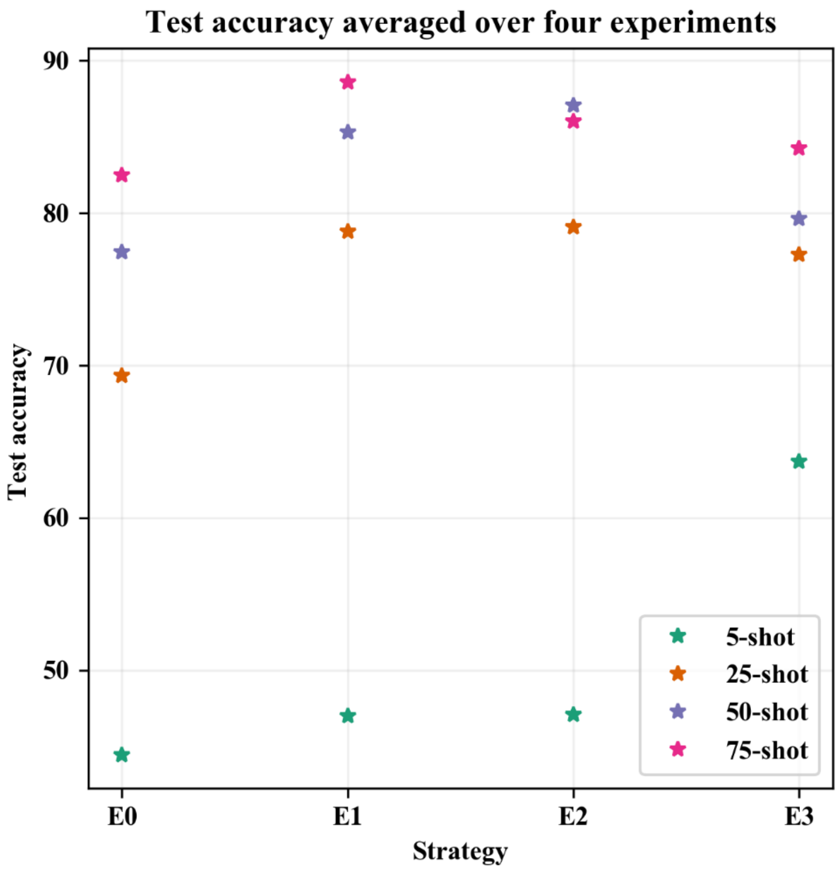


Fig 7. Test accuracies averaged over shots for four strategies

Conclusion

- Demonstrated the impact of data quality on prediction performance.
- Showed different data-centric performance improvement strategies
  - Classification improvement seen through strategy E1 and E2 – removal of 10 samples from the beginning and end of the phases and removal of initial half phase, respectively, before fine-tuning.
  - E1 had least influence on air misclassification and E3 the least juice misclassification.
- Susceptibility to contamination of MOX sensor data hinders rapid learning.
- Feasible for binary classification with distinct classes, e.g., Air vs. Juice.