ACS SENSORS

Article

Cite This: ACS Sens. 2019, 4, 2101−2108

pubs.acs.org/acssensors

Chemiresistive Sensor Array and Machine Learning Classification of Food

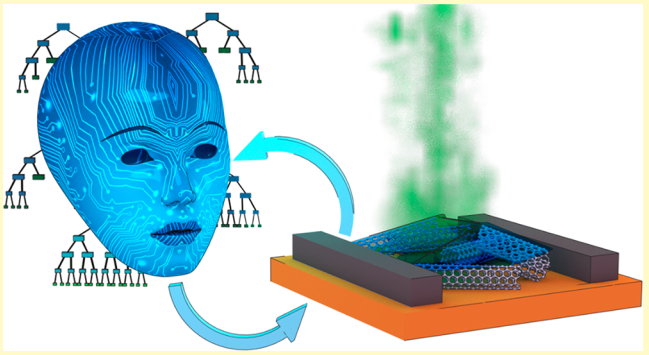
Vera Schroeder,[†,‡] Ethan D. Evans,[§,‡] You-Chi Mason Wu,[†] Constantin-Christian A. Voll,[†] Benjamin R. McDonald,[†] Suchol Savagatrup,[†] and Timothy M. Swager*[,†]

[†]Department of Chemistry and Institute for Soldier Nanotechnologies, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge Massachusetts 02139, United States

[§]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge Massachusetts 02139, United States

Supporting Information

ABSTRACT: Successful identification of complex odors by sensor arrays remains a challenging problem. Herein, we report robust, category-specific multiclass-time series classification using an array of 20 carbon nanotube-based chemical sensors. We differentiate between samples of cheese, liquor, and edible oil based on their odor. In a two-stage machine-learning approach, we first obtain an optimal subset of sensors specific to each category and then validate this subset using an independent and expanded data set. We determined the optimal selectors via independent selector classification accuracy, as well as a combinatorial scan of all 4845 possible four selector combinations. We performed sample classification using two models—a *k*-nearest neighbors model and a random forest model trained on extracted features. This protocol led to high classification accuracy in the independent test sets for five cheese and five liquor samples (accuracies of 91% and 78%, respectively) and only a slightly lower (73%) accuracy on a five edible oil data set.



KEYWORDS: *chemical sensor, sensor array, carbon nanotubes, electronic nose, authentication, time series classification, feature selection, nearest neighbors*

The three major functions of the human olfaction system are related to social communication,[1,2] avoidance of environmental hazards,[3] and ingestive behavior.[4−6] Our ability to successfully differentiate between stimuli in these applications is enabled by more than 1000 distinct olfactory receptors (ORs).[7] Activation of an OR by an odorant generates an electronic signal that is transmitted to the brain.[8] Each OR recognizes several odorant molecules, and each odorant molecule is recognized by several distinct ORs. The identification of a specific odorant molecule, or mixture thereof, is confirmed by the activation of a specific combination of ORs.[8,9] Thus, the olfaction system of humans and animals can identify complex mixtures of chemicals and not just pure chemicals against an odorless background.

Array-based chemical sensors follow a similar biomimetic approach. Detection of chemical inputs is achieved through a combination of multiple sensing channels where each channel responds to several analytes.[10,11] Single-walled carbon nanotube (SWCNT) chemiresistors and carbon nanotube-based electrochemical sensors have been shown to provide suitable platforms for array-based detection of various gases.[12] Several sensor arrays comprising CNT-based sensing channels have been used to discriminate between single volatile organic compound (VOC) vapors,[13−18] inorganic gases,[19,20] and biological samples.[21−24] However, few reports have been published on the differentiation between food samples, among them are the determination of caffeine content in coffee,[25] the electrochemical detection and differentiation between rice wines,[26] electrochemical determination of capsaicin content of hot sauces,[27] and chemiresistive differentiation of liquors using multiwalled CNT/polymer composites.[28,29]

Herein, we differentiate between complex odors using an array of 20 SWCNT-based chemiresistive sensors (Figure 1). As a proof-of-principle system, we investigated samples of different food item categories—cheese, liquor, and edible oil. The odor compounds of these food items include distinct combinations of a multitude of sulfur compounds, alcohols, ketones, aldehydes, esters, organic acids, alkanes, and aromatic compounds.[30−40] We strategically designed the sensor array to include selectors that can interact with odor compounds containing these functional groups. In addition to the choice of selectors, a similarly sophisticated analysis of the sensing data is necessary to enable successful differentiation.

Received: May 6, 2019
Accepted: July 11, 2019
Published: July 24, 2019

DOI: 10.1021/acssensors.9b00825
ACS Sens. 2019, 4, 2101−2108

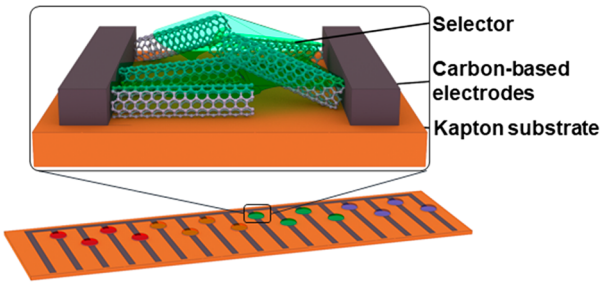


Figure 1. Schematic of sensing device with carbon-based electrodes deposited on a polyimide substrate. The active layer of SWCNTs and selectors is deposited between the electrodes. All 20 selectors are listed in Table S1.

Previously, classification in gas sensing applications has been performed by discriminant factor analysis (DFA) to diagnose disease,[22] principal component analysis (PCA) to differentiate between VOCs,[13] a neural network model to differentiate between formaldehyde and ammonia gas,[19] and a support vector machine to differentiate between NO_2, NH_3, EtOH, and acetone.[41] These methods generally only use a limited number of features from the sensing response (sometimes with additional metadata concerning the specifics of the sensing apparatus). In contrast, we use the entire time series of the sensor response or a large set of diverse features extracted from the exposure-recovery cycle to perform accurate classification.

Traditional time series classification has relied on nearest neighbor approaches via either Euclidean or dynamic time warping-based (a means of aligning time series) distance comparisons used to calculate a similarity between two time series.[42−44] Similarly, time series decompositions, where the data is represented as piecewise linear[45] or piecewise polynomial segments,[46] as well as feature extraction techniques, combined with more traditional machine learning models like random forests (RFs) or support vector machines[47] have also been proposed. Lately, ensembles of models have become a popular choice for high performance time series classification leveraging potentially multiple techniques with different models.[48] Often, improvement over the simple *k*-nearest neighbors algorithm (KNN) is difficult, questioning the need for these highly sophisticated methods[49] that often require larger training data sets. In this study, we found a RF model trained on features extracted from the sensing data (featurized-RF model, f-RF) to be a highly accurate method for analyzing time series data of very similar food samples. This model can classify samples of cheese, liquor, and edible oil with up to 91% accuracy.

EXPERIMENTAL SECTION

Sensor Array. The sensing substrates are made of carbon-based electrodes (1 mm gap width) on a polyimide support. One substrate contains 16 individual working electrodes with a shared counter electrode separated by a gap of 1 mm. The active material of the sensor channels consists of SWCNTs as a transduction material and a chemical selector (S1−S20) as a selective moiety. The active material is deposited between the electrodes in a one- or two-step process. For the one-step process, 1 μL of a dispersion of selectors S17, S18, or S19 and SWCNTs in *ortho*-dichlorobenzene (*o*-DCB) was drop-casted between the electrodes and dried in vacuo. For the two-step process, 1 μL of a dispersion of SWCNT in *o*-DCB was drop-casted between the electrodes and dried in vacuo, then 1 μL of a solution of the selector S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, or S20 was drop-casted on top of the SWCNT layer and again dried in vacuo. Chemical structures of the selectors and detailed experimental procedures can be found in the Supporting Information, and all solvents, concentrations, and dispersion parameters are listed in Table S1.

Gas Sensing Data. During the gas sensing experiment, 0.100 V was applied across the sensing electrodes and the resulting current was measured as a function of time. All gas sensing data was collected using air (approximately 40% relative humidity) as the carrier gas at a flow rate of 100 mL/min. The sample gas was generated by heating the food sample to 30−50 °C and collecting the headspace over the sample at a distance of 3 cm. During the gas measurement, the active material was exposed to the sample gas for 120 s, followed by a 600−900 s recovery period under air flow.

Machine Learning Model Training. We used 300 s of sensing data (120 s of exposure and 180 s of recovery) to train either a KNN model (k = 1) that directly used the time series data or a f-RF (100 estimators) using features extracted via tsfresh. Tsfresh extracts 794 features ranging from the coefficients of a continuous wavelet transform or fast Fourier transform to parameters like time series length, mean, max, and median among many others.[50] Data from each selector were used to train individual models for classification using Scikit-learn.[51] The 20 selector-three-class models were trained with a stringent 0.67: 0.33 train: test split, while the four selector-five-class models were trained with a 0.80: 0.20 train: test split. The combinatorial selector scan was performed on all 4845 possible combinations of 4 out of 20 selectors. Final model accuracy was assessed via a 50× repeated model training on randomly shuffled train: test splits. Complete protocols for the computational analysis and model building can be found in the Supporting Information.

RESULTS AND DISCUSSION

Collection of Sensing Data. To develop a sensing system that can differentiate between complex organic odor mixtures, the choice of selectors is critical. We assembled an array of 20 selectors including: transition metal complexes (S1, S2, S3, and S4) to bind organic acids[52] and sulfur-containing compounds, ionic liquids (S5, S6, S7, and S8) to interact with ketones, aldehydes, alkanes, and aromatic compounds;[53] porous polymers (S9, S10, S11, and S12) to sequester a large number of organic vapors;[54,55] cavitand molecules (S13, S14, S15, S16, and S20) for detection of aromatic compounds and alcohols with size exclusion properties;[15,17] and metalloporphyrins (S17 and S18) to bind amines, alcohols, ketones, alkanes, and aromatic compounds.[13,56]

The input data for our machine learning algorithms consist of the chemiresistive responses of the active material to the sample odor. We examined five categories of samples, including cheeses (cheddar, cream cheese, Cambozola, Mahón, and pecorino), liquors (gin, tequila, rum, vodka, and whiskey), and edible oils (canola, olive, coconut, toasted sesame, and walnut oil). During the gas measurement, the active material—SWCNT and selector—was exposed to the sample odor for 120 s, followed by a 600−900 s recovery period under air flow. The response is recorded as the conductance through the active material at a constant applied voltage (0.100 V). Figure 2 shows the representative response of average change in conductance normalized to the conductance at the start of the exposure ($\Delta G/G_0$) for one selector (S4) toward the odor of three different cheeses: cheddar, Mahón, and pecorino. The data used for classification consists of all time points in the 300 s time period after the start of the exposure (120 s of exposure, and 180 s of recovery, purple shaded area in Figure 2).

Selector Analysis on Three-Class Data Sets. When designing a sensor array, including sensor channels that have insignificant signal or the same signal for all samples weakens

DOI: 10.1021/acssensors.9b00825
ACS Sens. 2019, 4, 2101−2108

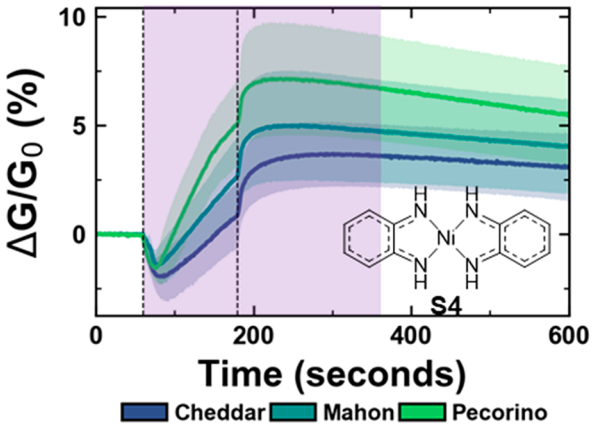


Figure 2. Example of sensing response for one selector (S4) toward cheddar, Mahón, and pecorino. The response is represented as a change in conductance normalized to the conductance at the start of the exposure ($\Delta G/G_0$). The exposure starts at t = 60 s and ends at t = 180 s (marked by dashed vertical lines). The response is an average of 12 separate sensing experiments; the blue/green shaded areas represent the standard deviation of the response. The purple shaded area represents the data used for classification.

the overall performance of the array. To ensure we only use selectors that have high predictive performance, we identified which of the 20 selectors demonstrates the highest accuracy in differentiating between the food samples. For this, we collected 12 sets of sensing data for each one of the 20 selectors and for three items from each category (2160 individual sensing traces). To evaluate the classification utility of the selectors in our sensor array, we built two models. The first model was the f-RF model trained on a set of features extracted using the tsfresh computational package[50] from each selector time series. The second model directly leverages the time-dependent nature of the data via a KNN model for which a Euclidean distance metric was used to measure time series similarity. Importantly and in contrast to typical analyses, both models rely on data from both the initial exposure and recovery time period, as the curvature and absolute values for both of these periods provides valuable time-based patterns. The models are used to differentiate between three food items from each category: for cheese, the classes include cheddar, Mahón, and pecorino; for liquor, the classes include rum, vodka, and whiskey; and for edible oil, the classes include canola, olive, and walnut.

Each of the 20 selectors was analyzed independently using both models on held-out test sets; 67% of the collected data sets were used for training the model (training set) and the

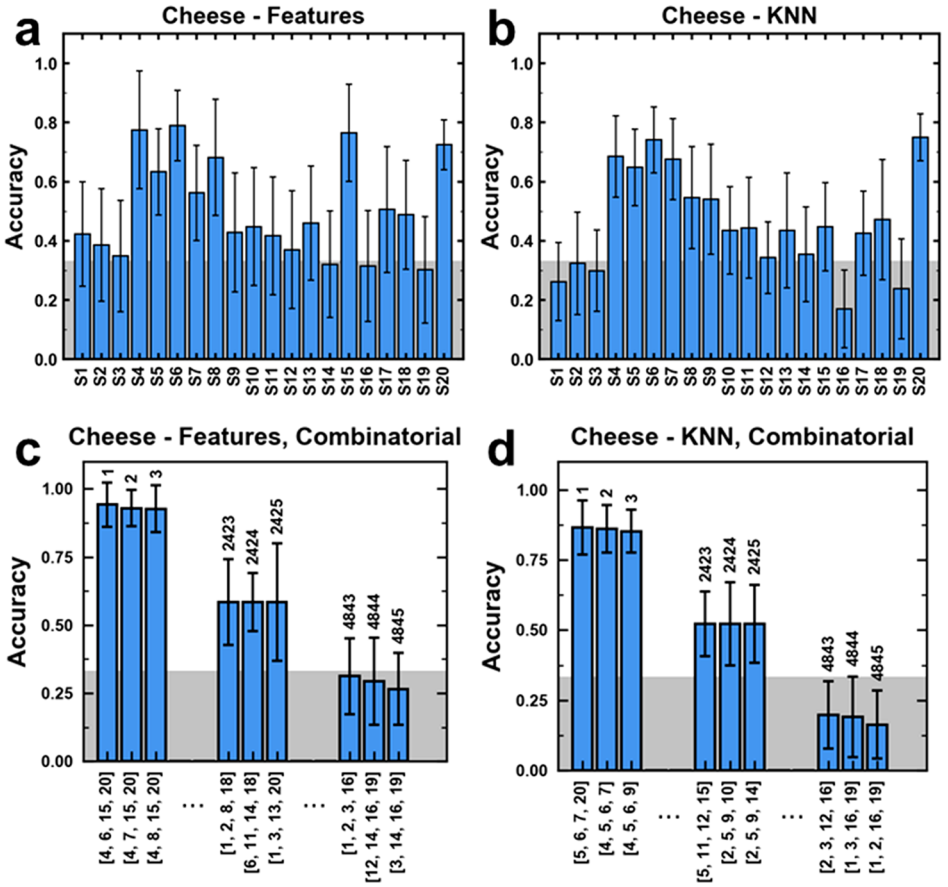


Figure 3. (a, b) Selector accuracies for both the (a) f-RF and (b) KNN models using single selectors for differentiating between cheddar, Mahón, and pecorino cheese. The shaded, gray area corresponds to random guessing (33% accuracy), selectors with accuracies around this threshold do not assist in this particular application. (c, d) Example results from the combinatorial selector scan for both the (a) f-RF and (b) KNN models on the cheese data set (cheddar, Mahón, and pecorino) using combinations of four selectors. Plotted are the top three (1−3), three medium (2423−2425), and the bottom three (4843−4845) combinations. Each selector combination was trained 27 times; the average accuracies and standard deviations are plotted.

DOI: 10.1021/acssensors.9b00825
ACS Sens. 2019, 4, 2101−2108

remaining 33% were used to test the model (test set). The models assign the samples of the test set into classes, such as cheddar, Mahón, and pecorino for the cheese samples.

The accuracy for each selector was then calculated as the fraction of correct assignment made by the f-RF or KNN model corresponding to that selector:

$$\text{accuracy} = \frac{\text{number of correct assignments}}{\text{number of total assignments}}$$

Figure 3a and b shows the accuracies for f-RF and KNN models when differentiating between three cheese samples (Figure S13 shows analogous graphs for liquor and edible oils). Selectors that performed well in an f-RF model often showed high accuracies in the KNN model, suggesting the importance of specific selectors for each use case. Similarly, selectors showing near random performance in one model, often display commensurate classification ability in the other model. S4 for example, demonstrates reasonable test set accuracy for classification of the three cheeses using the f-RF (0.78 ± 0.20) and the KNN (0.69 ± 0.14) model. Between different data sets (cheese vs liquor for instance) different selectors exhibit high performance; selectors working well in one use-case may have lower classification ability in another (Figure S13). This observation can be ascribed to the difference in chemical makeup of the items in these categories. One of the selectors showing high accuracy in the cheese use-case is S4, a nickel bis(ortho diiminosemiquinate) known to detect organic acids.[52] S5 and S6 are two methylimidazolium-based ionic liquids, 1-(2-(2-(2-hydroxyethoxy)ethoxy)ethyl)-3-methylimidazolium chloride and 1-(nonyl)-3-methylimidazolium hexafluorophosphate, designed to interact with aldehydes and ketones.[53] Organic acids, aldehydes, and ketones are all important aroma compounds with moderate volatility, which can be found in the odor of cheese.[31,57,58]

Combinatorial Selector Scan. While using each selector individually led to moderate success in classification, we suspected that using combinations of multiple selectors will enable classification with higher fidelity. To exhaustively determine an optimal set of selectors for model building, analysis, and collection of additional data sets, we trained and tested f-RF and KNN models on all possible combinations of 4 selectors (4845 combinations).

The combinatorial analysis revealed many selector combinations with similar scores, nonetheless common trends emerged internal to a data set category (Figure 3c,d, Figure S15). These trends often mapped to the results from the individual selector analysis. Across item categories, different selector combinations showed varying levels of usefulness for each individual classification problem (Figure S15), highlighting the need for tuning the selector panel when confronted with a new category of interest.

Focusing only on the highest accuracy selector combination in all three item categories, the f-RF models typically outperformed the simpler KNN models (Figure 4). We propose that this results from the extraction of more descriptive temporal features not accounted for in an isolated time point-to-time point distance calculation used in the KNN model in combination with a more expressive random forest model. Indeed, PCA on the features extracted from the three-item data sets often showed a degree of class separability using only the first three principal components (Figure S16).

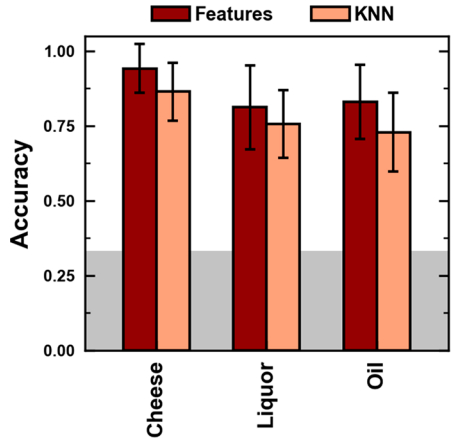


Figure 4. Selector scan results showing only the highest accuracy selector combinations for each use case.

On the basis of the information from the combinatorial selector analysis (Figure 3c and d), along with that of the individual selector analysis (Figure 3a and b), we picked four selectors for follow up analysis and validation for each sample category. For the cheese use-case, we identified S4, S5, S6, and S20 as a high-performing selector combination. We performed the analogous combinatorial analysis for the other two use-cases—liquor and edible oil—to identify use-case specific high-performing selector combinations (Figures S13 and S15). The selector combinations for liquor (S7, S8, S12, and S15) and the selector combination for edible oil (S5, S7, S8, and S13) both consist of three methylimidazolium-based ionic liquids and one calix[4]arene. The combinatorial approach leads to higher classification accuracy across all use-cases and both models. For example, the maximum test set accuracy for a single selector when differentiating between cheddar, Mahón, and pecorino cheese is 0.79 ± 0.12 (S6), whereas the combinatorial approach reaches 0.94 ± 0.08 (f-RF). Moving forward, all analysis was performed using the combinatorial approach.

Optimal Selector Combination Performance on Independent Five-Class Data Sets. With our optimized combination of selectors in hand, we significantly increased the difficulty of the classification problem by expanding the number of items in each category from three to five. Notably, there is the question whether the chosen selector combination can have utility for not only the presecreended items, but also for additional items from the same category. For cheese, the expanded selection includes cheddar cheese, Mahón, Cambozola, cream cheese, and pecorino; for liquor, the expanded selection includes rum, vodka, whiskey, gin, and tequila; and for edible oil, the expanded selection includes canola, olive, walnut, coconut, and toasted sesame oil. For this validation of our methodology, we collected new sets of data for all samples and retrained all models.

Even for these expanded item categories, the newly trained models show high levels of classification accuracy (Table 1). Between the three item categories, cheese was the easiest to classify with a test set accuracy of 0.91, followed by liquor at 0.78, with oil being the most challenging with a reasonable accuracy of 0.73 when using the f-RF model. The simpler KNN model displayed a reduced performance across the board with an accuracy of 0.73 for the cheese data set, while much lower values near 0.40 were observed for the liquor and oil data. These values are still significantly higher than that of

DOI: 10.1021/acssensors.9b00825
ACS Sens. 2019, 4, 2101−2108

Table 1. Optimal 4-Selector Test Set Accuracy[a] Analysis on the Five-Class Classification Problems for Both the f-RF and KNN Models

	cheese (S4, S5, S6, S20)	liquor (S7, S8, S12, S15)	oil (S4, S7, S9, S13)
f-RF	0.91 ± 0.05[b]	0.78 ± 0.08	0.73 ± 0.07
KNN	0.73 ± 0.06	0.40 ± 0.09	0.36 ± 0.08

[a]Average accuracies are calculated via 50 f-RF or KNN models trained and tested on shuffled data splits. [b]Error values are standard deviations.

random guessing (0.2) for a five-class problem. This again highlights the additional important information contained in the time series that a simple point-to-point distance measurement does not capture. Further, these results speak to the chemical compositions of the odor of the different samples, suggesting the volatiles in oil to be more similar between samples than those of cheese, making the classification problem more challenging.

Analyzing the average time course values for each sample and selector often revealed only subtle differences between the different classes with several notable cases. Typically, the overall shape of the selector response curve was similar between samples of a category (with differences in the response amplitude). For example, in the five-cheese data, the Cambozola cheese displayed similar temporal dynamics, but with a much higher selector response (Figure 5, purple line).

It should not be overlooked, however, that several visually distinct temporal traces were observed (S4 and S20 in the cheese data, Figure 5a and d; S7 in the liquor data, Figure S17a; S13 in the edible oil data, Figure S18d). In line with the lower accuracy of the five-oil data, the time course analysis revealed that many of the oils had overlapping temporal behavior in both the exposure and recovery periods with only subtle differences in the curvature. It is this subtle difference that the featurization protocol most likely leverages to gain a reasonable (0.73) accuracy but something the KNN Euclidean distance cannot overcome (0.36), leading to its significantly lower performance.

Extracted Feature Analysis. We took two approaches to better understand the extracted features, the first is a dimension reduction (PCA), while the second is a feature-by-feature importance analysis. PCA of extracted features of the five-class data sets again revealed decreased class separability going from cheese to liquor and finally to oil (Figure 6, Figure S19, S20, S21). As expected from the time course analysis, Cambozola cheese showed complete separation (along PC1). Nonetheless, after reduction of the sensing data to just a handful of principal components we observe significant overlap between the majority of the cheese classes (Figure 6). More class overlap was observed for the liquor samples, with gin and whiskey displaying the most separability

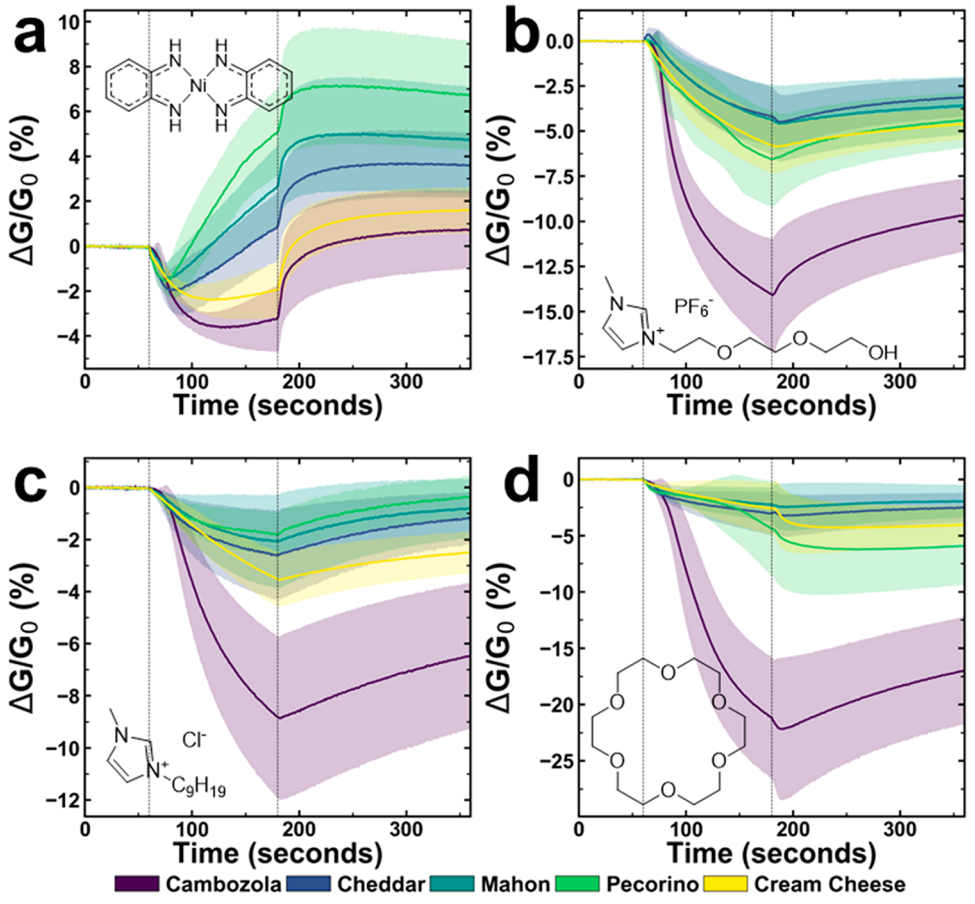


Figure 5. Sensing response for (a) S4, (b) S5, (c) S6, and (d) S20 toward five cheeses. The response is represented as a change in conductance normalized to the conductance at the start of the exposure ($\Delta G/G_0$). Each exposure starts at t = 60 s and ends at t = 180 s (marked by dashed vertical lines). Each response is an average of 40 separate sensing experiments, the shaded area represents the standard deviation of the response.

DOI: 10.1021/acssensors.9b00825
ACS Sens. 2019, 4, 2101−2108

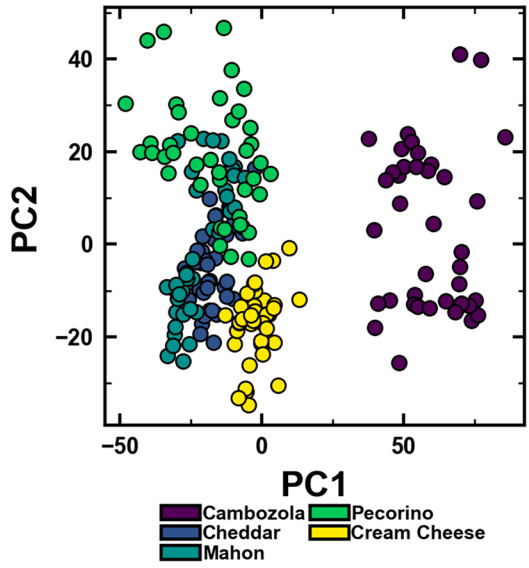


Figure 6. PCA of extracted features from the five-cheese data set showing the first two principal components.

from the other examples (Figure S20). Finally, in the edible oil data, nearly all classes showed significant overlap along the first two principal components, again corroborating the reduced model performance.

To better understand the f-RF models, we identified the features with the highest contribution to the prediction process. Figure 7 shows the 16 overall most important features—out of 794 total features—used in the f-RF model for the classification of five different cheeses. As expected from our findings of the PCA, no single feature can be used solely for classification with many features showing less than 1% feature importance. This observation holds true for all three categories (Figures 7, S23, and S24).

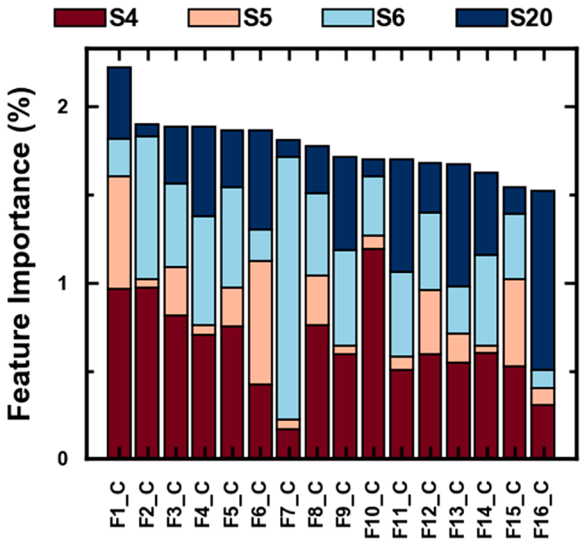


Figure 7. Top 16 overall most important features in the five-cheese f-RF model. The importance is averaged over 50 f-RF models trained and tested on shuffled data splits. The top 16 features for cheese, liquor, and edible oil are listed in Tables S2–S4.

The feature importance analysis also demonstrates that simple, descriptive features like maximum and minimum values, average values, and area under curve are not sufficient to perform this discriminative task (Table S2). In fact, the majority of the features used are coefficients from either a continuous wavelet transform (cwt features) or a fast Fourier transform (fFt features).

CONCLUSION

In this work, we demonstrated the classification of several complex odors using a chemiresistive sensing array in combination with a two-step machine learning approach. In doing so, we propose a general method for object classification that may be applied to a host of other challenging problems. With this method, we were able to differentiate between food samples with up to 91% accuracy. We envision this work to guide future research in two ways (1) our selector panel can be used to tackle a number of other challenging gas sensing problems, such as disease diagnostics, hazard detection, and food authentication, or (2) this general approach can also be used to quickly identify selectors from a large panel of potential molecules.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssensors.9b00825.

Descriptions of all chemical selectors used, additional details on sensor preparation, data collections and processing, computational models and data handling, and additional figures, including optimal selector time series, PCA, and feature importance analysis plots (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: tswager@mit.edu.

ORCID

Vera Schroeder: 0000-0002-6255-4418
Ethan D. Evans: 0000-0002-9383-2185
You-Chi Mason Wu: 0000-0002-6585-7908
Constantin-Christian A. Voll: 0000-0003-2769-3321
Timothy M. Swager: 0000-0002-3577-0510

Author Contributions

‡V.S. and E.D.E. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the KAUST sensor project CRF-2015-SENSORS-2719 and the Army Research Office through the Institute for Soldier Nanotechnologies and the National Science Foundation (DMR-1410718). S.S. was supported by an F32 Ruth L. Kirschstein National Research Service Award. We want to thank Dr. Nathan A. Romero and Dr. Monika Stolar for their useful suggestions and fruitful conversations.

REFERENCES

(1) Wysocki, C. J.; Preti, G. Facts, Fallacies, Fears, and Frustrations with Human Pheromones. *Anat. Rec.* **2004**, *281A*, 1201–1211.

DOI: 10.1021/acssensors.9b00825
ACS Sens. 2019, 4, 2101–2108

(2) Herz, R. S.; Cahill, E. D. Differential Use of Sensory Information in Sexual Behavior as a Function of Gender. *Hum. Nat.* **1997**, *8*, 275−286.
(3) Cain, W. S.; Turk, A. Smell of Danger: An Analysis of LP-Gas Odorization. *Am. Ind. Hyg. Assoc. J.* **1985**, *46*, 115−126.
(4) Stevenson, R. J. An Initial Evaluation of the Functions of Human Olfaction. *Chem. Senses* **2010**, *35*, 3−20.
(5) Zellner, D. A.; Rozin, P.; Aron, M.; Kulish, C. Conditioned Enhancement of Human's Liking for Flavor by Pairing with Sweetness. *Learn. Motiv.* **1983**, *14*, 338−350.
(6) Rozin, P.; Fallon, A.; Augustoni-Ziskind, M. The Child's Conception of Food. The Development of Contamination Sensitivity to "Disgusting" Substances. *Dev. Psychol.* **1985**, *21*, 1075−1079.
(7) Malnic, B.; Godfrey, P. a.; Buck, L. B. The Human Olfactory Receptor Gene Family. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2584−2589.
(8) Kajiya, K.; Inaki, K.; Tanaka, M.; Haga, T.; Kataoka, H.; Touhara, K. Molecular Bases of Odor Discrimination: Reconstitution of Olfactory Receptors That Recognize Overlapping Sets of Odorants. *J. Neurosci.* **2001**, *21*, 6018−6025.
(9) Malnic, B.; Hirono, J.; Sato, T.; Buck, L. B. Combinatorial Receptor Codes for Odors. *Cell* **1999**, *96*, 713−723.
(10) Albert, K. J.; Lewis, N. S.; Schauer, C. L.; Sotzing, G. A.; Stitzel, S. E.; Vaid, T. P.; Walt, D. R. Cross-Reactive Chemical Sensor Arrays. *Chem. Rev.* **2000**, *100*, 2595−2626.
(11) Peveler, W. J.; Yazdani, M.; Rotello, V. M. Selectivity and Specificity: Pros and Cons in Sensing. *ACS Sensors* **2016**, *1*, 1282−1285.
(12) Schroeder, V.; Savagatrup, S.; He, M.; Lin, S.; Swager, T. M. Carbon Nanotube Chemical Sensors. *Chem. Rev.* **2019**, *119*, 599−663.
(13) Liu, S. F.; Moh, L. C. H.; Swager, T. M. Single-Walled Carbon Nanotube−Metalloporphyrin Chemiresistive Gas Sensor Arrays for Volatile Organic Compounds. *Chem. Mater.* **2015**, *27*, 3560−3563.
(14) Schnorr, J. M.; van der Zwaag, D.; Walish, J. J.; Weizmann, Y.; Swager, T. M. Sensory Arrays of Covalently Functionalized Single-Walled Carbon Nanotubes for Explosive Detection. *Adv. Funct. Mater.* **2013**, *23*, 5285−5291.
(15) Wang, F.; Swager, T. M. Diverse Chemiresistors Based upon Covalently Modified Multiwalled Carbon Nanotubes. *J. Am. Chem. Soc.* **2011**, *133*, 11181−11193.
(16) Chatterjee, S.; Castro, M.; Feller, J. F. F. Tailoring Selectivity of Sprayed Carbon Nanotube Sensors (CNT) towards Volatile Organic Compounds (VOC) with Surfactants. *Sens. Actuators, B* **2015**, *220*, 840−849.
(17) Wang, F.; Yang, Y.; Swager, T. M. Molecular Recognition for High Selectivity in Carbon Nanotube/Polythiophene Chemiresistors. *Angew. Chem., Int. Ed.* **2008**, *47*, 8394−8396.
(18) Shirsat, M. D.; Sarkar, T.; Kakoullis, J.; Myung, N. V.; Konnanath, B.; Spanias, A.; Mulchandani, A. Porphyrin-Functionalized Single-Walled Carbon Nanotube Chemiresistive Sensor Arrays for VOCs. *J. Phys. Chem. C* **2012**, *116*, 3845−3850.
(19) Zhang, D.; Liu, J.; Jiang, C.; Liu, A.; Xia, B. Quantitative Detection of Formaldehyde and Ammonia Gas via Metal Oxide-Modified Graphene-Based Sensor Array Combining with Neural Network Model. *Sens. Actuators, B* **2017**, *240*, 55−65.
(20) Star, A.; Joshi, V.; Skarupo, S.; Thomas, D.; Gabriel, J.-C. P. Gas Sensor Array Based on Metal-Decorated Carbon Nanotubes. *J. Phys. Chem. B* **2006**, *110*, 21014−21020.
(21) Silva, G. O.; Michael, Z. P.; Bian, L.; Shurin, G. V.; Mulato, M.; Shurin, M. R.; Star, A. Nanoelectronic Discrimination of Non-malignant and Malignant Cells Using Nanotube Field-Effect Transistors. *ACS Sensors* **2017**, *2*, 1128−1132.
(22) Nakhleh, M. K.; Amal, H.; Jeries, R.; Broza, Y. Y.; Aboud, M.; Gharra, A.; Ivgi, H.; Khatib, S.; Badarneh, S.; Har-Shai, L. Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules. *ACS Nano* **2017**, *11*, 112−125.
(23) Peng, G.; Trock, E.; Haick, H. Detecting Simulated Patterns of Lung Cancer Biomarkers by Random Network of Single-Walled Carbon Nanotubes Coated with Nonpolymeric Organic Materials. *Nano Lett.* **2008**, *8*, 3631−3635.
(24) Peng, G.; Tisch, U.; Haick, H. Detection of Nonpolar Molecules by Means of Carrier Scattering in Random Networks of Carbon Nanotubes: Toward Diagnosis of Diseases via Breath Samples. *Nano Lett.* **2009**, *9*, 1362−1368.
(25) Bian, L.; Sorescu, D. C.; Chen, L.; White, D. L.; Burkert, S. C.; Khalifa, Y.; Zhang, Z.; Sejdic, E.; Star, A. Machine-Learning Identification of the Sensing Descriptors Relevant in Molecular Interactions with Metal Nanoparticle-Decorated Nanotube Field-Effect Transistors. *ACS Appl. Mater. Interfaces* **2019**, *11*, 1219−1227.
(26) Wei, Z.; Zhang, W.; Wang, J. Nickel and Copper Foam Electrodes Modified with Graphene or Carbon Nanotubes for Electrochemical Identification of Chinese Rice Wines. *Microchim. Acta* **2017**, *184*, 3441−3451.
(27) Kachoosangi, R. T.; Wildgoose, G. G.; Compton, R. G. Carbon Nanotube-Based Electrochemical Sensors for Quantifying the 'Heat' of Chilli Peppers: The Adsorptive Stripping Voltammetric Determination of Capsaicin. *Analyst* **2008**, *133*, 888−895.
(28) Wang, L. C.; Tang, K. T.; Chiu, S. W.; Yang, S. R.; Kuo, C. T. A Bio-Inspired Two-Layer Multiple-Walled Carbon Nanotube-Polymer Composite Sensor Array and a Bio-Inspired Fast-Adaptive Readout Circuit for a Portable Electronic Nose. *Biosens. Bioelectron.* **2011**, *26*, 4301−4307.
(29) Chiu, S.-W.; Wu, H.-C.; Chou, T.-I.; Chen, H.; Tang, K.-T. A Miniature Electronic Nose System Based on an MWNT−Polymer Microsensor Array and a Low-Power Signal-Processing Chip. *Anal. Bioanal. Chem.* **2014**, *406*, 3985−3994.
(30) Milo, C.; Reineccius, G. A. Identification and Quantification of Potent Odorants in Regular-Fat and Low-Fat Mild Cheddar Cheese. *J. Agric. Food Chem.* **1997**, *45*, 3590−3594.
(31) Singh, T. K.; Drake, M. A.; Cadwallader, K. R. Flavor of Cheddar Cheese: A Chemical and Sensory Perspective. *Compr. Rev. Food Sci. Food Saf.* **2003**, *2*, 166−189.
(32) Shimoda, M.; Nakada, Y.; Nakashima, M.; Osajima, Y. Quantitative Comparison of Volatile Flavor Compounds in Deep-Roasted and Light-Roasted Sesame Seed Oil. *J. Agric. Food Chem.* **1997**, *45*, 3193−3196.
(33) Conner, J.; Reid, K.; Richardson, G. SPME Analysis of Flavor Components in the Headspace of Scotch Whiskey and Their Subsequent Correlation with Sensory Perception. *ACS Symposium Series* **2001**, *782*, 113−122.
(34) Vichi, S.; Riu-Aumatell, M.; Mora-Pons, M.; Buxaderas, S.; López-Tamames, E. Characterization of Volatiles in Different Dry Gins. *J. Agric. Food Chem.* **2005**, *53*, 10154−10160.
(35) Benn, S. M.; Peppard, T. L. Characterization of Tequila Flavor by Instrumental and Sensory Analysis. *J. Agric. Food Chem.* **1996**, *44*, 557−566.
(36) Franitza, L.; Granvogl, M.; Schieberle, P. Influence of the Production Process on the Key Aroma Compounds of Rum: From Molasses to the Spirit. *J. Agric. Food Chem.* **2016**, *64*, 9041−9053.
(37) De Souza, M. D. C. A.; Vásquez, P.; Del Mastro, N. L.; Acree, T. E.; Lavin, E. H. Characterization of Cachaça and Rum Aroma. *J. Agric. Food Chem.* **2006**, *54*, 485−488.
(38) Wiśniewska, P.; Sliwińska, M.; Dymerski, T.; Wardencki, W.; Namieśnik, J. The Analysis of Vodka: A Review Paper. *Food Anal. Methods* **2015**, *8*, 2000−2010.
(39) Buttery, R. G.; Guadagni, D. G.; Ling, L. C. Flavor Compounds: Volatilities in Vegetable Oil and Oil-Water Mixtures. Estimation of Odor Thresholds. *J. Agric. Food Chem.* **1973**, *21*, 198−201.
(40) Raghavan, S. K.; Connell, D. R.; Khayat, A. Canola Oil Flavor Quality Evaluation by Dynamic Headspace Gas Chromatography. *ACS Symposium Series* **1994**, *558*, 292−300.
(41) Pugh, D. C.; Newton, E. J.; Naik, A. J. T.; Hailes, S. M. V.; Parkin, I. P. The Gas Sensing Properties of Zeolite Modified Zinc Oxide. *J. Mater. Chem. A* **2014**, *2*, 4758−4764.
(42) Chaovalitwongse, W. A.; Fan, Y.; Sachdeo, R. C. On the Time Series K-Nearest Neighbor Classification of Abnormal Brain Activity.

DOI: 10.1021/acssensors.9b00825
ACS Sens. 2019, 4, 2101−2108

IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans **2007**, *37*, 1005−1016.
(43) Lee, Y.; Wei, C.; Cheng, T.; Yang, C. Nearest-Neighbor-Based Approach to Time-Series Classification. *Decis. Support Syst.* **2012**, *53*, 207−217.
(44) Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21−27.
(45) Keogh, E.; Pazzani, M. An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification, Clustering. *Proc. Fourth Int. Conf. Knowl. Discovery Data Min.* **1998**, 239−243.
(46) Bar-Joseph, Z.; Gerber, G.; Gifford, D. K.; Jaakkola, T. S.; Simon, I. A New Approach to Analyzing Gene Expression Time Series Data. In *Proceedings of the Sixth Annual International Conference on Computational Biology—RECOMB '02*; ACM Press: New York, New York, USA, 2002; pp 39−48.
(47) Rodriguez, J. J.; Kuncheva, L. I. Time Series Classification: Decision Forests and SVM on Interval and DTW Features. In *Proc. Workshop on Time Series Classification, 13th International Conference on Knowledge Discovery and Data Mining*, 2007.
(48) Bagnall, A.; Lines, J.; Hills, J.; Bostrom, A. Time-Series Classification with COTE: The Collective of Transformation-Based Ensembles. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 2522−2535.
(49) Xi, X.; Keogh, E.; Shelton, C.; Wei, L.; Ratanamahatana, C. A. Fast Time Series Classification Using Numerosity Reduction. *Proc. 23nd Int. Conf. Mach. Learn.* **2006**, 1033−1040.
(50) Christ, M.; Braun, N.; Neuffer, J.; Kempa-Liehr, A. W. Time Series FeatuRe Extraction on Basis of Scalable Hypothesis Tests (Tsfresh − A Python Package). *Neurocomputing* **2018**, *307*, 72−77.
(51) Pedregosa, F.; Weiss, R.; Brucher, M. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(52) Lin, S.; Swager, T. M. Carbon Nanotube Formic Acid Sensors Using a Nickel Bis(Ortho -Diiminosemiquinonate) Selector. *ACS Sensors* **2018**, *3*, 569−573.
(53) Park, C. H.; Schroeder, V.; Kim, B. J.; Swager, T. M. Ionic Liquid-Carbon Nanotube Sensor Arrays for Human Breath Related Volatile Organic Compounds. *ACS Sensors* **2018**, *3*, 2432−2437.
(54) Moh, L. C. H.; Goods, J. B.; Kim, Y.; Swager, T. M. Free Volume Enhanced Proton Exchange Membranes from Sulfonated Triptycene Poly(Ether Ketone). *J. Membr. Sci.* **2018**, *549*, 236−243.
(55) Rifai, S.; Breen, C. A.; Solis, D. J.; Swager, T. M. Facile in Situ Silver Nanoparticle Formation in Insulating Porous Polymer Matrices. *Chem. Mater.* **2006**, *18*, 21−25.
(56) Liu, S. F.; Petty, A. R.; Sazama, G. T.; Swager, T. M. Single-Walled Carbon Nanotube/Metalloporphyrin Composites for the Chemiresistive Detection of Amines and Meat Spoilage. *Angew. Chem., Int. Ed.* **2015**, *54*, 6554−6557.
(57) Christensen, K. R.; Reineccius, G. A. Aroma Extract Dilution Analysis of Aged Cheddar Cheese. *J. Food Sci.* **1995**, *60*, 218−220.
(58) Qian, M.; Nelson, C.; Bloomer, S. Evaluation of Fat-Derived Aroma Compounds in Blue Cheese by Dynamic Headspace GC/Olfactometry-MS. *J. Am. Oil Chem. Soc.* **2002**, *79*, 663−667.

DOI: 10.1021/acssensors.9b00825
ACS Sens. 2019, 4, 2101−2108