

# Robust chemical analysis with graphene chemosensors and machine learning

<https://doi.org/10.1038/s41586-024-08003-w>

Received: 6 December 2023

Accepted: 30 August 2024

Published online: 9 October 2024

 Check for updates

Andrew Pannone<sup>1</sup>, Aditya Raj<sup>1</sup>, Harikrishnan Ravichandran<sup>1</sup>, Sarbassis Das<sup>2</sup>, Ziheng Chen<sup>1</sup>, Collin A. Price<sup>1</sup>, Mahmooda Sultana<sup>3</sup> & Saptarshi Das<sup>1,2,4</sup>✉

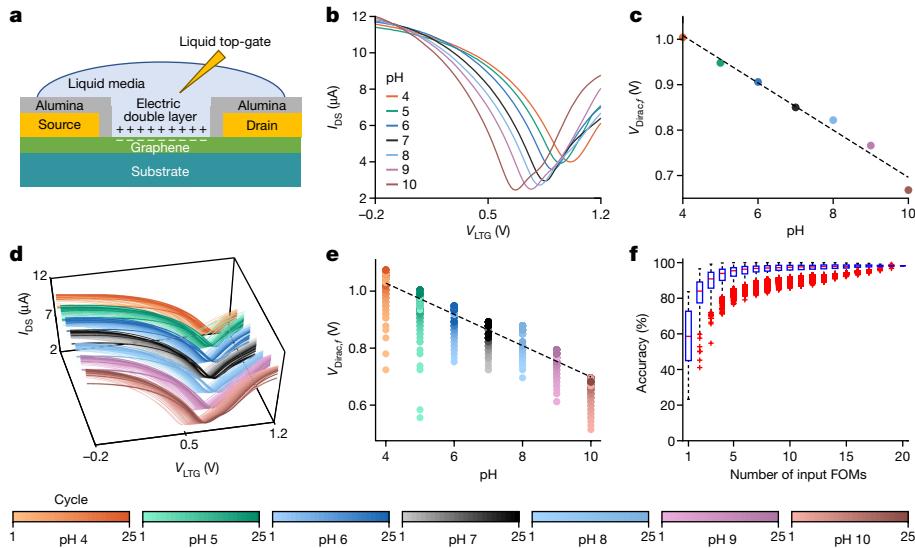
Ion-sensitive field-effect transistors (ISFETs) have emerged as indispensable tools in chemosensing applications<sup>1–4</sup>. ISFETs operate by converting changes in the composition of chemical solutions into electrical signals, making them ideal for environmental monitoring<sup>5,6</sup>, healthcare diagnostics<sup>7</sup> and industrial process control<sup>8</sup>. Recent advancements in ISFET technology, including functionalized multiplexed arrays and advanced data analytics, have improved their performance<sup>9,10</sup>. Here we illustrate the advantages of incorporating machine learning algorithms to construct predictive models using the extensive datasets generated by ISFET sensors for both classification and quantification tasks. This integration also sheds new light on the working of ISFETs beyond what can be derived solely from human expertise. Furthermore, it mitigates practical challenges associated with cycle-to-cycle, sensor-to-sensor and chip-to-chip variations, paving the way for the broader adoption of ISFETs in commercial applications. Specifically, we use data generated by non-functionalized graphene-based ISFET arrays to train artificial neural networks that possess a remarkable ability to discern instances of food fraud, food spoilage and food safety concerns. We anticipate that the fusion of compact, energy-efficient and reusable graphene-based ISFET technology with robust machine learning algorithms holds the potential to revolutionize the detection of subtle chemical and environmental changes, offering swift, data-driven insights applicable across a wide spectrum of applications.

Chemosensors that collect information related to the chemical composition of a liquid species and transduce it into an electrical signal play a crucial part in environmental monitoring<sup>5,6</sup>, healthcare diagnostics<sup>7</sup> and industrial process control<sup>8</sup>. Among chemosensors, ion-sensitive field-effect transistors (ISFETs) have emerged as a promising technology because of their remarkable sensitivity that comes in conjunction with a highly scalable design<sup>2</sup>. Bergveld first demonstrated ISFET operation in 1970 by measuring the concentration of  $\text{Na}^+$  ions in a solution using a silicon-based ISFET<sup>1,3</sup>. The advent of carbon nanotubes<sup>11</sup>, nanowires<sup>12</sup>, graphene<sup>13</sup> and related nanomaterials led to the incorporation of new channel materials into existing ISFET architectures<sup>14–17</sup>. Specifically, graphene serves as an excellent ISFET channel material because of its large surface-to-volume ratio, chemically inert basal plane, extraordinarily high carrier mobility and compatibility with covalent and non-covalent functionalization techniques that increase sensitivity and selectivity<sup>4,18–20</sup>. Recently, there have been several demonstrations of chemical sensor arrays consisting of differently functionalized graphene-based ISFETs that collectively determine the presence and concentration of ions such as  $\text{K}^+$ ,  $\text{Na}^+$ ,  $\text{Ca}^{2+}$  and  $\text{NH}_4^+$  (refs. 9,10). However, the reliability of ISFETs is impeded by various non-ideal factors including cycle-to-cycle variations, sensor-to-sensor differences and chip-to-chip disparities arising from manufacturing processes, material properties, environmental conditions and design

considerations. These challenges hinder the widespread adoption of ISFETs in commercial applications. Therefore, the development of a comprehensive approach to guarantee ISFET reliability could have a transformative impact.

In this study, we leverage the abilities of machine learning to mitigate the challenges listed above, using non-functionalized graphene-based ISFETs as a testbed. Although previous research using machine learning algorithms to analyse ISFET data has concentrated on enhancing sensitivity or selectivity, less emphasis has been placed on addressing the issue of response variation across a population of ISFETs. To begin, we use pH sensing as a benchmark to demonstrate the effectiveness of our approach. We use statistical analysis and straightforward machine learning algorithms to assess various human-derived figures of merit (FOMs) that are directly linked to the transport properties of graphene-based ISFETs. Then we attempt to correlate these human-derived FOMs with regions of interest obtained from maximally activated nodes in an artificial neural network (ANN) model trained on the same pH dataset. We observed that different regions of the graphene ISFET characteristics are used by the ANN to identify any given class, whereas other regions are used to differentiate between the classes. This underscores the efficacy of ANNs in recognizing features beyond human-derived FOMs. In other words, our sensing methodology transfers the responsibilities of variation mitigation

<sup>1</sup>Engineering Science and Mechanics, Penn State University, University Park, PA, USA. <sup>2</sup>Electrical Engineering, Penn State University, University Park, PA, USA. <sup>3</sup>Planetary Environments Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA. <sup>4</sup>Materials Science and Engineering, Penn State University, University Park, PA, USA. ✉e-mail: sud70@psu.edu



**Fig. 1 | pH sensitivity of graphene ISFETs, non-idealities and mitigation.**

**a**, Cross-sectional schematic of a representative graphene-based ISFET. **b**, Transfer characteristics, that is, source-to-drain current ( $I_{DS}$ ) plotted as a function of the liquid top-gate voltage ( $V_{LTG}$ ), of a representative ISFET for buffer solutions of integer pH values ranging from 4 to 10. **c**,  $V_{Dirac,f}$  plotted against the pH of standard buffer solutions. The fit line is created using linear regression. The sensitivity was determined to be  $S = 52 \text{ mV pH}^{-1}$  with a coefficient of determination (regression score) of  $R^2 = 0.98$ . **d**, Transfer characteristics of

seven graphene ISFETs across two separate chips. **e**, pH sensitivity of  $V_{Dirac,f}$  considering cycle-to-cycle, sensor-to-sensor and chip-to-chip variations. **f**, Box plots showing the minimum, 25th percentile, median, 75th percentile and maximum accuracies of a k-NN algorithm evaluated over all possible single and multivariate combinations of input FOMs. Red crosses indicate outliers with an accuracy that differs from the 25th or 75th percentile by a value larger than 1.5 times the interquartile range.

and classification to the machine learning model. Our results indicate that a single sensor design can be deployed across a range of application spaces without the need for sensor calibration or model retraining. Moreover, we expand the scope of our machine learning analysis beyond traditional classification tasks by introducing quantification through gradient training.

Finally, we apply these machine-learning-assisted graphene-based ISFETs to address challenges in the food industry. We create datasets that enable us to authenticate common food products, quantify instances of food adulteration and identify food safety concerns. Furthermore, we demonstrate that this sensing method can be used to simultaneously evaluate both the identity and spoilage status of various fruit juices. We believe that miniaturized graphene-based ISFET technology, enhanced by a suite of machine learning methodologies, can serve as a cost-effective platform for a wide range of chemical sensing applications in the food supply chain and beyond.

## Monitoring pH with human-derived FOMs

It has been widely demonstrated that ISFETs based on graphene exhibit sensitivity to the pH (potential of hydrogen) of a liquid solution<sup>15,21–26</sup>. However, the impacts of the aforementioned non-idealities on ISFET response variation are often disregarded, preventing their commercial adoption as a reliable technology. The cross-sectional schematic of a representative graphene-based ISFET is shown in Fig. 1a and a chip containing an array of ISFETs mounted on a printed circuit board is shown in Extended Data Fig. 1. Figure 1b shows the transfer characteristics, that is, the source-to-drain current ( $I_{DS}$ ) plotted as a function of the liquid top-gate voltage ( $V_{LTG}$ ), of a representative ISFET for buffer solutions of integer pH values. The applied  $V_{LTG}$  leads to the accumulation of complementary layers of charge carriers at the liquid–graphene interface that form an electric double layer (EDL)<sup>27–31</sup>. The EDL electrostatically modulates the conductance of the graphene channel as is reflected in the transfer characteristics. Note that several FOMs can be derived from the ISFET transfer characteristics in relation to the Dirac voltage ( $V_{Dirac}$ ), that is, the applied  $V_{LTG}$  that results in a minimum  $I_{DS}$ . In total,

we defined 20 physically relevant FOMs, which are explained in Supplementary Information Section 1.

In existing literature,  $V_{Dirac}$  is the most commonly used metric for evaluating the pH sensitivity of graphene ISFETs<sup>15,21,23–26,32</sup>. Figure 1c shows the extracted  $V_{Dirac,f}$  plotted against pH. The sensitivity was calculated as  $52 \text{ mV pH}^{-1}$  with a coefficient of determination,  $R^2 = 0.98$ , consistent with other reports<sup>23</sup>. The pH response observed in graphene ISFETs can be attributed to defects introduced during chemical vapour deposition growth, as defect-free graphene is unlikely to exhibit a pH response<sup>33</sup>. Nevertheless, the pH value predicted by  $V_{Dirac,f}$  deviates substantially from the true pH value when incorporating non-idealities such as cycle-to-cycle, sensor-to-sensor and chip-to-chip variations. Figure 1d shows the transfer characteristics of seven ISFETs across two chips measured for  $N = 25$  cycles at each pH value and Fig. 1e shows that the pH sensitivity of  $V_{Dirac,f}$  remains similar at  $55 \text{ mV pH}^{-1}$ , but  $R^2$  reduces substantially to 0.83 when including the above-mentioned variations. Therefore, identification of FOMs beyond  $V_{Dirac}$  that exhibit lesser variations due to non-idealities while retaining a high pH sensitivity and  $R^2$  value is necessary. Extended Data Fig. 2 presents the impact of each non-ideality on pH sensitivity showing that cycle-to-cycle variations, primarily attributed to drift, are most detrimental to pH sensitivity. Supplementary Information Sections 2–5 provide a detailed representation of the temporal evolution of FOMs considering these non-idealities and their impact on pH sensitivity. Extended Data Fig. 3 shows a ranking of FOMs based on their  $R^2$  value. These findings highlight the necessity of adopting compensation techniques such as drift correction because prolonged stabilization periods are not feasible for continuous pH monitoring. Previous reports have shown successful mitigation of temporal drift in ISFETs through functionalization and cross-compensation with machine learning analysis<sup>6,34</sup>. Nonetheless, these initiatives often incorporate extra layers of complexity into the sensor design<sup>35,36</sup>. We surpass existing approaches to well-known ISFET reliability challenges by emphasizing model explainability while using graphene ISFETs that have not been functionalized. Our approach fully leverages the capability of the ANN to assess sensor data and mitigate the associated non-idealities.

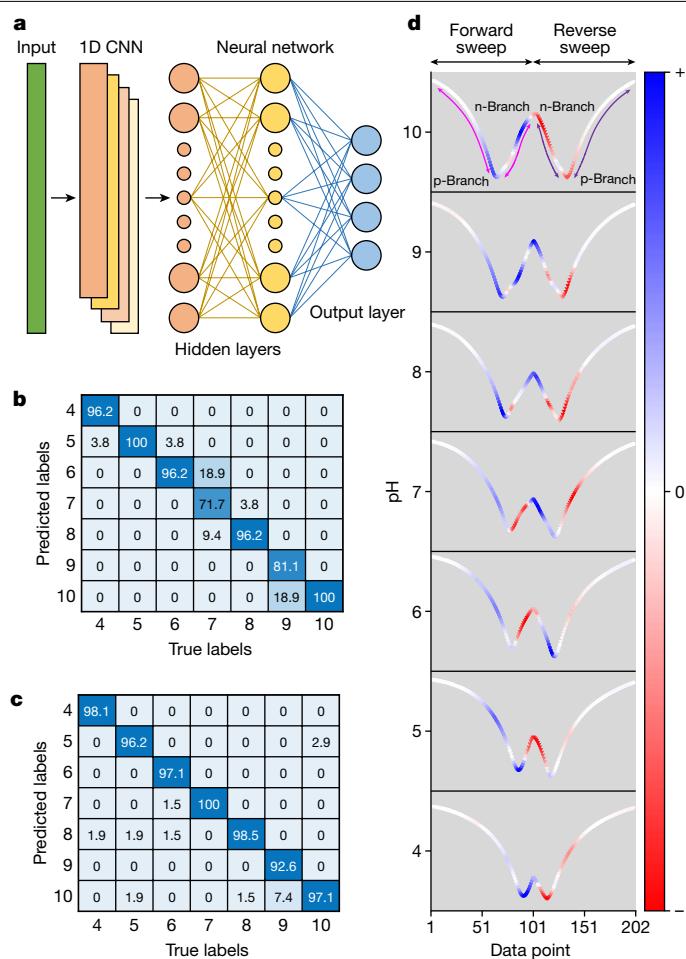
Although the classification of integer pH values through analysis with a single FOM remains challenging, we demonstrate that an analysis method that simultaneously considers multiple FOMs can easily differentiate pH classes, highlighting the importance of the rich parameter space provided by ISFETs. We used k-nearest neighbours (k-NN) (ref. 37) to evaluate the significance of each FOM with a simple classification algorithm that thrives in the increased parameter space. To eliminate the need for drift correction, all data corresponding to seven sensors, irrespective of stabilization, were collected and split randomly for training and testing. To explore the solution space, a brute force method was used to analyse the accuracy of a k-NN algorithm for all possible single and multivariate combinations of input FOMs. A box plot demonstrating the accuracy statistics is shown in Fig. 1f. We also ranked the FOMs based on k-NN accuracy as shown in Extended Data Fig. 3. The rankings largely agree with the regression scores derived earlier indicating that  $I_{\text{Dirac},f}$ ,  $I_{\text{hys}}$  and  $I_{\text{max/min},nf}$  are informative FOMs that can supplement  $V_{\text{Dirac}}$  in pH monitoring. Furthermore, Extended Data Fig. 4 shows scatter plots in three-dimensional space incorporating the three highest- and lowest-ranked FOMs as the primary axis variables.

### Use of an ANN to understand pH sensitivity

FOMs that play a notable part in enhancing the pH sensitivity of ISFETs amidst the presence of non-idealities were derived based on human expertise. It may be beneficial to develop models that are capable of learning FOMs independently through feature extraction. ANNs are universal approximators with the ability to learn any possible function mapping from input to output space<sup>38</sup>. In this section, we evaluate the performance of ANNs on the task of pH sensing using two types of input data. First, we consider FOMs extracted from ISFET transfer characteristics. Next, we allow the model to determine its own machine-derived features by inputting the entire ISFET transfer characteristics. We extensively evaluate the feature space of the trained ANN model in the form of SHAP (SHapley Additive exPlanations)<sup>39</sup> feature analysis and compare regions of importance to extracted FOMs. ANNs are often labelled as black boxes because of the lack of interpretable and explainable model predictions. In comparison, the evaluation of feature space in this work establishes the superior performance of the proposed ANN with explainability.

The architecture of the ANN is shown in Fig. 2a. We use a one-dimensional convolutional neural network as the feature extractor for both types of input data followed by three fully connected layers. The model is designed to classify input data into a total of seven classes representing each pH value. The performance of the model was evaluated using classification accuracy and SHAP feature analysis. Figure 2b shows the confusion matrix obtained after using FOMs as the training input in which an average accuracy of 91.64% was achieved with uneven prediction accuracies for different pH values. In comparison, the model trained directly on ISFET characteristics reached an average accuracy of 97.09% with balanced class-wise accuracies, as shown in Fig. 2c.

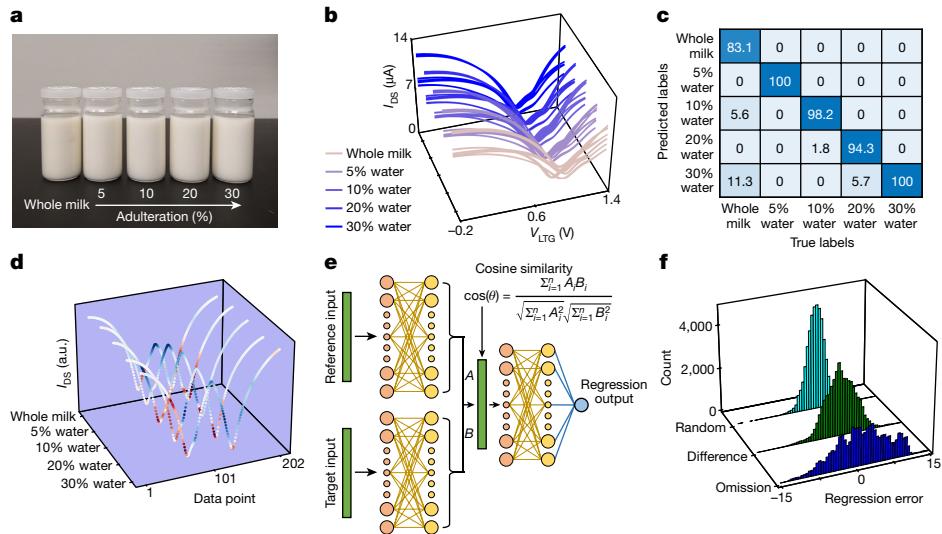
Class-wise SHAP feature analysis (Fig. 2d) was performed to identify the machine-derived regions of importance that facilitated high classification accuracy for the ANN trained directly on ISFET characteristics. Each data point in the input is regarded as a feature and assigned a SHAP value. Positive and negative SHAP values signify a proportional shift in the likelihood of the input being categorized into a particular prediction class. Uncoloured values close to zero do not contribute to the class prediction probability. Note that the SHAP values for each pH class have distinct positive blue regions that denote their identity as well as negative red regions that may discern them from other classes. For example, the Dirac region during the forward sweep is significant across most pH levels (except for pH 6 and 7), highlighting the relevance of  $V_{\text{Dirac},f}$  and  $I_{\text{Dirac},f}$  in detecting pH changes. Extended Data Fig. 5 shows



**Fig. 2 | Understanding pH sensitivity using an ANN.** **a**, ANN architecture consisting of a one-dimensional convolutional neural network (CNN) that acts as a feature extractor followed by three fully connected network layers. The model is designed to predict pH classes ranging from pH 4 to pH 10. **b,c**, Confusion matrices showing the classification accuracy of the model after training with either FOMs (**b**) or the entire graphene ISFET characteristics (**c**). **d**, SHAP features for each pH class with SHAP values assigned to each data point in a representative input. Positive SHAP values shown in blue denote a region of high importance, indicating a higher probability of prediction as the respective input class, whereas negative SHAP values, denoted in red, indicate a decrease in the probability of prediction as that class. Uncoloured values close to zero do not contribute to the class prediction probability.

SHAP features for correctly identified classes, illustrating that consistent regions are activated for each pH level in three representative test samples. This observation underscores that although human-derived FOMs are useful, they are vaguely correlated to SHAP features. Different areas of the transfer characteristics in ISFETs are crucial for differentiating between pH levels, highlighting the effectiveness of ANNs in identifying both inter-class and intra-class features and leveraging their cross-validation to enhance pH sensitivity.

We also explored three training scenarios: in the first two cases, the data from one chip was used as the training set for the other chip, whereas in the third case, all data were collected and split randomly for training and testing. Extended Data Fig. 6 shows the resulting confusion matrices, which notably demonstrate that the model achieves strong performance on the test data regardless of the chip used for training. This finding is important as it suggests that a model trained on ISFET data from one chip can be effectively applied to chips fabricated subsequently, eliminating the need for recalibration or retraining of the model.



**Fig. 3 | From classification to quantification of adulterated milk.** **a**, Samples of whole milk adulterated with different percentages of water not only look alike but also possess similar pH values. **b**, Transfer characteristics of six graphene ISFETs when tested with unadulterated whole milk and with milk at four different levels of adulteration, namely, 5%, 10%, 20% and 30%. **c**, Confusion matrix

showing the per-class accuracy of each milk category. **d**, SHAP feature analysis highlighting the attributes in the graphene ISFET characteristics that enable differentiation of varying percentages of adulteration in whole milk. **e**, Architecture of the ANN-based regression model. **f**, Histograms of error between the predicted level and true level of adulteration.

## Quantification of food adulterants

In this section, we broaden the application scope of machine-learning-assisted graphene ISFETs from pH sensing to the detection and quantification of food adulterants. Economically motivated adulteration of everyday dietary nutrition sources such as milk involves the addition of water or other harmful substances to increase volume or extend shelf life. This adulteration leads to products that are both diluted and nutritionally inferior<sup>40–42</sup>. Conventional detection methods are often inadequate because adulterated products closely resemble their pure counterparts in appearance and taste, as shown in Fig. 3a for whole milk adulterated with different percentages of water. Adulterated products also share chemical similarities as indicated by the corresponding pH values shown in Supplementary Information Section 6. Figure 3b shows the transfer characteristics of six graphene ISFETs when tested with unadulterated whole milk and with milk at four different levels of adulteration, namely, 5%, 10%, 20% and 30%. Note that, in all forthcoming figures, we have not included the cycle-to-cycle variations attributed to drift for clearer visibility of the ISFET characteristics. However, these variations were incorporated while training and testing the ANNs. Figure 3c presents the corresponding confusion matrix, which shows a high classification accuracy. Figure 3d presents SHAP feature analysis highlighting the regions in the graphene ISFET characteristics beyond single-valued FOMs that enable differentiation of varying percentages of adulteration in whole milk. Supplementary Information Section 7 provides further analysis of the features exploited by the ANN to achieve high accuracy.

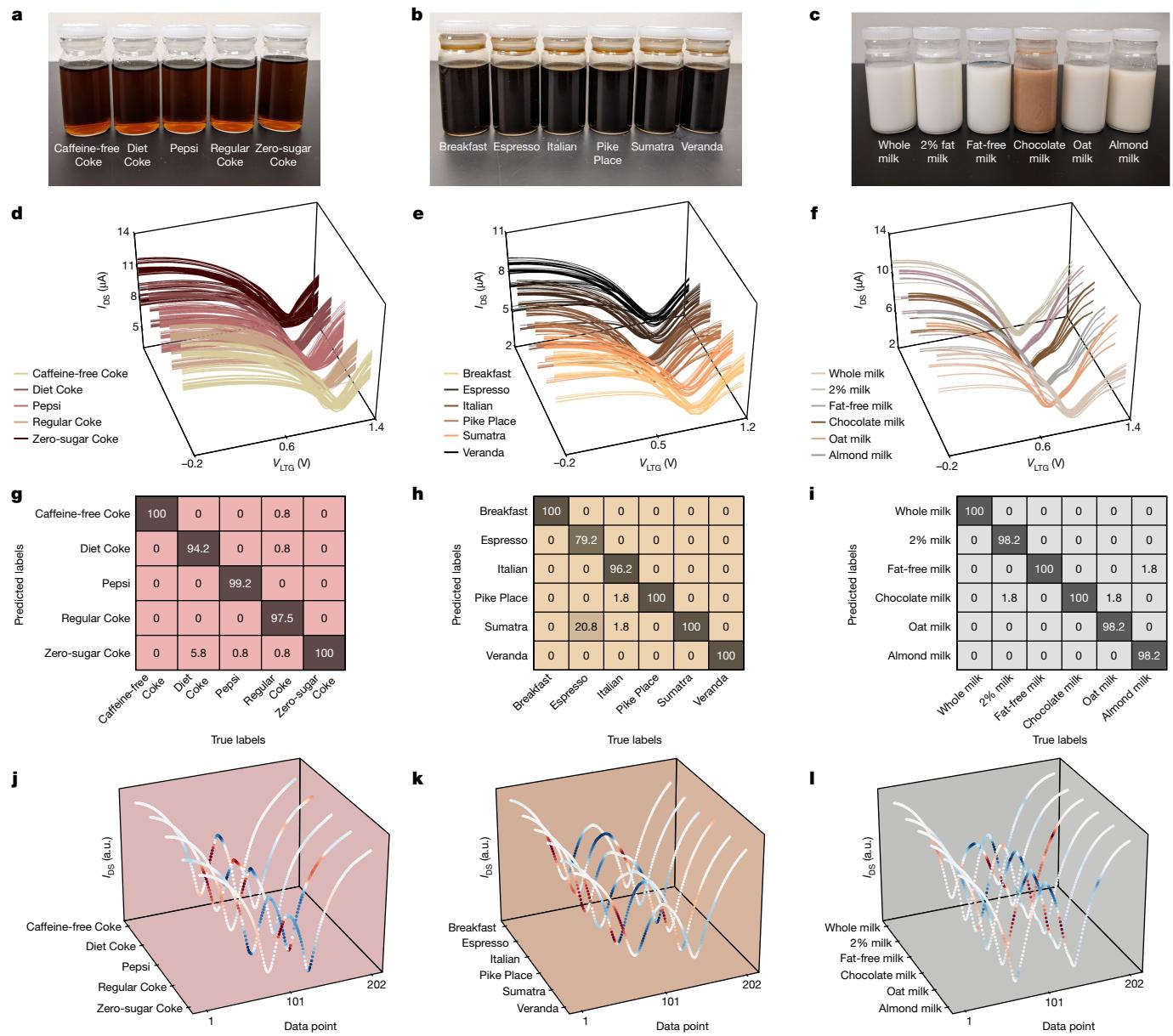
We also extracted FOMs and performed classification with k-NN on the milk adulteration dataset. The FOM rankings and k-NN accuracy are shown in Extended Data Figs. 7 and 8, respectively. Notably, the top three ranked FOMs,  $I_{\max,n}$ ,  $I_{\max/\min,rr}$  and  $I_{\max/\min,rf}$  differ from the top three FOMs obtained during pH classification as the pH values of differently adulterated milk samples are similar. This further emphasizes the versatile parameter space ISFETs offer, enabling sensitive detection across diverse chemical solutions. Contrary to the k-NN approach, ANNs are particularly powerful because the feature identification process is delegated to the model itself. However, precisely quantifying adulteration is a continuous value problem that necessitates the development of a model that can measure the

extent of adulteration, instead of merely labelling it as present in fixed percentages. We used an ANN-based regression model that predicts the adulteration level of a given sample with the help of a reference sample. The adulteration level of the reference sample is known. The model output is the difference between the adulteration level of the unknown sample and the reference sample. We supplemented the existing dataset with samples corresponding to intermediate adulteration levels, namely, 26%, 24%, 17%, 12% and 3%. Following this, all pairs of samples with different adulteration levels were generated, in which one of the samples arbitrarily acts as the reference sample. Identical two-layer fully connected ANNs were used to extract features from each input. Then, the cosine similarity metric was computed between the extracted features and fed as the input to an ANN-based regression model. The entire ANN architecture is shown in Fig. 3e. The total dataset was divided into train and test sets in the following methods: random split, difference-based split and random omission of two adulteration levels. Omitted adulteration levels used for testing were not seen by the model during training. Despite this, the model achieved reasonable accuracy while performing inference on approximately 46,000 samples from the omitted adulteration levels. Histograms depicting error between the predicted and true level of adulteration for all testing cases are shown in Fig. 3f. The root mean squared errors for random split, difference-based split and random omission were found to be about 3%, 3.8% and 5.7%, respectively.

## Food authentication and safety

Another prevalent challenge in the food industry is the increasing occurrence of food fraud incidents that necessitate diligent product authentication practices<sup>43,44</sup>. Authentication becomes especially challenging for products that share similar tastes and appearances, such as various soft drinks (caffeine-free Coke, diet Coke, Pepsi, regular Coke and zero-sugar Coke), coffee blends (Breakfast, Espresso, Italian, Pike Place, Sumatra and Veranda) and types of milk (whole milk, 2% milk, fat-free milk, chocolate milk, oat milk and almond milk) as shown in Fig. 4a–c, respectively. Figure 4d–f shows the transfer characteristics of several graphene ISFETs when tested with different soft drinks, coffee blends and milk varieties, respectively. Figure 4g–i shows the corresponding confusion matrices, which reveal a high classification

# Article

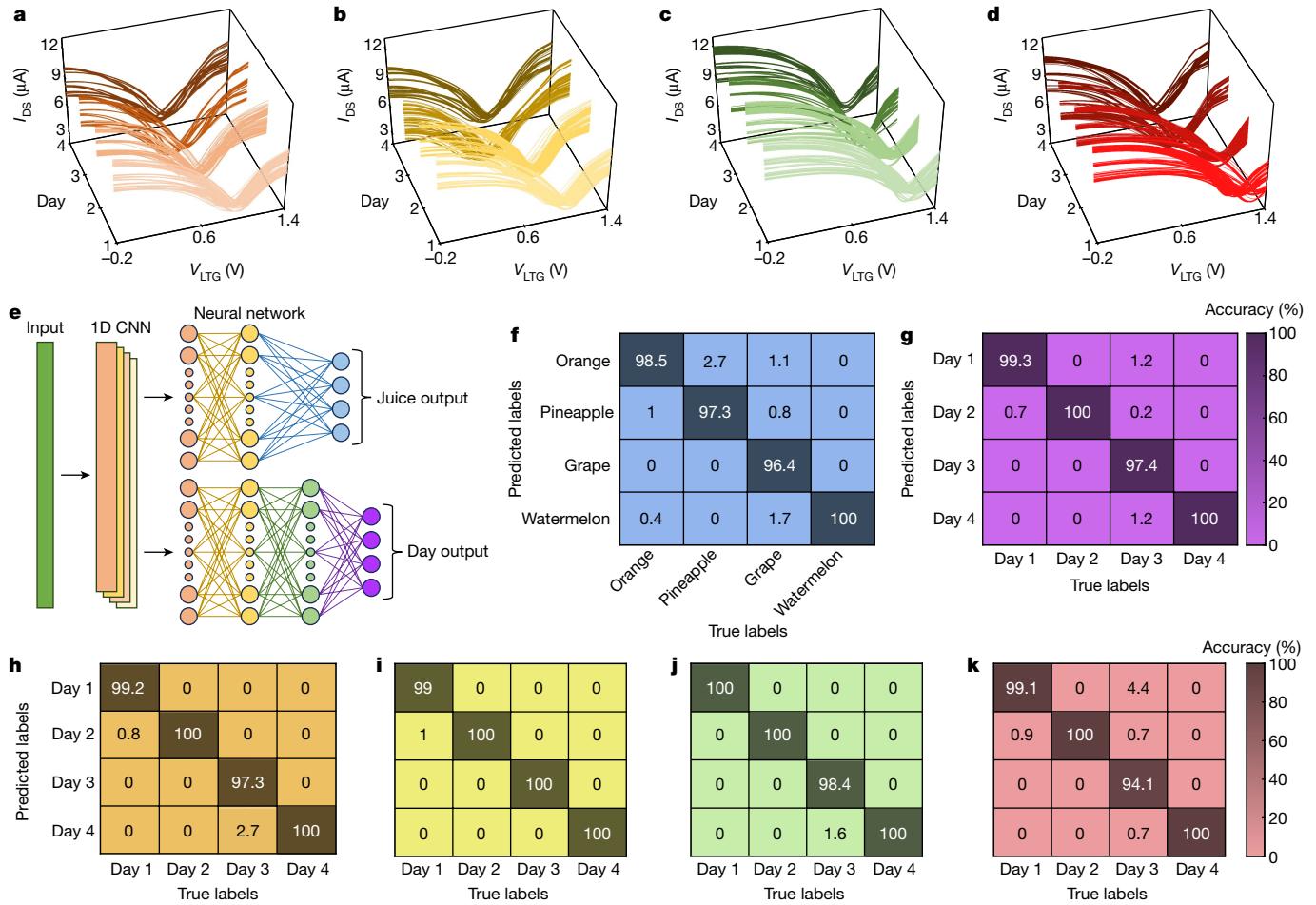


**Fig. 4 | Food authentication with machine-learning-aided graphene ISFETs.** **a–c**, Optical images showing various soft drinks (**a**), coffee blends (**b**) and types of milk (**c**). **d–f**, Transfer characteristics of several graphene ISFETs tested with soft drinks (**d**), coffee blends (**e**) and types of milk (**f**).

**g–i**, Confusion matrices showing per-class accuracy of ANNs trained using graphene ISFET data to authenticate soft drinks (**g**), coffee blends (**h**) and types of milk (**i**). **j–l**, SHAP feature analysis for soft drinks (**j**), coffee blends (**k**) and types of milk (**l**).

accuracy for each item in its respective product category. This implies that ANNs, when trained on data from graphene ISFETs, are highly capable of accurately authenticating food items. Figure 4j–l and Supplementary Information Sections 8–10 present the SHAP values for each item in the product classes. FOMs derived from graphene ISFET characteristics for the different soft drinks, coffee blends and milk varieties were used as inputs to the k-NN algorithm. The resultant FOM rankings and k-NN accuracies are shown in Extended Data Figs. 7 and 8, respectively. The combination of top FOMs differs for each product category, further emphasizing the significance of using the expanded parameter space provided by ISFETs. Extended Data Fig. 9a–d shows the authentication of three types of wine (red, white and sparkling) and two types of water (spring and sparkling), transfer characteristics of a set of six ISFETs measured using these solutions, a confusion matrix demonstrating near-perfect classification accuracy and associated SHAP feature analysis.

Next, ensuring food safety is a critical issue, with the timely detection of harmful contaminants in food production and distribution posing a continuous challenge. For instance, more than 1,400 potentially hazardous per- and polyfluoroalkyl substances (PFAS) are pervasive in industrial processes and found in consumer products<sup>45–47</sup>. PFAS has the potential to accumulate in the environment, including in drinking water, highlighting the need for effective PFAS detection. We demonstrate the detection of perfluorohexanoic acid (PFHxA) in water at concentrations down to 2.5 ppb. Extended Data Fig. 9e–h shows solutions of deionized water and PFHxA solutions at four concentrations (2,500 ppb, 250 ppb, 25 ppb and 2.5 ppb), transfer characteristics of a set of four ISFETs measured using these solutions, a confusion matrix depicting perfect classification accuracy and associated SHAP feature analysis. These results highlight that machine-learning-aided graphene ISFETs can be applied to tackle a broad spectrum of challenges in the food industry.



**Fig. 5 | Monitoring food freshness using machine-learning-assisted graphene ISFETs.** **a–d**, Transfer characteristics of 14 graphene ISFETs, each measured over 4 consecutive days using orange (**a**), pineapple (**b**), grape (**c**) and watermelon (**d**) juices. **e**, ANN model consisting of a feature extraction layer followed by two classifiers that simultaneously predict juice identity and age.

**f,g**, Confusion matrices demonstrating per-class accuracy for the classification of fruit identity across all ages (**f**) and the classification of age across all fruit identities (**g**). **h–k**, Confusion matrices showing per-class accuracy for the age classification of orange (**h**), pineapple (**i**), grape (**j**) and watermelon (**k**).

## Food freshness monitoring

Although corrective actions exist for food adulteration and contamination incidents, monitoring food freshness is more challenging because of time-varying and complex chemical compositions present in food. Spoilt food is dangerous to consume and possesses a diminished nutritional value.

Here we show that machine-learning-aided ISFETs can be deployed to assess the freshness of consumer food products. We use juices extracted from four different fruits—orange, pineapple, grape and watermelon—and test them over 4 consecutive days as they gradually perish. The simultaneous detection of fruit juice identity and age is a non-trivial problem as 4 fruit juices measured over 4 days form 16 closely related classes. Figure 5a–d shows the transfer characteristics of 14 ISFETs, each measured over 4 consecutive days using orange, pineapple, grape and watermelon juice, respectively. Note that two separate unknowns, that is, fruit juice identity and age, exist in the dataset presented to the ANN model. The number of output nodes required to express all combinations of these two solution domains in an ANN would scale rapidly as a product of the number of juice classes and the freshness timesteps that are considered. By contrast, our approach simplifies the problem by using a multi-output ANN to efficiently extract features from the input and perform the fruit identity and age classification tasks simultaneously.

The model, shown schematically in Fig. 5e, consists of a shared one-dimensional convolutional neural network that acts as a feature extractor followed by separate fully connected layers for each task, that is, classification of fruit identity and age. The fruit identity classifier uses a three-layer densely connected ANN, and the age classifier is a two-layer densely connected ANN. After training, the final accuracy of this network for the task of fruit identity–age classification reached a value of 97.67%. Figure 5f,g shows the confusion matrices demonstrating the per-class accuracy after training for fruit identity excluding age information and fruit age excluding identity information, respectively. The accuracy values were found to be about 98% and 99%, respectively, for classification of fruit identity and age. Moreover, the per-class accuracy of the age class for each fruit identity—that is, orange, pineapple, grape and watermelon—can be found in the confusion matrices shown in Fig. 5h–k, respectively. Note that the multithread classification problem involving the determination of fruit juice identity and age represents a rather broad and complex class of chemosensing problems. This is because each fruit juice is composed of many chemicals such as water, carbohydrates, proteins, lipids, acid, vitamins and minerals and their decomposition can be driven by many causes such as oxidation, bacteria and fungi. Yet, the machine learning algorithm can classify the juices and their freshness based on sensor data collected by a graphene ISFET array without any previous knowledge of either the composition or the ageing mechanism. The advantage of using an ANN is that it allows for the classification of various classes and sub-classes while

compensating for sensor non-idealities. Extended Data Table 1 summarizes the resulting accuracies of classification with k-NN and an ANN trained with either human-derived FOMs or the entire ISFET transfer characteristics across several different training scenarios for all applications and Supplementary Information Section 11 provides a tabular description of all datasets.

## Conclusion

In conclusion, this study highlights the transformative potential of integrating machine learning algorithms with graphene ISFETs in chemosensing applications. ANNs trained with data generated by non-functionalized graphene ISFET arrays have demonstrated remarkable abilities in classifying and quantifying chemical changes, especially in the context of food authenticity, adulteration, safety and spoilage. Furthermore, this study also provides insights on the interpretability of ANNs, traditionally considered as opaque black boxes. By using machine learning algorithms, this research overcomes traditional challenges observed in ISFETs, such as sensor variability and the need for extensive calibration, thereby enhancing the reliability and applicability of the technology.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-08003-w>.

1. Bergveld, P. Development of an ion-sensitive solid-state device for neurophysiological measurements. *IEEE Trans. Biomed. Eng.* **BME-17**, 70–71 (1970).
2. Bergveld, P. Thirty years of ISFETOLOGY: what happened in the past 30 years and what may happen in the next 30 years. *Sens. Actuators B Chem.* **88**, 1–20 (2003).
3. Bergveld, P. Development, operation, and application of the ion-sensitive field-effect transistor as a tool for electrophysiology. *IEEE Trans. Biomed. Eng.* **19**, 342–351 (1972).
4. Fu, W., Jiang, L., van Geest, E. P., Lima, L. M. C. & Schneider, G. F. Sensing at the surface of graphene field-effect transistors. *Adv. Mater.* **29**, 1603610 (2017).
5. Saba, G. K. et al. The development and validation of a profiling glider deep ISFET-based pH sensor for high resolution observations of coastal and ocean acidification. *Front. Mar. Sci.* **6**, 664 (2019).
6. Margarit-Taulé, J. M., Martín-Ezquerro, M., Escudé-Pujol, R., Jiménez-Jorquera, C. & Liu, S.-C. Cross-compensation of FET sensor drift and matrix effects in the industrial continuous monitoring of ion concentrations. *Sens. Actuators B Chem.* **353**, 131123 (2022).
7. Gao, W. et al. Fully integrated wearable sensor arrays for multiplexed *in situ* perspiration analysis. *Nature* **529**, 509–514 (2016).
8. Weston, M., Geng, S. & Chandrawati, R. Food sensors: challenges and opportunities. *Adv. Mater. Technol.* **6**, 2001242 (2021).
9. Xue, M. et al. Integrated biosensor platform based on graphene transistor arrays for real-time high-accuracy ion sensing. *Nat. Commun.* **13**, 5064 (2022).
10. Fakih, I. et al. Selective ion sensing with high resolution large area graphene field effect transistor arrays. *Nat. Commun.* **11**, 3226 (2020).
11. Iijima, S. Helical microtubules of graphitic carbon. *Nature* **354**, 56–58 (1991).
12. Treuting, R. G. & Arnold, S. M. Orientation habits of metal whiskers. *Acta Metall.* **5**, 598 (1957).
13. Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
14. Heller, I. et al. Influence of electrolyte composition on liquid-gated carbon nanotube and graphene transistors. *J. Am. Chem. Soc.* **132**, 17149–17156 (2010).
15. Ang, P. K., Chen, W., Wee, A. T. S. & Loh, K. P. Solution-gated epitaxial graphene as pH sensor. *J. Am. Chem. Soc.* **130**, 14392–14393 (2008).
16. Knopfmacher, O. et al. Nernst Limit in Dual-Gated Si-Nanowire FET Sensors. *Nano Lett.* **10**, 2268–2274 (2010).
17. Wang, K. et al. Carbon nanotube field-effect transistor based pH sensors. *Carbon* **205**, 540–545 (2023).
18. Fu, W. et al. High mobility graphene ion-sensitive field-effect transistors by noncovalent functionalization. *Nanoscale* **5**, 12104–12110 (2013).
19. Shang, X., Park, C. H., Jung, G. Y., Kwak, S. K. & Oh, J. H. Highly enantioselective graphene-based chemical sensors prepared by chiral noncovalent functionalization. *ACS Appl. Mater. Interfaces* **10**, 36194–36201 (2018).
20. Li, X. et al. Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324**, 1312–1314 (2009).
21. Ohno, Y., Maehashi, K., Yamashiro, Y. & Matsumoto, K. Electrolyte-gated graphene field-effect transistors for detecting pH and protein adsorption. *Nano Lett.* **9**, 3318–3322 (2009).
22. Lei, N., Li, P., Xue, W. & Xu, J. Simple graphene chemiresistors as pH sensors: fabrication and characterization. *Meas. Sci. Technol.* **22**, 107002 (2011).
23. Lee, M. H. et al. Apparent pH sensitivity of solution-gated graphene transistors. *Nanoscale* **7**, 7540–7544 (2015).
24. Jung, S.-H. et al. Super-Nernstian pH sensor based on anomalous charge transfer doping of defect-engineered graphene. *Nano Lett.* **21**, 34–42 (2021).
25. Mailly-Giacchetti, B. et al. pH sensing properties of graphene solution-gated field-effect transistors. *J. Appl. Phys.* **114**, 084505 (2013).
26. Gao, J. et al. Graphene-based field-effect transistors integrated with microfluidic chip for real-time pH monitoring of seawater. *J. Mater. Sci., Mater. Electron.* **31**, 15372–15380 (2020).
27. Helmholz, H. Studien über electrische Grenzschichten. *Ann. Phys. Chem.* **243**, 337–382 (1879).
28. Gouy, M. Sur la constitution de la charge électrique à la surface d'un électrolyte. *J. Phys. Theor. Appl.* **9**, 457–468 (1910).
29. Chapman, D. L. LI. A contribution to the theory of electrocapillarity. *Lond. Edinb. Dubl. Philos. Mag. J. Sci.* **25**, 475–481 (1913).
30. Wang, Z. L. & Wang, A. C. On the origin of contact-electrification. *Mater. Today* **30**, 34–51 (2019).
31. Wu, J. Understanding the electric double-layer structure, capacitance, and charging dynamics. *Chem. Rev.* **122**, 10821–10859 (2022).
32. Salvo, P. et al. Graphene-based devices for measuring pH. *Sens. Actuators B Chem.* **256**, 976–991 (2018).
33. Fu, W. et al. Graphene transistors are insensitive to pH changes in solution. *Nano Lett.* **11**, 3597–3600 (2011).
34. LeBow, N. et al. Real-time edge neuromorphic tasting from chemical microsensor arrays. *Front. Neurosci.* **15**, 771480 (2021).
35. Chang, K.-M., Chang, C.-T., Chao, K.-Y. & Lin, C.-H. A novel pH-dependent drift improvement method for zirconium dioxide gated pH-ion sensitive field effect transistors. *Sensors* **10**, 4643–4654 (2010).
36. Sinha, S. et al. Temperature and temporal drift compensation for Al<sub>2</sub>O<sub>3</sub>-gate ISFET-based pH sensor using machine learning techniques. *Microelectron. J.* **97**, 104710 (2020).
37. Larose, D. T. & Larose, C. D. in *Discovering Knowledge in Data: An Introduction to Data Mining* 149–164 (Wiley, 2014).
38. Hornik, K., Stinchcombe, M. & White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366 (1989).
39. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *In Proc. 31st International Conference on Neural Information Processing Systems Vol. 30*, 4768–4777 (ACM, 2017).
40. Chauhan, S. L., Priyanka, Mandal, K. D., Paul, B. R. & Maji, C. Adulteration of milk: a review. *Int. J. Chem. Stud.* **7**, 2055–2057 (2019).
41. Techane, T. Effect of adulterants on quality and safety of cow milk: a review. *Int. J. Diabates Metab. Disord.* **8**, 277–287 (2023).
42. Das, S., Goswami, B. & Biswas, K. Milk adulteration and detection: a review. **14**, 4–18 (2016).
43. Danezis, G. P., Tsagkaris, A. S., Camin, F., Brusic, V. & Georgiou, C. A. Food authentication: techniques, trends & emerging approaches. *Trends Anal. Chem.* **85**, 123–132 (2016).
44. Aung, M. M. & Chang, Y. S. Traceability in a food supply chain: safety and quality perspectives. *Food Control* **39**, 172–184 (2014).
45. Wang, Z., DeWitt, J. C., Higgins, C. P. & Cousins, I. T. A never-ending story of per- and polyfluoroalkyl substances (PFASs)? *Environ. Sci. Technol.* **51**, 2508–2518 (2017).
46. Glüge, J. et al. An overview of the uses of per- and polyfluoroalkyl substances (PFAS). *Environ. Sci. Process. Impacts* **22**, 2345–2373 (2020).
47. Cousins, I. T. et al. The concept of essential use for determining when uses of PFASs can be phased out. *Environ. Sci. Process. Impacts* **21**, 1803–1815 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

## Methods

### Dataset descriptions

**pH dataset.** Standard pH buffer solutions were commercially acquired from Sigma Aldrich. The solutions are accurate to within  $\pm 0.02$  pH for solutions of pH 4–9. The standard reference solution for pH 10 was reported to be accurate to within  $\pm 0.05$  pH. The dataset comprises measurements collected from seven ISFET devices across two separate chips. Chip 1 contains three ISFET devices whereas chip 2 contains four ISFET devices. ISFETs on each chip were measured sequentially. The first ISFET on a given chip was measured 50 times per solution (except for pH 4 and pH 5, which used 25 iterations for the first ISFET), whereas the following ISFETs were measured 25 times for all solutions. This dataset contains 1,475 total measurement cycles (and 297,950 total  $I_{DS}$  datapoints), that is, 225 measurement cycles for pH 6, 7, 8, 9 and 10 and 175 measurement cycles for pH 4 and 5 with 202  $I_{DS}$  datapoints per measurement cycle.

For data collection here, and in all other datasets, we define measurement as a single application of the forward and reverse voltage sweeps between  $-200$  mV and  $1,200$  mV or  $1,400$  mV that will provide transfer characteristics. This measurement consists of exactly 202  $I_{DS}$  datapoints, in which the first and second sets of 101 points correspond to the forward and reverse sweeps, respectively.

**Milk quantification dataset.** For the task of milk adulterant quantification, a dataset was created containing ISFET measurements of unadulterated commercially available whole milk and the same whole milk at four different levels of adulteration, that is, 5%, 10%, 20% and 30% of total solution volume was replaced with added water. For example, 100 ml of a 5% adulterated solution would contain 95 ml of whole milk with 5 ml of added water. Measurements were collected from six ISFETs from one chip. Each class had 175 samples consisting of 875 total measurements. Moreover, to aid in the training of the ANN regression model, a set of six additional adulteration levels were measured, that is, 0%, 3%, 12%, 17%, 24% and 26%. Measurements were collected from six ISFETs on one chip. Each added class had 175 samples consisting of 1,050 total measurements.

**Soda dataset.** For the task of soda authentication, a dataset was created containing ISFET measurements of five commercially available soda product classes, including Zero-sugar Coke, Regular Coke, Pepsi, Diet Coke and Caffeine-free Coke. Measurements were collected from 12 ISFETs across two separate chips. Each class had 400 samples with a total of 2,000 measurements.

**Coffee dataset.** For the task of coffee authentication, a dataset was created containing ISFET measurements of six classes, including coffees brewed from commercially available Starbucks brand Veranda, Sumatra, Pike Place, Italian, Espresso and Breakfast blends. Measurements were collected from six ISFETs on one chip. Each class had 175 samples with a total of 1,050 measurements.

**Milk authentication dataset.** For the task of milk authentication, a dataset was created containing ISFET measurements of six classes, including commercially available whole milk, 2% milk, fat-free milk, chocolate milk, oat milk and almond milk. Measurements were collected from six ISFETs on one chip. Each class had 175 samples with a total of 1,050 measurements.

**Perfluorohexanoic acid dataset.** For the task of perfluorohexanoic acid (PFHxA) detection, a dataset was created containing ISFET measurements of five classes, including deionized water and PFHxA at concentrations of 2.5 ppb, 25 ppb, 250 ppb and 2,500 ppb. Measurements were collected from four ISFETs on one chip. Each class had 125 samples with a total of 625 measurements.

**Juice dataset.** For the task of juice identity and age classification, a dataset was created containing ISFET measurements of four classes, including grape, pineapple, orange and watermelon across 4 consecutive days beginning on the day the fruits were purchased and juice was squeezed. Measurements were collected from 14 ISFETs across two separate chips. Watermelon had 350 measurements on day 2. Grape and watermelon had 375 measurements on day 4. All other fruit and day combinations contained 400 measurements adding to a total of 6,300 measurements.

**Wine and water dataset.** For the task of wine and water classification, a dataset was created containing ISFET measurements of five classes, including commercially acquired red wine (Clos du Bois Merlot), white wine (Clos du Bois Chardonnay), sparkling wine (Berlucci), Club Soda (Seagram's) and spring water (Deer Park). Measurements were collected across six ISFETs on one chip. Each class had 175 samples with a total of 875 measurements.

### Model descriptions

**k-NN classification.** All scripts that perform k-NN analyses were written and evaluated in MATLAB. Standard MATLAB functions were used to create a script that executes the k-NN algorithm. All k-NN algorithms use an L1 distance calculation to determine the distance between a target point and its neighbours in space. For pH, milk quantification, soda, coffee, milk classification, PFAS, wine and juice tasks, k values of 30, 25, 35, 30, 30, 20, 30 and 65, respectively, were used when splitting the data randomly. For training with a random split, 70% of samples from each class were extracted to serve as a training set, whereas 30% of samples were reserved for testing. All FOMs were independently normalized using the z-transform method. Training and testing sets were normalized separately.

**ANN classification.** All ANN models were created in Python using the PyTorch library. SHAP analysis was performed using the SHAP Library<sup>39</sup>. When dividing the data into train and test sets, 70% of samples from each class were extracted, and then combined to form the training set, the rest 30% of samples from each class were used as the unseen test set. The exact number of cycles, sensors and chips in each dataset can be found in Supplementary Information Section 11. For preprocessing, z-transform was applied to train and test sets independently. The ANN architecture in which the entire ISFET transfer curve was taken as an input included a one-dimensional convolutional layer (CONVID) with four filters each having a dimension of 2 and a stride of 1. Following this, the output of the CONVID layer was flattened and fed into a three-layer neural network with ReLU as the activation in between the layers and sigmoid as the activation at the output layer. The input layer of the neural network had 512 nodes, the hidden layer had 128 nodes and the output layer had the number of nodes corresponding to the number of classes. The ANN architecture was trained using the Adam optimizer with a learning rate of  $1 \times 10^{-3}$ , weight decay of  $1 \times 10^{-5}$  and the batch size was fixed at 64. The loss function used was mean squared error. These hyperparameters, model architecture and the train–test split scheme were kept identical across all classification tasks. pH, soda, coffee, milk classification, wine, PFAS and juice ANNs used 250, 250, 800, 800, 200, 500 and 800 epochs, respectively, for training. The model was evaluated on the unseen test set with parameters learned at the end of the training epoch for all tasks.

For ANN experiments with FOMs as the input, the model architecture was kept the same, except that the number of nodes in the first fully connected layer following the CONVID layer was 64 nodes. The batch size was fixed at 64 and the model was trained for a total of 3,000 epochs on all datasets with Adam as the optimizer and mean squared error as the loss function. The learning rate for the optimizer was set to  $1 \times 10^{-4}$  with a weight decay of  $1 \times 10^{-5}$ . The inference on the unseen test set was carried out with the trained model after completion of

# Article

the training epochs. These hyperparameters were kept consistent across all datasets.

For the task of multihead classification with juice and age data from FOMs, the ANN model architecture was designed to have a CONVID layer for feature extraction followed by two fully connected networks for classification. The juice classifier model consisted of three fully connected layers with 64 nodes in the input layer, 128 nodes in the hidden layer and 4 nodes in the output layer followed by a sigmoid activation. Similarly, for the age classifier model, two fully connected layers were used with 128 nodes in the input layer and 4 nodes in the output layer followed by a sigmoid activation. The intermediate layer activation was ReLU. This model was trained for 800 epochs using Adam as the optimizer and mean squared error as the loss function. The batch size was set to 64. The learning rate for the optimizer was  $1 \times 10^{-4}$  with a weight decay of  $1 \times 10^{-5}$ .

**ANN quantification.** In the difference of adulteration-based split, the model takes as input a reference sample for which the adulteration level is known and an unknown sample for which the model predicts the adulteration level difference. After inference, we compute the adulteration of the unknown sample by simply adding the predicted difference and the adulteration level of the reference sample. We construct pairs of samples from the existing milk quantification dataset and collect 300,000 total pairs from the distribution of all possible differences. This was split into a 200,000 pair training set and a 100,000 pair test set. Notably, there were about 30 unique adulteration difference values. For preprocessing, the z-transform was applied to train and test sets independently. The ANN-based model included two identical sub-networks each having two fully connected layers for feature extraction from each sample in the input pair. The number of nodes in the hidden layer was 512, followed by the second layer with an output of size 1,024. Following this, the extracted features from both networks were used to compute the cosine similarity metric and this was given as an input to the regressor model. The regressor was designed with three fully connected layers with tanh as the activation for the output. The number of nodes in the first layer was 512, followed by 128 nodes in the hidden layer and one node in the output layer. The interlayer activation for all sub-networks was ReLU. The model was trained for 50 epochs with a batch size of 256. The Adam optimizer was used with a learning rate of  $1 \times 10^{-4}$  and weight decay of  $1 \times 10^{-5}$ . The inference was performed with the trained model obtained at the end of the training epochs.

**Ranking of FOMs using k-NN.** The accuracy of a k-NN algorithm was evaluated for all possible single and multivariate combinations of the 20 input FOMs. Accordingly,  $\binom{20}{n}$  total accuracies were recorded in the range of  $n = 1$  to  $n = 20$ , in which  $n$  is the number of input FOMs considered. The top 1,000 accuracies were considered at each  $n$  except for  $n = 1, 2, 18, 19$  and  $20$  in which all 20, 190, 190, 20 and 1 combination(s) were considered. For any given  $n$ , the combinations corresponding to

the top 1,000 accuracies were deconstructed into their constituent FOMs and averaged for each FOM. The composite accuracy for each FOM was averaged across  $n = 1$  to  $n = 20$ . The rank for each FOM was determined based on its composite accuracy. These rankings can be found in Extended Data Fig. 7.

## Electrical characterization

A Keysight B1500A parameter analyser was used for all electrical measurements. This parameter analyser has a measurement resolution of 0.1 fA. Moreover, noise added by the packaged printed circuit board readout module and SMU cables has been measured and found to be in the range of 1 pA. This is several orders of magnitude below the minimum recorded ISFET currents. Therefore, we believe that any added noise or error from this tool or measurement setup is negligible.

## Computational details

All scripts that perform k-NN analyses were written and evaluated in MATLAB. Standard MATLAB functions were used to create a script that executes the k-NN algorithm while the curve-fitting toolbox was used for regression. In MATLAB, all computations were carried out using double-precision floating point format numbers. At this level of precision, we can assume that the error introduced by computation during k-NN or regression is negligible. All ANN models were created in Python using the PyTorch library. SHAP analysis was performed using the SHAP Library<sup>39</sup>.

## Data availability

The datasets generated during and/or analysed during the current study are freely available at GitHub (<https://github.com/Andrew-Pannone/Chemosensing>).

## Code availability

The codes used to analyse the data are freely available at GitHub (<https://github.com/Andrew-Pannone/Chemosensing>).

**Acknowledgements** This work was supported by a NASA Space Technology Graduate Research Opportunity (NSTGRO) grant.

**Author contributions** Saptarshi Das, A.P., H.R. and Sarbassis Das conceived the idea and designed the experiments. A.P., H.R., Sarbassis Das, Z.C. and C.A.P. performed the experiments. A.P. and A.R. developed the machine learning models. All authors contributed to the preparation of the paper.

**Competing interests** The authors declare no competing interests

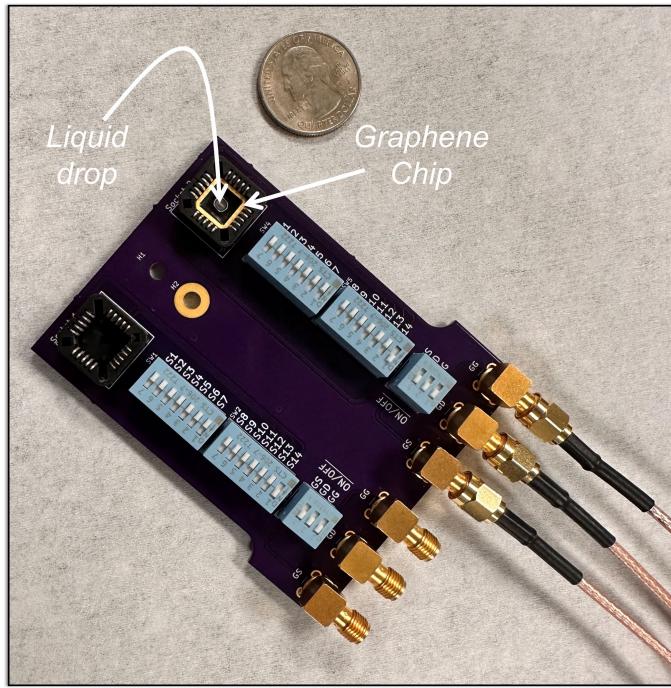
## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-08003-w>.

**Correspondence and requests for materials** should be addressed to Saptarshi Das.

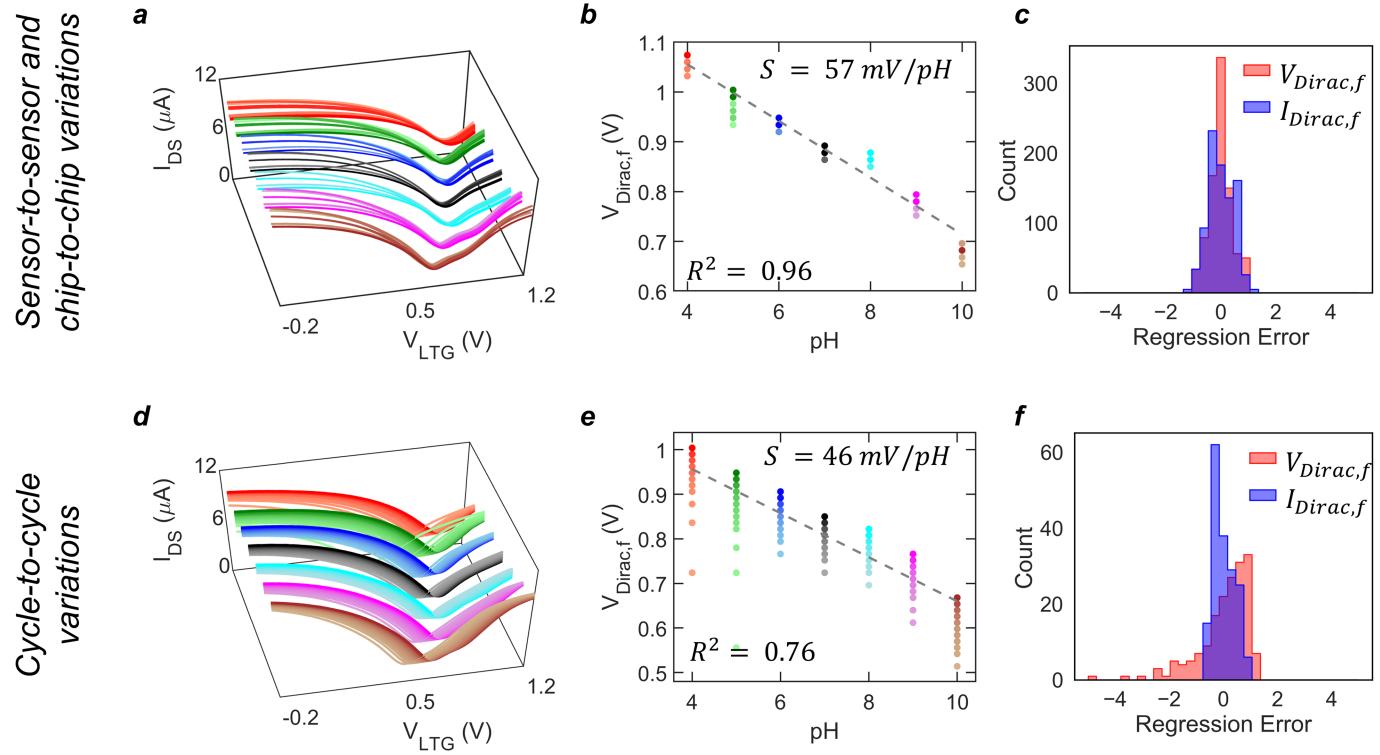
**Peer review information** *Nature* thanks Marcos Teixeira and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig. 1 | Graphene ISFET chip mounted on a printed circuit board (PCB).** Optical image of a chip containing an array of graphene-based ISFETs mounted on a PCB.

# Article

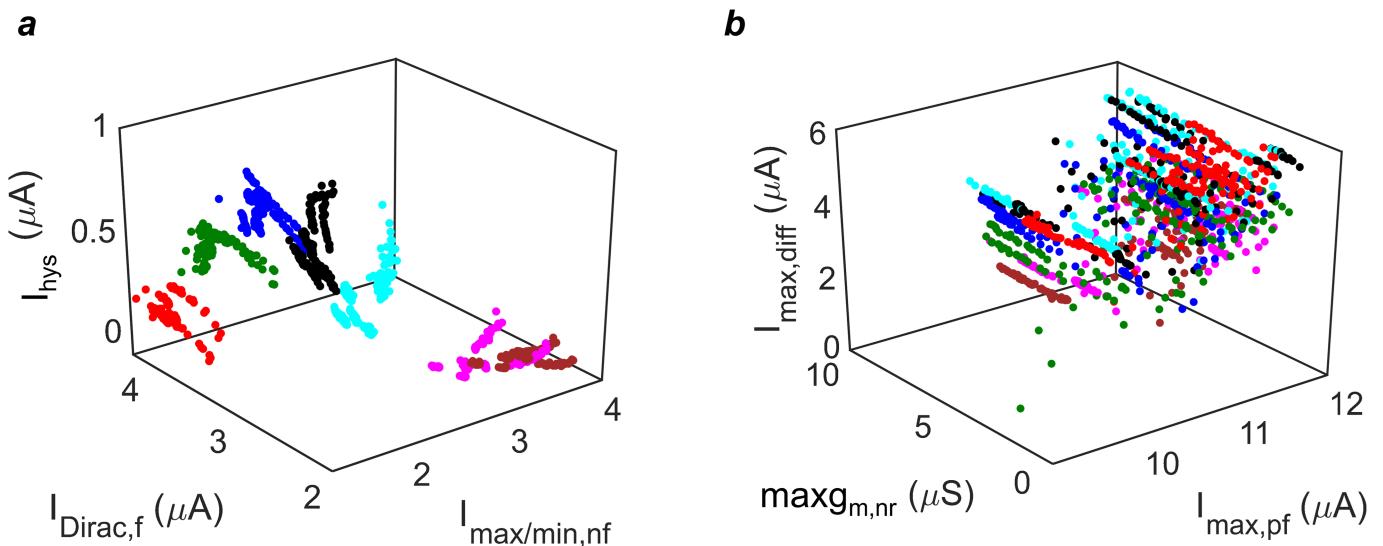


**Extended Data Fig. 2 | Impact of ISFET non-idealities on pH sensitivity.**  
**a)** Transfer characteristics for a set of five graphene ISFETs across two separate chips. **b)** pH sensitivity of  $V_{Dirac,f}$  and corresponding **c)** histograms of the error for  $V_{Dirac,f}$  and  $I_{Dirac,f}$  considering sensor-to-sensor and chip-to-chip variations.

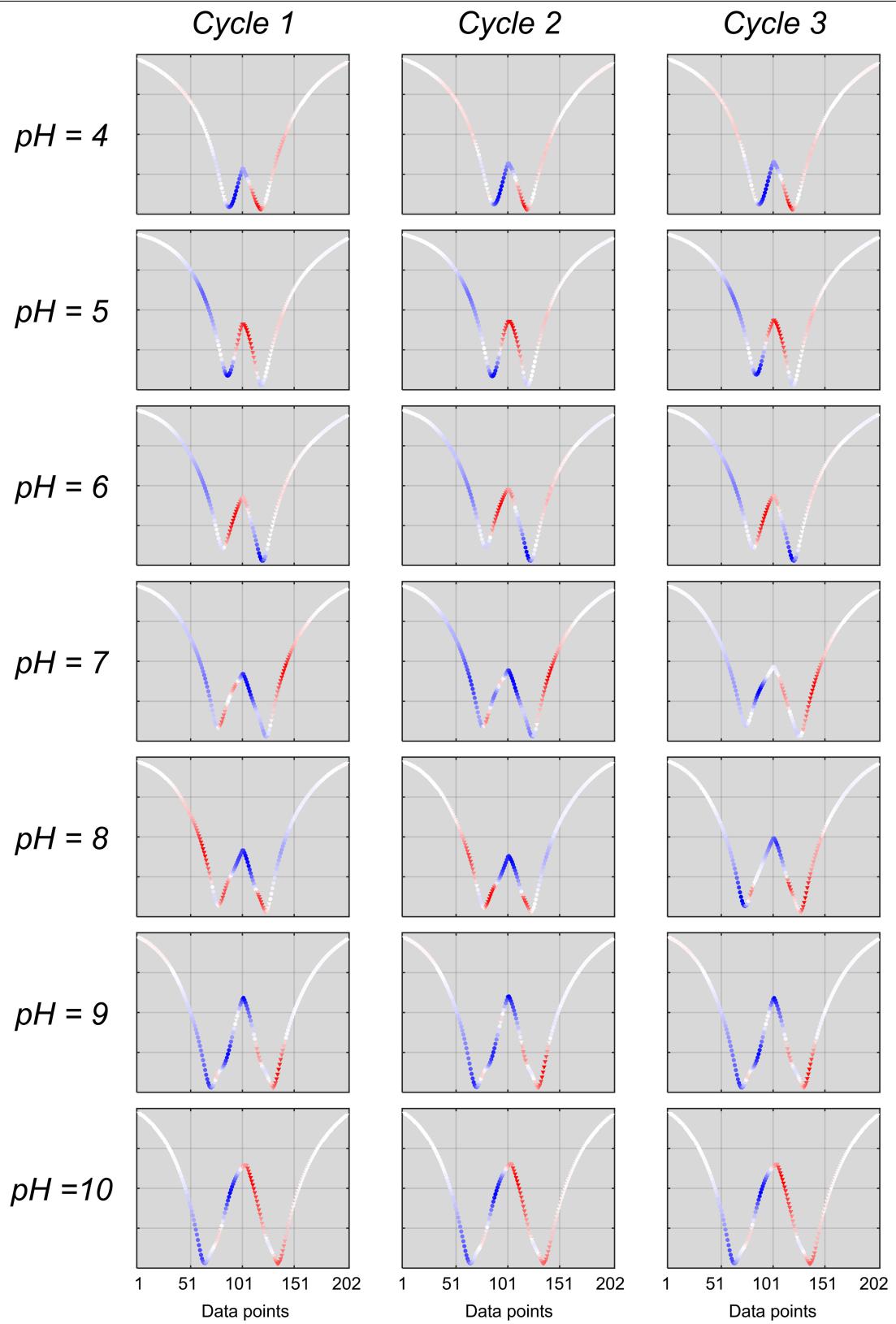
**d)** Transfer characteristics of a graphene ISFET measured for  $N = 25$  cycles depicting the cycle-to-cycle variations. **e)** pH sensitivity of  $V_{Dirac,f}$  and corresponding **f)** histograms of the error for  $V_{Dirac,f}$  and  $I_{Dirac,f}$  accounting for cycle-to-cycle variation.

FOM	$R^2$	Regression Rank	k-NN Score	k-NN Rank
$V_{Dirac,f}$	0.791	7	0.972	5
$V_{Dirac,r}$	0.788	8	0.963	10
$I_{Dirac,f}$	0.919	2	0.981	1
$I_{Dirac,r}$	0.383	15	0.968	8
$V_{hys}$	0.512	13	0.960	13
$I_{hys}$	0.898	4	0.976	2
$I_{max,n}$	0.646	9	0.956	14
$I_{max,pf}$	0.165	18	0.942	18
$I_{max,pr}$	0.322	16	0.948	17
$I_{max/min,nf}$	0.899	3	0.975	3
$I_{max/min,rr}$	0.844	6	0.967	9
$I_{max/min,pf}$	0.927	1	0.975	4
$I_{max/min,pr}$	0.533	12	0.971	6
$maxg_{m,nf}$	0.247	17	0.961	12
$maxg_{m,rr}$	0.061	19	0.463	20
$maxg_{m,pf}$	0.877	5	0.969	7
$maxg_{m,pr}$	0.533	11	0.961	11
$maxg'_{m,r}$	0.018	20	0.954	16
$maxg'_{m,r}$	0.561	10	0.955	15
$I_{max,diff}$	0.399	14	0.898	19

**Extended Data Fig. 3 | FOM rankings for pH classification.** FOMs ranked based on their coefficient of determination ( $R^2$ ) and k-NN score.



**Extended Data Fig. 4 | Three-dimensional scatter plots of highest and lowest ranked FOMs.** Three-dimensional scatter plots that utilize combinations of **a)** the highest ranked FOMs, that is,  $I_{\text{Dirac,f}}$ ,  $I_{\text{hys}}$ , and  $I_{\text{max/min,nf}}$  and **b)** the lowest ranked FOMs, that is,  $I_{\text{max,pf}}$ ,  $I_{\text{max,diff}}$ , and  $\text{max}g_{m,\text{nr}}$ .



**Extended Data Fig. 5 | SHAP features of three representative cycles for all pH classes.** SHAP features of three representative cycles for all pH classes showing that consistent regions in the ISFET characteristics are activated for each pH level.

# Article

		Train Chip1					Train Chip2					Random split: Chip1+Chip2				
		Test Chip2					Test Chip1									
Predicted Labels	4	98	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	95	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	94.4	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	5.6	100	2.4	0	0	0	0	0	0	0	0	0	0
	8	2	0	0	0	0	93.6	0	0	0	0	0	0	0	0	0
	9	0	3	0	0	4	92.8	0	0	0	0	0	0	0	0	0
	10	0	1	0	0	0	0	7.2	100	0	0	0	0	0	0	0
		4	5	6	7	8	9	10		4	5	6	7	8	9	10
		True Labels														

## Extended Data Fig. 6 | Confusion matrices depicting pH classification

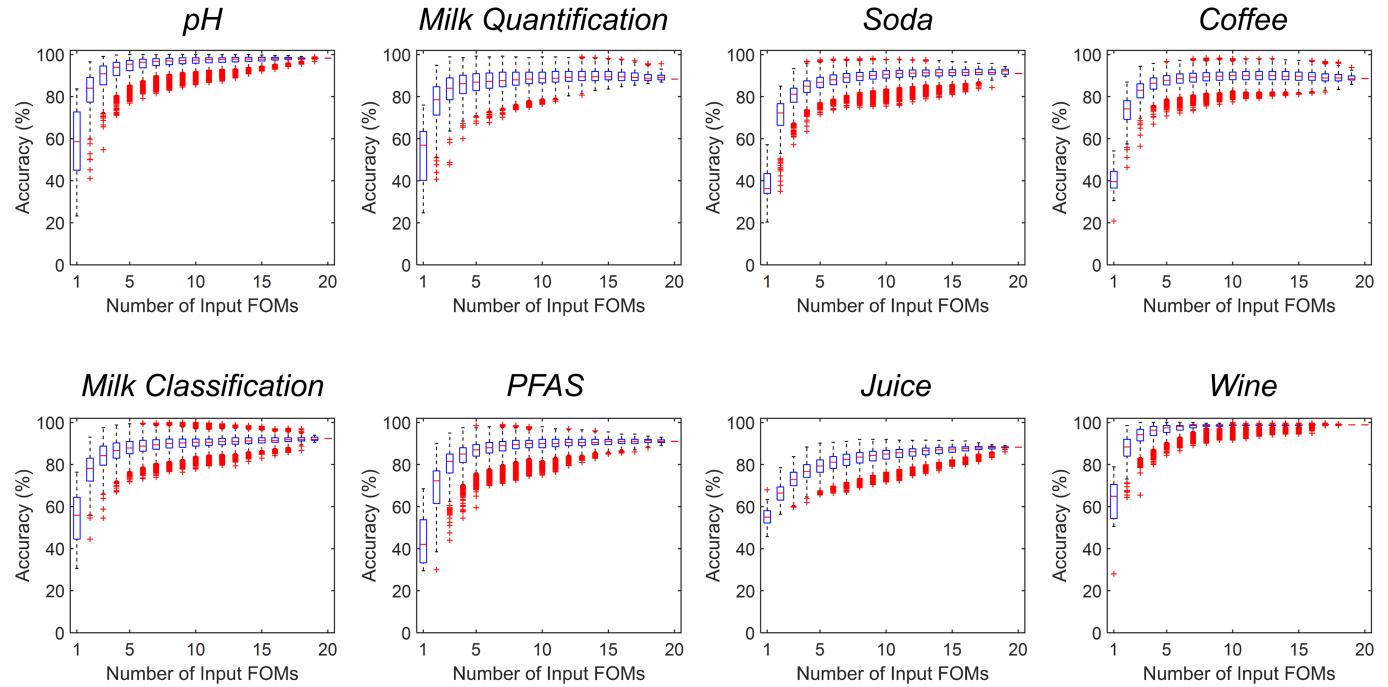
**accuracy in different training scenarios.** Confusion matrices for three training scenarios: in the first two cases, the data from one chip was used as the training set for the other chip while in the third case all data was collected and split randomly for training and testing in the ratio of 70:30. The model achieves

similar performance on the test data regardless of which chip was used for training. This finding is significant as it suggests that a model trained on graphene ISFETs from one chip can be effectively applied to chips fabricated subsequently, eliminating the need for recalibration, or retraining of the model.

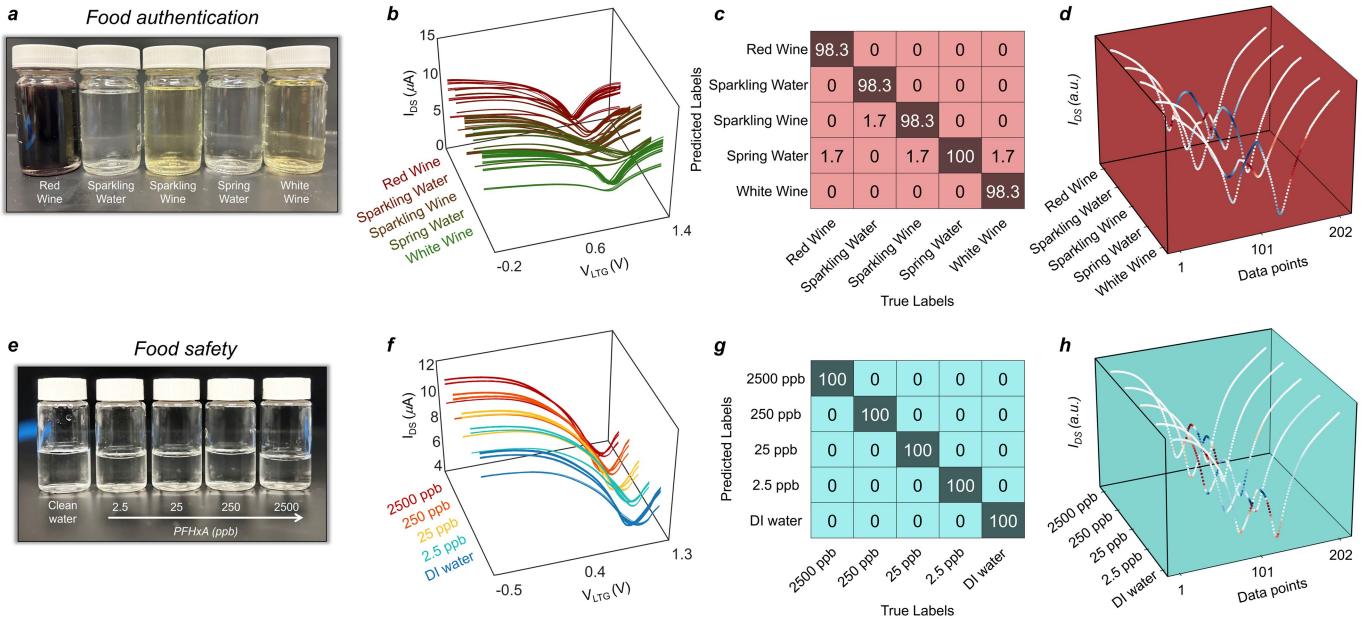
FOM	pH	Milk Quantification	Soda	Coffee	Milk Classification	PFAS	Juice	Wine
$V_{Dirac,f}$	5	18	19	7	14	17	4	12
$V_{Dirac,r}$	10	17	18	6	18	13	1	20
$I_{Dirac,f}$	1	12	14	13	8	14	15	16
$I_{Dirac,r}$	8	7	11	11	3	8	16	9
$V_{hys}$	13	19	17	12	19	18	2	18
$I_{hys}$	2	14	2	9	15	1	19	5
$I_{max,n}$	14	1	6	18	6	7	14	14
$I_{max,pf}$	18	16	3	19	17	19	9	13
$I_{max,pr}$	17	15	1	15	16	16	6	17
$I_{max/min,nf}$	3	3	13	8	10	6	3	15
$I_{max/min,nr}$	9	2	9	1	2	4	5	2
$I_{max/min,pf}$	4	9	8	2	5	11	18	3
$I_{max/min,pr}$	6	4	12	3	1	2	12	1
$maxg_{m,nf}$	12	8	15	5	13	5	10	10
$maxg_{m,nr}$	20	20	20	20	20	20	20	19
$maxg_{m,pf}$	7	10	7	4	4	9	7	7
$maxg_{m,pr}$	11	5	5	10	11	12	8	8
$maxg'_{m,r}$	16	6	4	17	7	10	11	4
$maxg'_{m,r}$	15	13	10	16	12	3	17	11
$I_{max,diff}$	19	11	16	14	9	15	13	6

**Extended Data Fig. 7 | FOM rankings for all datasets.** FOMs extracted from each dataset and ranked based on their k-NN score.

# Article



**Extended Data Fig. 8 | k-NN accuracy statistics for all datasets.** Box plots depicting the minimum, 25th percentile, median, 75th percentile and maximum accuracies of a k-NN algorithm evaluated over all possible single and multivariate combinations of input FOMs.



**Extended Data Fig. 9 | Enabling food authentication and safety with machine-learning-aided graphene ISFETs.** **a)** Optical images showing three varieties of wine (red, sparkling and white) and two varieties of water (spring and sparkling). **b)** Transfer characteristics of six graphene ISFETs tested with solutions. **c)** Confusion matrix showing per-class accuracy for authenticating wine and water based on an ANN trained using graphene ISFET data. **d)** SHAP feature analysis results highlighting the attributes in the graphene ISFET characteristics that enable the ANN to differentiate between wine and water

varieties. **e)** Optical images showing solutions of PFHxA at concentrations of 2500 ppb, 250 ppb, 25 ppb, and 2.5 ppb. **f)** Transfer characteristics of several graphene ISFETs tested with these PFAS solutions. **g)** Confusion matrix showing per-class accuracy for detecting various PFAS levels based on an ANN trained using graphene ISFET data. **h)** SHAP feature analysis results highlighting the attributes in the graphene ISFET characteristics that enable the ANN to differentiate between different PFAS concentrations.

# Article

**Extended Data Table 1 | Resulting accuracies of classification with k-NN and an ANN trained with either human-derived FOMs or the entire ISFET transfer characteristics as the input across all applications**

Task	k-NN (FOM)				ANN (FOM)				ANN (ISFET transfer characteristic)			
	Random Split	Split by Sensor	Split by Chip	Split Temporally	Random Split	Split by Sensor	Split by Chip	Split Temporally	Random Split	Split by Sensor	Split by Chip	Split Temporally
pH	98.26	94.57	96.73	100	91.64	93.75	88.21	99.03	97.09	100	96.24	100
			96.62				93.64				94.61	
Milk Quantification	88.36	40.80	-		85.09	80.75	51.25	-	72.27	95.09	48.80	-
Soda	91.00	70.00	55.20	95.24	91.00	52.71	35.75	91.00	98.17	90.60	58.00	98.02
			50.50				60.73				52.00	
Coffee	88.61	48.33	-	93.92	81.94	44.79	-	74.11	95.91	91.33	-	100
Milk Classification	92.42	86.67	-	91.52	89.62	70.49	-	76.89	99.06	86.00	-	100
PFAS	91.00	68.80	-	94.87	80.00	22.5	-	69.14	100	96.00	-	97.50
Juice Identity	90.53	88.06	75.91	93.60	93.49	85.87	66.84	93.20	98.05	90.50	71.69	98.67
			79.29				59.98				71.84	
Juice Age	95.93	90.88	84.44	96.78	85.68	78.78	71.18	87.70	99.18	94.44	77.69	99.56
			89.32				81.81				87.77	
Wine	98.91	99.60	-	100	98.11	90.83	-	100	98.67	99.60	-	100

We investigated four training scenarios where we split the measured ISFET dataset into training and testing data 1) randomly in a 70:30 ratio, 2) by sensor, 3) by chip, and 4) temporally with the first 70% of cycles from each sensor serving as the training data and the last 30% serving as the testing data. When splitting the dataset by chip, we report two accuracies corresponding to both possible combinations of utilizing one chip for training and one chip for testing.