

Cite this: DOI: 00.0000/xxxxxxxxxx

## Machine learning methods for the detection of explosives, drugs and precursor chemicals gathered using a colorimetric sniffer sensor<sup>†</sup>

Deena P. Francis<sup>\*a</sup>, Milan Laustsen<sup>b</sup>, Eleftheria Dossi<sup>c</sup>, Tuule Treiberg<sup>d</sup>, Iona Hardy<sup>c</sup>, Shai Hvid Shiv<sup>d</sup>, Bo Svarrer Hansen<sup>b</sup>, Jesper Mogensen<sup>e</sup>, Mogens H. Jakobsen<sup>d</sup>, and Tommy S. Alstrøm<sup>a</sup>

Received Date  
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Colorimetric sensing technology for the detection of explosives, drugs, and their precursor chemicals is an important and effective approach. In this work, we use various machine learning models to detect these substances from colorimetric sensing experiments conducted in controlled environments. The detection experiments based on the response of a colorimetric chip containing 26 chemo-responsive dyes indicate that homemade explosives such as HMTD, TATP, and MEKP used in improvised explosives devices are detected with true positive rate (TPR) of 70–75%, 73–90% and 60–82% respectively. Time series classifiers such as Convolutional Neural Networks (CNN) are explored, and the results indicate that improvements can be achieved with the use of kinetics of the chemical responses. The use of CNNs is limited, however, to scenarios where a large number of measurements, typically in the range of a few hundred, of each analyte are available. Feature selection of important dyes using the Group Lasso (GPLASSO) algorithm indicated that certain dyes are more important in discrimination of an analyte from ambient air. This information could be used for optimizing the colorimetric sensor and extend the detection to more analytes.

### 1 Introduction

When dealing with explosives, illicit drugs, and their precursor chemicals, “prevention is better than the cure.” Those who manufacture, transport, and use these substances take steps to obscure them from law enforcement authorities. When these substances are not intercepted, incidents such as the May 2017 Manchester Arena bombing that used the homemade explosive TATP<sup>1</sup> and increases in drug misuse deaths – majority heroin and/or morphine related – occur<sup>2</sup>. As such, the detection of these substances is vital in preventing immediate harm to people as well as damage to infrastructure and mitigating long-term negative economic and social impacts on communities.

The primary method of detection is sniffer dogs which have many advantageous properties for detection, including noses with parts-per-million (ppm) to parts-per-trillion (ppt) sensitivity to vapor<sup>3</sup>. Their disadvantages come from being living beings, they are limited to the number of substances and mixtures they can detect, cannot discriminate well between them, and require rest<sup>4</sup>. Among the instrumental detection techniques are Ion Mobility Spectroscopy (IMS) and Gas Chromatography-Mass Spectrometry (GC-TMS)<sup>5,6</sup>. These potentially destructive techniques have low detection limits, high selectivity and discrimination, but miniaturization is required to achieve portability, which is often expensive and requires trained operators to prepare and process the samples<sup>7,8</sup>. To overcome the disadvantages of the techniques while still including their necessary advantages, an alternative mechanism of detection was applied in the creation of the CRIM-TRACK sensing device<sup>9</sup> in a 2015 EU FP7 development project: colorimetric sensing. Colorimetric sensing has found applications in multiple areas such as the detection of explosives<sup>10,11</sup>, the detection of microbes<sup>12,13</sup> and in the food industry<sup>14</sup> showing capability for easy use, high sensitivity vapor detection.

The implementation of a colorimetric sensor system in the portable CRIM-TRACK device has ensured the detection of trace amounts of substances (i.e., homemade explosives), at parts-per-

<sup>a</sup> DTU Compute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark; E-mail: dfra@dtu.dk

<sup>b</sup> CrimTrack ApS, Denmark.

<sup>c</sup> Centre for Defence Chemistry, Cranfield University, Defence Academy of United Kingdom, Shrivenham, SN6 8LA, UK.

<sup>d</sup> DTU Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.

<sup>e</sup> Danish Emergency Management Agency, Chemical Division, Nørre Allé 67, 2100 Copenhagen, Denmark

<sup>†</sup> Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/cXsm00000x/

trillion (ppt) level in the air, with near real-time detection capability and no contact operator-threat<sup>11,15</sup>. It incorporates similar sensitivity and portability of sniffer dogs with the discriminatory ability of instrumentation without the use of sample preparation or a highly trained operator.

In this work, optical arrays based on color-changing (chromic) dyes are used (described in Section 2.1). Colorimetric arrays exploit supramolecular chemistry and intermolecular interactions such as acid-base, hydrogen-bonding, charge-transfer,  $\pi-\pi$  stacking, multipolar interactions, van der Waals and physical adsorption<sup>16</sup>. Their multiple potential methods of chromism, in addition to the multiple types of intermolecular interaction or reaction, “provides a high dimensionality to chemical sensing that permits high sensitivity”<sup>17</sup> and requires resilient data processing in order to fully exploit the detection capabilities.

In the context of machine learning methods for the detection of explosives and precursors, previous works have used colorimetric sensor data with machine learning classifiers like Random Forest (RF), k-nearest neighbors (KNN), and Logistic Regression (LR)<sup>11</sup>. In follow-up work,<sup>15</sup> used time series classifiers like Convolutional Neural Networks (CNN) to obtain improved detection capabilities for explosives and their precursors using the kinetic or time series nature of the data.

### 1.1 Scope of the work

In this work, we consider a version of the colorimetric chip (CC) developed by the Technical University of Denmark (DTU) and Cranfield University (CU)<sup>11</sup>. The dyes on this chip were selected based on their chemo-selective nature. The Crim-Track sniffer device is used in conjunction with this chip for the actual collection of data through “sniffing experiments”. The data collected is stored and then ultimately used for the purpose of gaining information about the presence/absence of explosives, drugs, and precursors. Machine learning models are used to perform this detection.

### 1.2 Data and code

The data generated from the sniffing experiments are available at [Will be available after acceptance]. The code for the experiments in this work are available at [Will be available after acceptance].

### 1.3 Challenges and problems addressed

There are several challenges associated with the detection of explosives in real-life. The main challenges and goals of this work are to address the following key points frequently asked by the stakeholders:

- Can the data from the sensor help in detecting explosives, drugs, and their precursor chemicals with sufficient detection performance?
- What type of analytes can the sensor detect reliably, and what kinds are harder to detect?
- Can the target analyte be detected even under differing measuring conditions with varying ambient noise, humidity, and

temperature?

In terms of machine learning and improving detection capability, this work aims to address the following questions.

- Can we use a machine learning model that can learn the kinetic nature of the chemical reaction from the multidimensional time series and improve detection capabilities?
- Which dyes on the sensor are the most useful in detecting specific analytes?

## 2 Materials

The materials used in this work are named as analytes (A) according to their role in the acquisition and analysis of detection data processes. A comprehensive list of analytes with details regarding their classification as threats is shown in Table 1. Ethylene Glycol (EG), Benzyl Methyl Ketone (BMK), Ephedrine (EPH), Hexamine (HEX), Cocaine free base (COfb) were purchased from Sigma-Aldrich and used as provided. 10 – 15% hydrogen peroxide solution in water (H<sub>2</sub>O<sub>2</sub>) was made by diluting a 35% Hydrogen Peroxide solution (CAS 7722-84-1). Triacetone Triperoxide (TATP), Hexamethylene Triperoxide diamine (HMTD), and Methyl Ethyl Ketone Peroxide (MEKP) were synthesized at DTU, CU, and Danish Emergency Management Agency (DEMA) laboratories following internal risk assessed protocols. Ambient Air (AA) is a control analyte that is composed of the lab air or the ambient conditions under which the measurements are made.

### 2.1 Dyes and colorimetric chip (CC) description

Twenty-six chemo-responsive dyes with eight repetitions (spots in total) code-named as Dye<sub>1</sub>, Dye<sub>2</sub>, ..., Dye<sub>26</sub> and seventeen blank spots (Dye<sub>27</sub>) spots, were randomly placed in a 15 × 15 grid (CC) using a nanoplotter<sup>41</sup>. The placement of the dyes and their replicas are consistent across the chips that are produced. The chip layout is as follows: 15 × 15 array, 225 spots, 26 chemo responsive dyes in 8 replicates each, spot diameter 0.7 mm, center to center distance 1 mm (Figure 1). The dyes were formulated using diacetone alcohol (CAS 123-42-2) as the main solvent (CAS 123-42-2), and after printing, solvents were evaporated at room temperature, and the chips were placed in a glove box filled with nitrogen and filtered air using a combination of High-Efficiency Particulate Air (HEPA) filtering and an active carbon filter to remove solvent residues and dust particles.

The mono-use CCs were individually packed in sealed polypropylene bags after staying in the glove box for at least 24 hours and kept in the dark until use.

## 3 Data acquisition


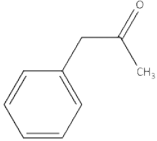
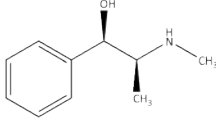
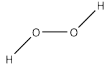
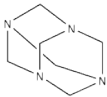
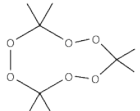
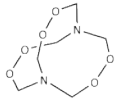
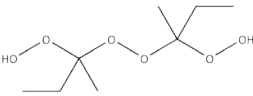
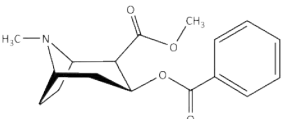
### 3.1 Detection experiments

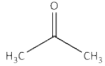
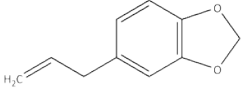
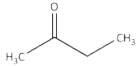
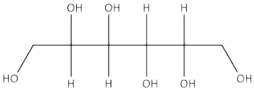
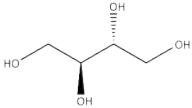
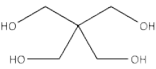
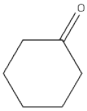
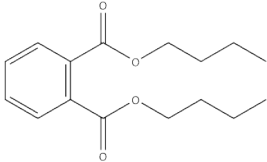
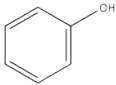
White, transparent polypropylene boxes<sup>42</sup> were used to generate the test samples of air with the vapors of the single analytes

\*\* Controlled chemicals, authorising licence for use needed.

†† 10-15% Aqueous solution.

§§ Nitrogen, oxygen, water vapor, argon, and carbon dioxide account for about 99% of the composition of air.

Analyte				Use	Vapor pressure at 25°C (Pa)
Name and CAS No	Acronym	Chemical structure	Physical state at 25°C		
Ethylene glycol 107-21-1	EG		Liquid	Precursor of explosive ethylene glycol dinitrate	~ 10 <sup>18</sup>
Benzyl methyl ketone also known as phenylacetone 103-79-7	BMK **		Liquid	Precursor of methamphetamine	21.3 <sup>19</sup>
(1R,2S)-(-)-Ephedrine 299-42-3	EPH **		Solid	Precursor of methamphetamine	0.16 <sup>20</sup>
Hydrogen peroxide	H <sub>2</sub> O <sub>2</sub>		Liquid <sup>††</sup>	Precursor of TATP, 10-15% concentration	~16 <sup>21</sup>
Hexamethylenetetramine also known as hexamine 100-97-0	HEX		Solid	Precursor of HMTD	~0.1 <sup>22</sup>
Triacetone triperoxide	TATP		Solid	Homemade explosive	6.2 <sup>23</sup>
Hexamethylene triperoxide diamine	HMTD		Solid	Homemade explosive	~6.5e-6 Pa <sup>24</sup>
Methyl ethyl ketone peroxide	MEKP		Liquid	Homemade explosive	1.3 <sup>25</sup>
Cocaine free base 50-36-2	COfb **		Solid	Controlled drug	0.0003 <sup>26</sup>

Acetone 67-64-1	Ac		Liquid	Precursor of homemade explosive, Manufacturing solvent	30800 <sup>27</sup>
Safrole 94-59-7	SA		Liquid	Precursor of methamphetamine	9.3 <sup>28</sup>
Methyl ethyl ketone 78-93-3	MEK		Liquid	Precursor of MEKP	12100 <sup>29</sup>
Mannitol 69-65-8	MA		Solid	Precursor of homemade explosive	Data unavailable (Very <sup>30</sup> low)
meso-Erythritol 149-32-6	ER		Solid	Precursor of homemade explosive	Data unavailable (Very <sup>31</sup> low)
Pentaerythritol 115-77-5	PER		Solid	Precursor of homemade explosive	1.07e - 7 <sup>32</sup>
Cyclohexanone 108-94-1	Cy		Liquid	Manufacturing solvent	667 <sup>33</sup>
Dibutyl phthalate 84-74-2	DP		Liquid	Additive in explosive formulations	1.9e-3 <sup>34</sup>
Phenol 108-95-2	Ph		Solid	Toxic. Important chemical intermediate in many industrial processes	53 <sup>35</sup>


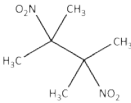
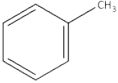
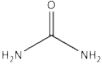
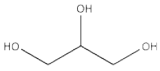
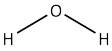
Ammonium nitrate 6484-52-2	AN		Solid	Oxidizer for homemade explosive formulations	0.32 <sup>36</sup>
2,3-Dimethyl-2,3-dinitrobutane 3964-18-9	DMNB		Solid	Additive (taggant) in explosive formulations	0.27 <sup>37</sup>
Toluene 108-88-3	Tol		Liquid	Solvent	3800 <sup>38</sup>
Urea 57-13-6	Ur		Solid	Precursor of explosives	1.6e-3 <sup>39</sup>
Glycerin 56-81-5	Gly		Liquid	Precursor of home-made explosives	321e3 <sup>40</sup>
Ambient air	AA <sup>§§</sup>		Gas	Chemical	
Water	H <sub>2</sub> O		Liquid	Control analyte and precursor of home-made explosives in aqueous solution	

Table 1 Analytes used for the colorimetric detection experiments with Crim-Track sniffer device. Typical quantities of 25 – 50 mg of solid analytes and 0.1 ml of liquid analytes or solution of analytes were used for the detection experiments



Fig. 1 The colorimetric chip.

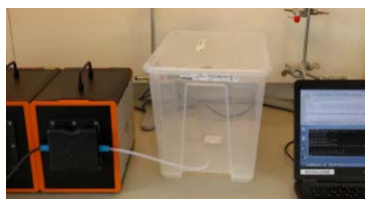


Fig. 2 Sampling of analyte from the box (middle) into the Crim-Track sniffer device (left). The experiment(s) were performed by using a laptop (right).

with the aim of mimicking as close as possible realistic detection scenarios. Small samples of the analytes, typically 25-50 mg of solid analytes and 0.1 ml of liquid analytes (or solutions), were placed inside the boxes' on an aluminum dish and left to evaporate/sublimate for 24h. The test sample boxes were handled in ambient air, and the lids were not sealed. This allows exchange with the ambient air to some extent and to better represent the real detection scenario (i.e., traveling bag at airport security control). After 24h, a sampling tube was connected to the box, and the air containing the analyte was sucked from the box directly into the Crim-Track sniffer device<sup>9</sup>. During the acquisition of data, the boxes were left unsealed to the ambient air for pressure equalization.

### 3.2 Datasets

The data collected by Crim-Track sniffer campaigns were organized into datasets. Within each dataset are a number of data points of an analyte's interaction with the chip, measured under similar temperature and humidity conditions in a laboratory. Each dataset has a number of analyte's data, selected based on the aim of the given experiment. For example, a dataset containing HMTD, TATP, H<sub>2</sub>O<sub>2</sub>, EG, and AA would be created for the purpose of distinguishing between explosives, precursors, and control analytes. Table 2 shows the dataset names presented in this work, their constituent analytes, and their purpose.

Corresponding to each of the datasets generated by the sniffing experiments, we can consider two kinds of data for the analysis.

- Diff data: this data is the result of computing the differences between the first image of the chip before exposure to the analyte and the last image of the chip, taken after exposure to the analyte. In order to aggregate the color changes of a dye, the median value of its replicas is computed. The size of this type of dataset is  $N \times 78$ , where  $N$  is the number of

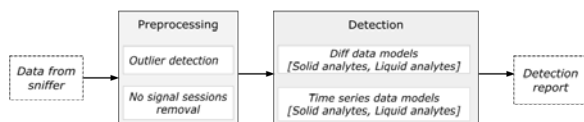


Fig. 3 Overall setup of the detection process using sniffer data

measurement sessions made within a dataset and  $78 = 26$  dyes  $\times 3$  color channels, Red (R), Green (G) and Blue (B).

- Time series data: this data is the result of capturing the evolution of dye colors on the chip through the course of the measurement. Similar to the diff data, the median value of the replicas of each dye is computed. The size of this dataset is  $N \times T \times 78$ , where  $N$  is the number of measurement sessions in a dataset,  $T$  is the number of time points during which the images of the chip are captured, and  $78 = 26$  dyes  $\times 3$  color channels.

There are advantages to using both kinds of data. The diff data is easy to measure and can be analyzed with both off-the-shelf machine learning algorithms and statistical methods. In fact, this kind of data is prevalent in colorimetric sensing applications and is considered state-of-the-art. Time series data requires collecting images of the chip at selected time points and also requires machine learning methods that deal with sequentially ordered data. Recently, time series data has shown helpful in improving the detection capability of several explosives and their precursors<sup>15</sup>. This kind of data is helpful in understanding the temporal evolution of the dye color changes as the reaction progresses. Table 3 shows the number of measurements of each analyte present in each dataset.

## 4 Methods

The data collected by the sniffer undergoes a data processing and detection pipeline that provides an indication of the presence of the target analyte. The focus of this work is on using machine learning methods to provide predictions with sufficient detection performance, which entails obtaining a high true positive rate (TPR).

Depending on the type of data, different machine learning methods are employed. Figure 3 shows the overall setup of the detection process from the sniffer data.

### 4.1 Pre-processing

In the data pre-processing phase, an outlier method is employed to detect and remove outlier or non-conforming measurements. During the measurement process, several errors can creep in, which can cause the data to be significantly different from the norm. In this context, outlier detection can be considered a method for quality assurance. Outlier detection is a critical part of most applications that deal with real data. This is particularly true in the case of colorimetric sensor data. Figure 4 shows the time series of a single dye on the chip when exposed to the explosive TATP in two different measurement sessions. In the second session (bottom figure), the dye behaves in a different manner

Dataset		Analytes			
Name	ID	Control	Chemicals	Aim of the experiment	
Explosives 1	Exp1	EG, AA	HMTD, MEKP, TATP, H <sub>2</sub> O <sub>2</sub>	Detecting peroxide explosives	
Explosives 2	Exp2	H <sub>2</sub> O, AA	TATP, H <sub>2</sub> O <sub>2</sub> , HEX	Detecting peroxide explosive and its precursor	
Precursors 1	Pc1	EG, AA	BMK, Ac, H <sub>2</sub> O <sub>2</sub>	Detecting precursors of drugs and peroxide explosives	
Precursors 2	Pc2	EG, AA	Ph, DMNB, MEK, H <sub>2</sub> O <sub>2</sub>	Detecting precursors of peroxide explosives and taggant	
Precursors 3	Pc3	EG, AA	Gly, Tol, Ur, H <sub>2</sub> O <sub>2</sub>	Detecting precursors of peroxide explosives, solvents	
Precursors 4	Pc4	EG	MA, ER, PER, H <sub>2</sub> O <sub>2</sub>	Detecting precursors of explosives and peroxide explosives	
Precursors 5	Pc5	EG	AN, DP, H <sub>2</sub> O <sub>2</sub> , Cy	Detecting precursors of peroxide explosives and additives	
Drugs and precursors 1	D1	EG, AA	EPH, COfb, Ph, SA, H <sub>2</sub> O <sub>2</sub>	Detecting controlled drugs and their precursors	
Drugs and precursors 2	D2	AA	COfb, SA	Detecting drugs and precursors of controlled drugs	

Table 2 Datasets from measurements obtained when the chip is exposed to specific chemicals.

Analyte/Dataset	Ex1	Ex2	Pc1	Pc2	Pc3	Pc4	Pc5	D1	D2
TATP	63	9							
HMTD	68								
MEKP	33								
BMK			10						
H <sub>2</sub> O <sub>2</sub>	33	8	10	23	20	15	13	16	
HEX		9							
EG	32		10	14	18	10	12	17	
Ac			9						
H <sub>2</sub> O		9							
Ur					20				
Ph				8					
MA						5			
ER						6			
PER						14			
AN							13		
Cy							12		
EPH								17	
COfb								18	24
DMNB				21					
DP							14		
SA								18	23
MEK				19					
Gly					18				
Tol					17				
AA	30	8	8	22	20	10	12	17	24

Table 3 The number of measurements of each analyte in each dataset.



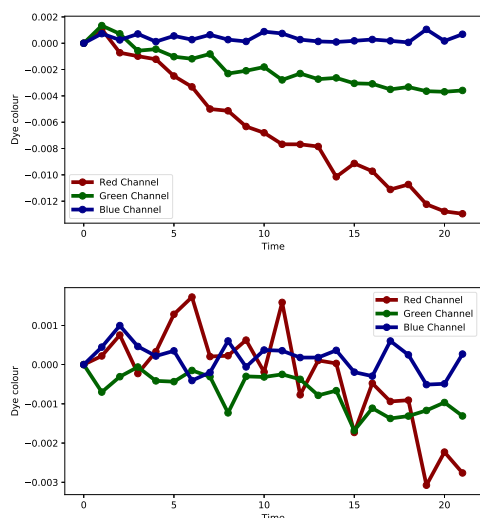


Fig. 4 An example of a normal (top) and outlier (bottom) measurement session for TATP with responses on the red channel, green channel, and blue channel.

than in the first session (top figure).

The outlier detection method is based on the Gaussian Mixture Model (GMM)<sup>43</sup>. The method for outlier detection assumes that in the case of outlier sessions, the final color values of the dyes are widely different from the normal sessions. This variation in the final color values can be captured by constructing an unsupervised probabilistic model that can capture the presence of the “normal” data measurements and the “outlier” measurements. A mixture of  $k$  Gaussian distributions modeling the entire data is learned. The data points are labeled as outliers if their log likelihoods exceed a threshold, otherwise they are labeled as normal.

## 4.2 Visualizations

Visualizing the data is a useful method to inspect the result of an experimental measurement manually. It also aids in interpreting the results of the detection phase. In this work, we present the following visualizations of the data.

- Raw data visualizations
  - **Difference maps:** these visualizations<sup>44</sup> are created by subtracting the pixel values of the last image of the chip (after exposure to the analyte) from the first image (before exposure to the analyte). The differences are then scaled to an appropriate value to make them visible to the naked eye.
  - **Raw time series signals:** these visualizations are plotted with the raw RGB values of the individual dye spots on the chip over time.
- Compressed visualizations
  - **t-distributed Stochastic Neighbor Embedding (t-SNE):** Unsupervised dimensionality reduction methods can be used to extract the most informative components of datasets without the need to consider the

labels/class or names of the analytes it is associated with. This can be used to visualize multidimensional datasets, such as those considered in this work. t-SNE<sup>45</sup> is one technique that embeds high-dimensional data into a lower-dimensional space<sup>46</sup>. It uses the notion of conditional probability distributions as a measure of similarity between data points.

## 4.3 Detection

The detection of explosives and precursors using supervised machine learning models is the focus of this section. We use the term *training set* to refer to data that the machine learning model learns from and *test set* to refer to the data that is used to test the performance of the model. The two sets of data are disjoint; that is, they do not contain any overlapping measurements.

**Stratified k-fold cross validation:** Cross-validation is a procedure that creates random splits in the given dataset and produces two disjoint sets of data points: the *training set* and the *test set*. The training of the model is done on the training set, and prediction is made on the test set. A  $k$ -fold cross-validation procedure creates  $k$  such random splits. Training is done on  $k - 1$  splits, and prediction is made on the remaining split, which yields  $k$  prediction results. Stratification is a procedure that ensures that any test set formed as a result of any of the  $k$  splits contains approximately the same distribution of classes.

**Multi-class classification:** Most of the classifiers used in this work are defined for performing a binary (two-class) classification task. To perform multi-class classification, we use the *one-versus-rest* classification strategy. In this method, a classifier is trained to detect a single class, and it learns to distinguish between data points of this class and all other classes.

### 4.3.1 Diff data classification

For the Diff data, we use classifiers such as K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM) with a squared exponential kernel, Logistic Regression with L1 penalty (LR1), Logistic Regression with L2 penalty (LR2) GPLASSO and Gaussian Process Classification (GPC).

**KNN classifier:** It is a non-parametric distance-based algorithm that assigns classes to instances based on the notion of distance to its neighbors<sup>47</sup>. It has a parameter  $k$  that defines the number of data points it must be similar to in order to belong to this class of data points.

**RF classifier:** This classifier<sup>48</sup> is an ensemble of decision tree classifiers. A decision tree uses a feature partitioning mechanism in order to predict the class of a given test data point. The partitioning into regions  $R$  is typically done using a tree structure that makes “decisions” to split the data points based on a splitting criterion such as Gini index.

**SVM classifier:** This classifier discriminates between the different classes of data by means of a maximum margin hyperplane<sup>49</sup>. This hyperplane is learned such that its distance from data on either side of it is maximised.

**LR1 classifier:** This classifier models the conditional probability of a class label given the data<sup>50</sup>. This is done by using a logistic function with a parameter  $\beta$ . Learning this model involves solv-



ing an optimization problem involving the norm of  $\beta$ . If the norm chosen is the  $l1$  norm, then the model is referred to as LR1.

**LR2 classifier:** This classifier is a logistic regression model, similar to the LR1 classifier. In the LR2 classifier, the norm of the parameter  $\beta$  is chosen to be the  $l2$  norm.

**GPLASSO classifier:** This classifier combines a group lasso penalty term in a logistic regression setting to achieve feature selection in conjunction with classification<sup>51-53</sup>. This ensures that groups of features that are important for the classification task are learned.

**GPC classifier:** This classifier is based on the idea of using a distribution of functions to model the data distribution. The input  $\mathbf{x}$  and the output  $y$  is assumed to be related by a latent function  $f$  as  $y = \sigma(f(\mathbf{x})) + \varepsilon$ , where  $\sigma$  is a squashing function like a sigmoid function,  $\varepsilon$  is a noise term. In GPC,  $f$  is assumed to be drawn from a Gaussian process,  $f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ , where  $m(\mathbf{x})$  is the mean function and  $k(\mathbf{x}, \mathbf{x}')$  is the covariance function. The covariance function  $k$  gives this classifier its power to learn complex latent functions, and hence better discriminatory abilities<sup>54</sup>.

### 4.3.2 Time series classification

For the time series data, we use classifiers such as CNN, Fully Convolutional Neural Network (FCN), Temporal Convolutional Network (TCN-ED), Long Short Term Memory network (LSTM), KNN with Dynamic Time Warping (KNN-TS) and Support Vector Machine (SVM) with a time series kernel (SVM-TS).

**CNN classifier:** In general, this classifier is composed of one or more layers of convolution filters and pooling layers. The convolution filter is a matrix of scaling factors. The convolution operation amounts to the application of this linear filter on the input matrix using a sliding mechanism wherein the filter is multiplied with various parts of the input matrix. This convolution operation has been found to be very beneficial in extracting important features from the data, particularly from images<sup>55</sup>. The pooling operation is a summarizing mechanism that condenses information from a given layer. In this work we use a CNN architecture consisting of a layer of 1-D convolutions followed a layer of average pooling.

**FCN classifier:** The FCN is a variant of a CNN with a special architecture<sup>56</sup>. In this work we use an architecture consisting of a block of a layer of 1-D convolutions, a batch normalization layer, repeated thrice, followed by an average pooling layer.

**TCN-ED classifier:** This classifier uses temporal convolution (convolution along time dimension)<sup>57</sup>. In this work, the encoder has the architecture of a 1-D convolution layer and a max pooling layer, and the decoder consists of an upsampling layer followed by 1-D convolution layer.

**LSTM classifier:** This classifier has a graphical structure that consists of feedback connections, cells, input and forget gates<sup>58</sup>. This model is capable of handling data with long-range dependencies such as time series.

**KNN-TS classifier:** This classifier uses the same basic idea of a KNN classifier, but uses a time series distance metric such as Dynamic Time Warping (DTW)<sup>59</sup> to compute the distance between the data points, which are time series in this instance.

**SVM-TS classifier:** This classifier uses the SVM classifier with a time series kernel called Global Alignment Kernel (GAK)<sup>60</sup>.

## 5 Experimental results

This section describes the results of sniffing experiments, exploratory analysis, and detection obtained using the datasets and the machine learning classifiers.

### 5.1 Setup

The code was written in python. The scikit-learn<sup>61</sup> package was used for the diff data classifiers KNN, RF, LR1, LR2 and GPC. The group-lasso package<sup>62</sup> was used for the GPLASSO classifier. The time series classifiers: CNN, FCN, TCN-ED and LSTM were implemented in tensorflow<sup>63</sup>. The tslearn package<sup>64</sup> was used for the other time series classifiers KNN-TS and SVM-TS.

For the SVM classifier, we used the squared exponential kernel. In the case of GPC classifier, the predictive distribution is intractable and schemes such as Laplace approximation, Expectation propagation and Markov Chain Monte Carlo are typically used to preform approximate inference. In this work, we use Laplace approximation. For this classifier we used the squared exponential kernel.

### 5.2 Raw data visualizations

We present visualizations for diff data and for time series.

#### 5.2.1 Difference maps

Figure 5 shows the various difference maps of a subset of the analytes. By visual inspection of these images, we see certain dyes that “light up” or react with the analyte it is exposed to. Although these images serve as a quick method for visually inspecting the result of a reaction, there are certain disadvantages with these maps. One such problem is that not all dye variations are clearly visible. This makes it difficult to determine if the analyte has produced a valid response from the chip. Another problem with these maps is that the manual inspection of several such images is tedious and error-prone. Finally, a dye can change color in multiple ways that cannot be visualized with only the difference maps.

#### 5.2.2 Time series plots

The plots of raw time series signals obtained from the chip when exposed to various analytes are shown in Figure 6. For demonstration purposes, responses of the most reactive dyes on the chip alone are shown. Figure 6 shows the time series obtained when the chip is exposed to various analytes for a period of approximately 5 minutes with a total of 22 images.

### 5.3 Unsupervised visualizations

The t-SNE plots of the datasets are shown in Figure 7. This figure shows certain distinct clustering or grouping among the analytes, which is indicative of the ability of the measured data to differentiate between the various analytes. For example, Figure 7 (b) shows a clear separation of TATP from HEX.

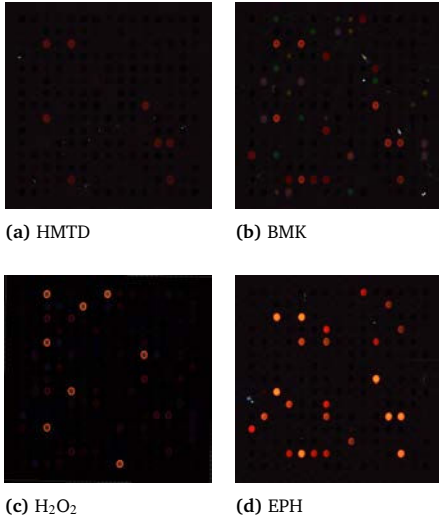


Fig. 5 Difference maps of the experiments with HMTD, BMK, H<sub>2</sub>O<sub>2</sub> and EPH considered in this work.

#### 5.4 Diff data classification

Depending on the phase (solid or liquid) of the analytes in the dataset, a classifier is trained on analytes of the same phase. From a practical viewpoint, when using the sniffer to detect an analyte in real life, the operator can visually observe the phase of the analyte, and hence this information serves as additional information that can potentially improve the detection capability. Therefore, in our prediction step, we divide the analytes from the same dataset into two groups: solids and liquids, and both train and predict each of these sub-categories of data separately. The per-dataset performance of the best performing classifier is explored in detail in the following section. The values of the metrics are shown as the mean of the metric  $\pm$  the standard deviation computed over 5-fold cross-validation splits.

The values of the metrics are computed from a confusion matrix obtained after the testing phase of the classification. In this work, we follow the convention of having the true labels along the columns of the confusion matrix and the predicted labels along the rows of the matrix. The metrics used while reporting the results are as follows, where TP: True Positives, FN: False Negatives, FN: False Negatives, TN: True Negatives, and FP: False Positives:

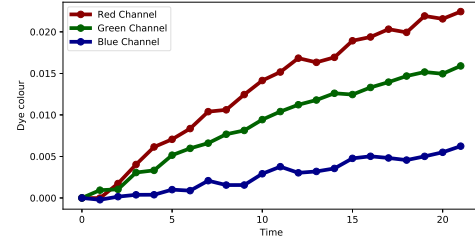
- True Positive Rate (TPR) or Sensitivity or Recall (Eqn. 1): It is a measure of how well the correct number of true positives are identified by the model.

$$TPR = \frac{TP}{TP + FN} \quad (1)$$

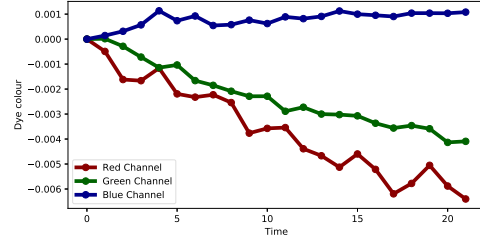
- False Negative Rate (FNR) or Miss Rate (Eqn. 2): It is a measure of the mistakes made by the model while identifying the correct number of true positives.

$$FNR = \frac{FN}{FN + TP} \quad (2)$$

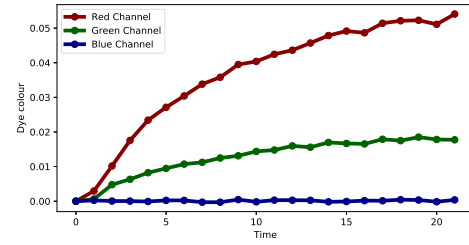
- False Positive Rate (FPR) (Eqn. 3): It is a measure of the



(a) Response of Dye 13 with BMK



(b) Response of Dye 16 with BMK



(c) Response of Dye 17 with BMK

Fig. 6 Plot of time series obtained from the chip for BMK for selected dyes. The RGB channel responses of the dyes are shown as red, green, and blue lines respectively.

rate of false predictions made by the model.

$$FPR = \frac{FP}{FP + TN} \quad (3)$$

- True Negative Rate (TNR) or Specificity (Eqn. 4): It is a measure of how well the correct number of false negatives are identified by the model.

$$TNR = \frac{TN}{TN + FP} \quad (4)$$

- F1-score (Eqn. 5): It is a measure that combines precision and recall. The metric of precision measures how well the model does in identifying the correct positives from all the positive identifications made by it.

$$F1 \text{ score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

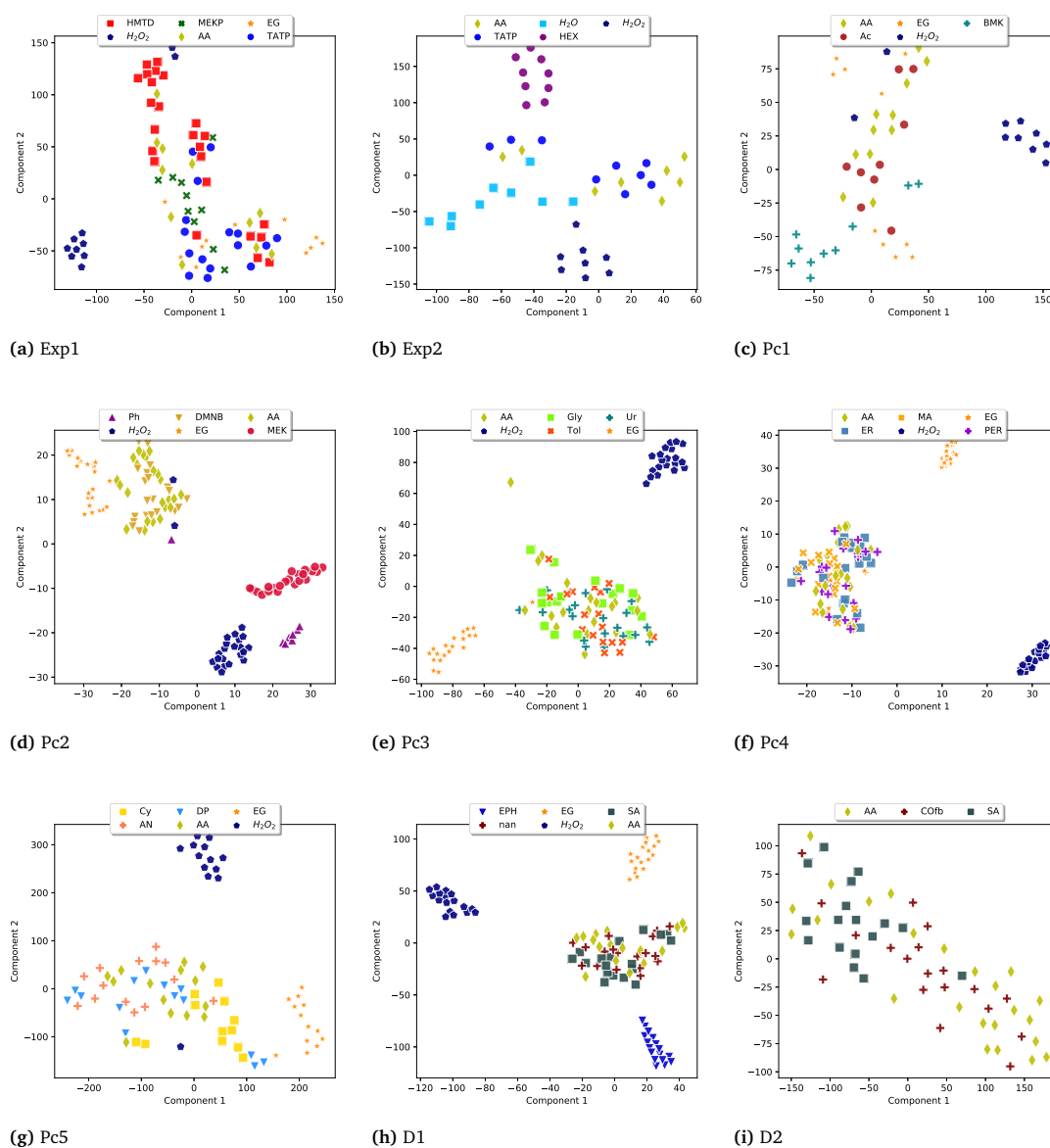


Fig. 7 t-SNE plots of all the datasets.

### 5.4.1 Per dataset classification

In all results discussed in this section, the classification has been done with the analytes in the dataset, including AA as a control.

Table 4 shows the values of the performance metrics obtained by classifying the Ex1 dataset. In this table, the results of GPLASSO (for solid analytes) and RF (for liquid analytes), the best performing classifiers for the specific configurations, are reported. We see that TATP and HMTD are detected reasonably well (TPR of 71%, 73% respectively), however, both the FPR and FNR are also high values.

To investigate this result further, we look at the confusion matrix (Table 5) and observe that there are several misclassifications between TATP and HMTD, and false alarms are caused by all AA datapoints. The misclassification between HMTD and TATP is not a serious concern because both these analytes are peroxide explo-

sives, and the goal has to some degree been achieved in dataset Ex1. For the liquid analytes, the detection performance of MEKP explosive is good (TPR of 82% and FNR of 18%), H<sub>2</sub>O<sub>2</sub> and EG are also detected well with TPR of 80% and 68% respectively.

For the Ex2 dataset, the results in Table 6 show that TATP is detected with good TPR = 90% and FNR of 10%. It is interesting to note that HEX is detected perfectly (TPR of 100% and FNR of 0%). Both H<sub>2</sub>O<sub>2</sub> and H<sub>2</sub>O are detected with good TPR of 100% and 80% respectively. A detailed look at the confusion matrix is provided in Table 7. There is just a single instance of TATP that is misclassified as AA.

For the Pc1 dataset, the results of the RF classifier are reported in Table 8. All the analytes in this dataset are detected with good TPR ranging from 70% to 100%. The high value of FNR (30%) for Ac is due to the misclassifications with EG and AA as shown in the confusion matrix in Table 9.

Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	HMTD	71.76±15.02	28.24±15.02	39.77±10.78	60.23±10.78	63.38±12.19
	TATP	73.21±11.19	26.79±11.19	29.74±10.71	70.26±10.71	67.02±10.00
Liquid	EG	68.00±0.26	32.00±0.26	5.00±0.08	95.00±0.08	74.00±0.28
	H <sub>2</sub> O <sub>2</sub>	80.00±0.32	20.00±0.32	3.00±0.04	97.00±0.04	81.00±0.30
	MEKP	82.00±0.17	18.00±0.17	13.00±0.03	87.00±0.03	74.00±0.13

Table 4 Classification results for Ex1 dataset obtained by using the classifier. The best classifier's results for solids (GPLASSO) and liquid (RF) analytes are shown separately.

	TATP	AA	HMTD
TATP	55	13	23
AA	0	0	0
HMTD	21	23	59

Table 5 Confusion matrix of solid analyte classification of Ex1 dataset with the GPLASSO classifier.

For the Pc2 dataset, the taggant DMNB is not detected with good detection performance, but the liquid analytes are detected with good TPR ranging from 90% to 100% as shown in Table 10.

For the Pc3 dataset, the results of SVM and RF classifiers are reported in Table 11. The solid precursor Ur is detected with a TPR of 75% and the liquid analytes except Gly are detected with good TPR ranging from 82% to 95%. The misclassifications for the liquid analytes is shown in Table 12. It can be observed that Gly, which has the worst detection performance is often misclassified with AA, and vice versa. This indicates the inability of the classifier to distinguish between the data of the two analytes.

For the Pc4 dataset, the results for SVM and KNN classifiers are shown in Table 13. Almost all solid analytes (MA, ER) are misclassified as PER as can be seen in Table 14. The liquid analytes are classified with good TPR ranging from 90% to 100%.

For the Pc5 dataset, the results of RF and KNN classifier are shown in Table 15. The solid analyte, AN is detected with a TPR of 93%. This indicates that it can be distinguished from AA by the classifier. The liquid analytes are classified with perfect TPR and FNR of 100% and 0% respectively.

For the D1 dataset, the results of KNN and RF classifiers are shown in Table 16. The solid analytes COfb and EPH are detected with good TPR of 75% and 100% respectively, while the liquid analytes achieve good TPR of classification.

For the D2 dataset, the results of SVM and KNN classifiers are shown in Table 17. The solid analyte COfb is detected with a TPR of 72%. This analyte is misclassified with AA as shown in Table 18. The liquid analyte, SA is classified with good TPR.

Summary of per-dataset classification:

- Among the explosives, TATP is detected with a TPR of 73 – 90% (Tables 4, 6), HMTD has a TPR of 71% (Table 4). The explosive oxidizer AN is also detected with a high TPR (~93%) (Table 15). The liquid explosive MEKP is also detected with a good TPR of 82% (Table 4).
- The precursors HEX, EPH, BMK, and MEK are detected per-

fectly (100% TPR) (Tables 6, 8, 16). Other precursors such as H<sub>2</sub>O<sub>2</sub>, EG, Ac, Ur, SA are detected with varying accuracies, but have TPR > 70%. The toxin Ph is detected perfectly (TPR of 100% and FPR of 0%) (Table 10) and the solvent Tol has been detected with good TPR of 82% and FPR of 13% (Table 11).

- The drug COfb was detected with a TPR ranging from 72 – 75% (Tables 16, 17).
- The results obtained in this section are consistent with the t-SNE plots in Figure 7. For instance, MEK is perfectly separated from all other analytes in the plot for Pc2, and it is detected perfectly as shown in Table 10. Similarly, for H<sub>2</sub>O<sub>2</sub>, EG and EPH visibly separable clusters in the plots correspond to good detection performance.

#### 5.4.2 Combined dataset classification

In this section, we combine all the datasets and generate a single dataset with 25 analytes. This allows us to explore the robustness of the models with respect to variations across datasets. The objective of this is to study the effect of different experimental conditions, such as humidity and temperature, on the detection capability. The stratified 5-fold cross-validation results for the best classifier (RF) are reported in Table 19. In this dataset, solid analytes such as HMTD, HEX, and EPH obtain good TPR, while TATP, COfb, and AN are not detected with good detection performance compared to their results in Section 5.4.1. We consider two primary factors as the cause. The first has to do with the number of analytes in the dataset. The combined dataset presents a much more challenging problem compared to a two or three-class problem. The second factor is the variation in measurements of the same analyte. In this combined dataset, measurements of the same analyte from different datasets, and hence different measuring conditions, are utilized. In order to investigate further, we examine the confusion matrix for the solid analytes as shown in Table 20. It can be observed that TATP is misclassified mostly with HMTD and vice versa. For the liquid analytes, most of the precursor analytes as well as Ph are detected with good detection performance despite the different measuring conditions. The confusion matrix of the solid analytes for the combined dataset is shown in Table 20.

We summarize the findings:

Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	HEX	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	TATP	90.00±0.20	10.00±0.20	47.00±0.12	53.00±0.12	63.00±0.11
Liquid	H <sub>2</sub> O <sub>2</sub>	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	H <sub>2</sub> O	80.00±0.24	20.00±0.24	5.00±0.10	95.00±0.10	80.00±0.16

Table 6 Classification results for Ex2 dataset obtained by using the classifier. The best classifier’s results for solids (KNN) and liquid (RF) analytes are shown separately.

	TATP	HEX	AA
TATP	10	0	10
HEX	0	11	0
AA	1	0	0

Table 7 Confusion matrix of solid analyte classification of Ex2 dataset with KNN classifier.

- Liquid analytes such as H<sub>2</sub>O<sub>2</sub>, BMK, Ph, MEK, MEKP, and EG are detected with TPR > 80% (Table 19) even with different measuring conditions.
- Solid analytes such as HMTD, HEX, and EPH are detectable with TPR > 70% (Table 19) even in the presence of varying measuring conditions. TATP is detected, but there are several misclassifications with HMTD. COfb is misclassified mostly with AA which could be attributed to its low vapor pressure.
- It is expected that analytes such as AN and PER will not be detected using the current chip because of their very low vapor pressure.

### 5.5 Time series data classification

In this section, we explore the benefits of using the kinetic (or time series) nature of the colorimetric sensing data. Table 21 shows the results of the multi-class classification of the combined dataset with all the analytes. For the solid analytes, DMNB, EPH, HEX, and HMTD are detected with good TPR ranging from 70% to 93%. When compared to the diff data results (Table 19) there are some improvements in the detection performance of solids such as DMNB, HEX, and EPH with 55%, 10% and 3% improvement, respectively. An improvement of 60% in the detection performance of Ac was observed for the liquid analytes.

### 5.6 A study on dye variation and importance

In addition to detection performance, it is useful to understand the most important dyes in a colorimetric sensing experiment. There are two perspectives to studying the importance of dyes. The first is a statistical perspective, where the aim is to answer the question of *which features vary during an experimental session?*. The second perspective is a classification viewpoint wherein the underlying question is, *which features are the most informative in terms of distinguishing between various analytes given the dataset in question?*

From a statistical viewpoint, we can estimate the names of the dyes that vary when exposed to analytes. A simple test of variation would be to test if the diff data can be fitted to a linear regression model. Table 22 reports the dyes that are found important using this technique.

The chip contains a range of 26 chemo-selective dyes and an empty dye which vary in their function and responsiveness. There are some dyes, such as dyes 15, 16, and 17, that respond to almost all analytes, while there are dyes (4, 12, 13, and 14) that respond to only a few analytes. In the prediction stage, data from all the dyes are utilized. This may sometimes be detrimental to the performance because the presence of irrelevant or redundant information can adversely affect the discriminatory power of the classifier. Selecting and using the most important features for classification can improve detection capabilities. There are several feature selection methods in literature<sup>65–67</sup>. In this work, we chose to work with the method of GPLASSO.

There are attractive properties of the GPLASSO estimator that motivate us to use this method in our application. This method has the benefit of performing feature selection at a group level. In the context of colorimetric sensor data, the features are the RGB channel values of the dyes, and the groups correspond to the dyes. For example, group 1 corresponds to the set of features extracted from dye 1, red, green, and blue channel values. For this analysis, we considered the combined datasets and perform several binary classifications of an analyte versus all ambient air measurements. Table 23 shows the dyes selected as being the most important by the GPLASSO algorithm for distinguishing between any analyte and ambient air. These results give insights into the detectability of an analyte. It can be observed that most of the analytes are detectable from ambient air, with F1-score > 71% (Table 23). The dyes selected by the algorithm as being important can be used in optimizing the sensor, for example adding more dye spots of the most important dyes and fewer of the others. It can be observed that only a few dyes are needed from a classification viewpoint. These dyes are important for distinguishing an analyte from AA. However, if we are interested in distinguishing between two generic analytes A and B, for example, TATP and HMTD, then the sets of dyes that are important for this task would be different from the ones identified in Table 23.

## 6 Overall discussion

**Question 1:** *Can the data from the sensor help in detecting explosives, drugs, and their precursor chemicals with sufficient true positive rate (TPR > 70%)?*

Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Liquid	Ac	70.00±0.40	30.00±0.40	8.00±0.10	92.00±0.10	-
	BMK	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	EG	90.00±0.20	10.00±0.20	3.00±0.06	97.00±0.06	90.00±0.20
	H <sub>2</sub> O <sub>2</sub>	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00

Table 8 Classification results for Pc1 dataset obtained by using the RF classifier. This dataset only has liquid analytes.

	BMK	EG	H <sub>2</sub> O <sub>2</sub>	Ac	AA
BMK	12	0	0	0	0
EG	0	11	0	1	0
H <sub>2</sub> O <sub>2</sub>	0	0	12	0	0
Ac	0	1	0	9	3
AA	0	0	0	1	7

Table 9 Confusion matrix of liquid analytes classification of Pc1 dataset with RF classifier.

The current colorimetric sensor produced data that can be used to detect various explosives such as TATP, HMTD, and MEKP with varying accuracies. It was found that these analytes could be detected when they are measured under similar conditions of humidity and temperature. For example in Exp1 and Exp2, the TPR of TATP ranges from 73 – 90% (Tables 4 and 6). To understand this further, a confusion matrix for the analytes that belong to each category is created and presented in Tables 24, 25, 26. These tables correspond to the confusion matrices derived for the Ex1, Ex2, and combined dataset for diff data, respectively. In this table, the following abbreviations are used: D: Drugs, E: Explosives, EA: Explosive Additive, EP: Explosive Precursor, DP: Drug Precursor. It can be observed that explosives and explosive precursors are detected with good TPR of > 90% (this value was computed from the confusion matrices), since they have very few misclassifications when compared to the correct classifications.

**Question 2:** *What type of analytes can the sensor detect easily and what kind are harder to detect?*

Liquid analytes are in general easier to detect. The liquid analytes we used in this work generally have higher vapor pressure with respect to the solid analytes. From the results obtained it can be observed that liquid analytes like H<sub>2</sub>O<sub>2</sub>, MEK, Ph, BMK, EG can be detected with very good TPR. Although solid analytes are harder to detect, within a measurement experiment, where the temperature and humidity are the in a fixed range, explosives like TATP, MEKP, and HMTD are detected with good TPR. This also applies to the precursor EPH. It was observed that the drug COfb was detected with TPR 72 – 75% (Tables 16, 17) and SA was detected with a TPR of 72 – 86% (Tables 16, 17).

Some analytes that were found to be not detectable are Ur, DMNB, Gly, MA, ER, PER, DP, and Cy. In the per-dataset classification experiments, the analytes AN, SA, Tol, and Ac were detected with reasonable detection performance. However, in the combined dataset results this was no longer the case.

**Question 3:** *Can the target analyte be detected even with different measuring conditions with varying ambient air, humidity, and*

*temperature?*

According to the results obtained, the combined dataset's results show that there are analytes such as H<sub>2</sub>O<sub>2</sub>, EG, HEX, and EPH that are detected with good TPR of > 70% (Tables 19, 21) despite the change in measuring conditions. Explosives such as TATP, HMTD, and MEKP despite having good results in their individual datasets, do not have good results when combined. But, the misclassifications for these analytes are mostly between each other, which indicates that when the datasets are combined, the ML model tends to generate discrimination between explosives of similar properties and other analytes. This is useful in practice because the objective of the detection is to detect explosives, and this is precisely what can be done with the current chip. Finally, there are analytes such as MEK, Ph, and BMK that achieve 90 – 100% (Tables 8, 10, 19, 21) TPR regardless of the type of experiment (per dataset or combined) we run. This indicates that the data from these analyte measurements are very distinctive and consistent, both of which result in perfect classification results.

**Question 4:** *Can we use a machine learning model that can learn the kinetic nature of the chemical reaction from the multidimensional time series and improve detection capabilities?*

The results obtained indicate that the kinetic nature of the data is useful in improving the detection capability of explosives and precursors as shown by previous work<sup>15</sup>. However, the use of deep learning models requires access to large amounts of data measurements. Thus, the use of such models is limited to the case where we have a sufficient number of data instances.

In the combined dataset time series classification results that were presented in Section 5.5, it was observed that explosives such as TATP, HMTD, and MEKP did not have good TPR. In order to explore these results further, a condensed confusion matrix is created for the same results and presented in Table 27. It can be observed that the majority of the explosives are classified correctly for both solids and liquids. The kinetic nature of the data in conjunction with the time series classifiers could indeed be used to improve the detection of these explosives further.

**Question 5:** *Which dyes on the sensor are the most useful in detecting specific analytes?*

The GPLASSO classifier was used to estimate the dyes that are most important from a classification viewpoint. It was observed that most of the analytes could be distinguished from AA, with a few of the dyes being the most informative. This information together with a parallel investigation of the authors on the analyte-dye interactions could be used for further sensor optimization.



Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	DMNB	53.00±0.20	47.00±0.20	37.00±0.26	63.00±0.26	54.00±0.16
	EG	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
Liquid	H <sub>2</sub> O <sub>2</sub>	90.00±0.20	10.00±0.20	0.00±0.00	100.00±0.00	93.00±0.13
	MEK	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	Ph	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00

Table 10 Classification results for Pc2 dataset obtained by using the classifier. The best classifier's results for solids (KNN) and liquid (RF) analytes are shown separately.

Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	Ur	75.00±0.27	25.00±0.27	50.00±0.35	50.00±0.35	65.00±0.17
	EG	90.00±0.12	10.00±0.12	0.00±0.00	100.00±0.00	94.00±0.07
Liquid	Gly	40.00±0.16	60.00±0.16	12.00±0.05	88.00±0.05	42.00±0.15
	H <sub>2</sub> O <sub>2</sub>	95.00±0.10	5.00±0.10	0.00±0.00	100.00±0.00	97.00±0.06
	Tol	82.00±0.15	18.00±0.15	13.00±0.07	87.00±0.07	68.00±0.17

Table 11 Classification results for Pc3 dataset obtained by using the classifier. The best classifier's results for solids (SVM) and liquid (KNN) analytes are shown separately.

	Gly	H <sub>2</sub> O <sub>2</sub>	To	EG	AA
Gly	9	0	2	2	7
H <sub>2</sub> O <sub>2</sub>	0	23	0	0	0
Tol	4	0	16	0	9
EG	0	0	0	20	0
AA	9	1	2	0	8

Table 12 Confusion matrix of liquid analytes classification of Pc3 dataset for KNN classifier.

## 7 Conclusions and future work

In this work, the data generated by colorimetric chip sensing experiments using the Crim-Track sniffer was used for the detection of explosives, drugs, and their precursor chemicals. The data were organized as datasets, where each dataset has its own measurement conditions and analytes. Several ML classifiers were used to perform a multi-class classification, and the results for the best classifier were presented in this work. The experimental results indicate that the explosive HMTD was detected with TPR 70 – 75% (Tables 4, 19, 21), TATP was detected with TPR 73 – 90% (Tables 4, 6), and MEKP was detected with TPR in the range of 60 – 82% (Tables 4, 19, 21). The precursors such as H<sub>2</sub>O<sub>2</sub>, HEX, MEK, BMK, EG, EPH, and the toxin Ph were detected with very high TPR. This indicates the power of the current chip in aiding in the detection of homemade explosives and their precursors. There are some analytes such as AN, COfb, DMNB, DP, ER, PER, Cy, Ac, Ur, To, SA, MA, and Gly were found to be not detectable with the current chip. Among these analytes are chemicals with low vapor pressure, different sublimation properties, and chem-

ical structures. It is concluded that the current chip is less sensitive to these analytes, and further addition of chemo-selective dyes into the chip would improve the sensitivity and broaden the range of detection.

Feature selection and selection of important dyes using the GPLASSO algorithm showed that certain dyes are more important in the detection of an analyte from ambient air. This information can be used for optimizing the sensor.

For future work, we consider the following ideas. In terms of sensor optimization, by introducing additional chemo-selective dyes, the robustness and sensitivity of the chip could be improved. In the context of machine learning models, an interesting study would be to better learn the signature of the analyte responses for the dye-analyte combinations that produce weak signals from the chip, and are therefore hard to detect. Such an approach would be beneficial in a scenario where the ambient noise in the measuring environment is high but the signal strength is weak.

## Author Contributions

Deena P. Francis: Conceptualisation, Data collection, Formal Analysis, Investigation, Writing - original draft, review and editing. Milan Laustsen: Conceptualisation, Chip fabrication, Data collection, Formal Analysis, Investigation. Eleftheria Dossi: Conceptualisation, Supervision, Investigation, Writing - original draft, review, and editing. Tuule Treiberg: Chip fabrication, Data collection, Investigation. Iona Hardy: Data collection, Writing. Shai Hvid Shiv: Data collection, Writing. Jesper Mogensen: Data collection, Supervision. Bo Svarrer Hansen: Conceptualisation, Supervision. Mogens H. Jakobsen: Conceptualisation, Supervision, Formal Analysis, Investigation, Writing review, and editing. Tommy S. Alstrøm: Conceptualisation, Supervision, Formal Analysis, Investigation, Writing, review, and editing.



Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	ER	0.00±0.00	100.00±0.00	0.00±0.00	100.00±0.00	-
	MA	0.00±0.00	100.00±0.00	0.00±0.00	100.00±0.00	-
	PER	100.00±0.00	0.00±0.00	100.00±0.00	0.00±0.00	56.90±0.06
Liquid	EG	90.00±0.20	10.00±0.20	0.00±0.00	100.00±0.00	93.00±0.13
	H <sub>2</sub> O <sub>2</sub>	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00

Table 13 Classification results for Pc4 dataset obtained by using the KNN classifier. The best classifier's results for solids (SVM) and liquid (KNN) analytes are shown separately.

	MA	ER	PER	AA
MA	0	0	0	0
ER	0	0	0	0
PER	6	8	16	12
AA	0	0	0	0

Table 14 Confusion matrix of solid analytes classification of Pc4 dataset with SVM classifier.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgments

We acknowledge the support from the Danish Defense, contract number 4600005161 for the PORDEX project. The authors thank the Cranfield Defence and Security School for funding the PhD of Iona Hardy. The authors would also like to thank Jens K Munk, DTU Nanotech for his contribution to the first fabrication of the Crim-Track chip.

## Notes and references

- 1 *The road to the Manchester Arena bombing*, <https://www.bbc.co.uk/news/uk-51908280>, Accessed: 2020-03-01.
- 2 N. C. Agency, (*No*, 2018, 58.
- 3 J. M. Johnston, *Institute for Biological Detection Systems, Auburn University*, 1999, **1**, 1–7.
- 4 C. Fredericx, F. Verheggen and E. Haubruge, *Biotechnologie, Agronomie, Société et Environnement*, 2011, **15**, 449–458.
- 5 R. Bogue, *Sensor Review*, 2015, **35**, 237–243.
- 6 G. A. Buttigieg, A. K. Knight, S. Denson, C. Pommier and M. B. Denton, *Forensic science international*, 2003, **135**, 53–59.
- 7 G. Eiceman and H. Schmidt, in *Aspects of Explosives Detection*, Elsevier, 2009, pp. 171–202.
- 8 M. J. Kangas, R. M. Burks, J. Atwater, R. M. Lukowicz, P. Williams and A. E. Holmes, *Critical reviews in analytical chemistry*, 2017, **47**, 138–153.
- 9 J. K. Munk, O. T. Buus, J. Larsen, E. Dossi, S. Tatlow, L. Lässig, L. Sandström and M. H. Jakobsen, *Optics and Photonics for Counterterrorism, Crime Fighting, and Defence XI; and Optical Materials and Biomaterials in Security and Defence Systems Technology XII*, 2015, pp. 40–44.
- 10 T. Alstrøm and J. Larsen, *Ph.D. thesis*, PhD thesis, 2013.
- 11 L. L. Mølgaard, O. T. Buus, J. Larsen, H. Babamoradi, I. L. Thygesen, M. Laustsen, J. K. Munk, E. Dossi, C. O'Keeffe, L. Lässig *et al.*, *Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XVIII*, 2017, pp. 38–45.
- 12 J. R. Carey, K. S. Suslick, K. I. Hulkower, J. A. Imlay, K. R. Imlay, C. K. Ingison, J. B. Ponder, A. Sen and A. E. Wittrig, *Journal of the American Chemical Society*, 2011, **133**, 7571–7576.
- 13 Q. Chen, H. Li, Q. Ouyang and J. Zhao, *Sensors and Actuators B: Chemical*, 2014, **205**, 1–8.
- 14 Z. Li and K. S. Suslick, *Acs Sensors*, 2016, **1**, 1330–1335.
- 15 D. P. Francis, M. Laustsen, H. Babamoradi, J. Mogensen, E. Dossi, M. H. Jakobsen and T. S. Alstrøm, *Artificial Intelligence and Machine Learning in Defense Applications III*, 2021, pp. 107–114.
- 16 M. C. Janzen, J. B. Ponder, D. P. Bailey, C. K. Ingison and K. S. Suslick, *Analytical chemistry*, 2006, **78**, 3591–3600.
- 17 Z. Li, J. R. Askim and K. S. Suslick, *Chemical reviews*, 2018, **119**, 231–292.
- 18 *1,2-Ethanediol*, <https://webbook.nist.gov/cgi/cbook.cgi?ID=107-21-1>, Accessed: 23-06-2022.
- 19 *Phenylacetone*, <https://pubchem.ncbi.nlm.nih.gov/compound/Phenylacetone>, Accessed: 24-06-2022.
- 20 *Ephedrine*, <https://pubchem.ncbi.nlm.nih.gov/compound/Ephedrine>, Accessed: 24-06-2022.
- 21 *H2O2*, <http://www.h2o2.com/technical-library/physical-chemical-properties/physical-properties/default.aspx?pid=25&name=Vapor-Pressures>, Accessed: 20-06-2022.
- 22 I. Stranski, G. Klipping, A. Bogenschuetz, H. Heinrich and H. Maennig, in *Advances in Catalysis*, Elsevier, 1957, vol. 9, pp. 406–414.
- 23 H. Östmark, S. Wallin and H. G. Ang, *Propellants, Explosives, Pyrotechnics*, 2012, **37**, 12–23.
- 24 M. J. Aernecke, T. Mendum, G. Geurtsen, A. Ostrinskaya and R. R. Kunz, *The Journal of Physical Chemistry A*, 2015, **119**, 11514–11522.
- 25 *MEKP*, <https://pubchem.ncbi.nlm.nih.gov/compound/2-Butanone-peroxide#section=Vapor-Density>, Accessed: 24-06-2022.

Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	AN	93.00±0.13	7.00±0.13	56.90±0.39	43.00±0.39	76.00±0.10
Liquid	EG	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	H <sub>2</sub> O <sub>2</sub>	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00

Table 15 Classification results for Pc5 dataset obtained by using the classifier. The best classifier's results for solids (RF) and liquid (KNN) analytes are shown separately.

Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	COfb	75.00±0.27	25.00±0.27	31.00±0.13	69.00±0.13	61.00±0.17
	EPH	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
Liquid	EG	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	SA	72.00±0.27	28.04±0.27	16.00±0.11	84.00±0.11	65.00±0.22
	H <sub>2</sub> O <sub>2</sub>	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00

Table 16 Classification results for D1 dataset obtained by using the RF classifier. The best classifier's results for solids (KNN) and liquid (RF) analytes are shown separately.

Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	COfb	72.00±0.18	28.00±0.00	35.00±0.06	65.00±0.06	68.00±0.09
Liquid	SA	86.00±0.20	14.00±0.20	40.00±0.22	60.00±0.22	74.00±0.13

Table 17 Classification results for D2 dataset obtained by using the RF classifier. The best classifier's results for solids (SVM) and liquid (KNN) analytes are shown separately.

	AA	COfb
AA	14	10
COfb	10	14

Table 18 Confusion matrix of solid analytes classification of D2 dataset with SVM classifier.

- 26 A. B. Dindal, M. V. Buchanan, R. A. Jenkins and C. K. Bayne, *Analyst*, 2000, **125**, 1393–1396.
- 27 *Acetone*, <https://pubchem.ncbi.nlm.nih.gov/compound/acetone#section=Vapor-Pressure>, Accessed: 1-7-2022.
- 28 *Safrole*, <https://pubchem.ncbi.nlm.nih.gov/compound/safrole#section=Vapor-Pressure>, Accessed: 1-7-2022.
- 29 *MEK*, <https://pubchem.ncbi.nlm.nih.gov/compound/Methyl-ethyl-ketone#section=Vapor-Pressure>, Accessed: 1-7-2022.
- 30 *Mannitol*, <https://cameochemicals.noaa.gov/chemical/20585>, Accessed: 1-7-2022.
- 31 M. Bilde, K. Barsanti, M. Booth, C. D. Cappa, N. M. Donahue, E. U. Emanuelsson, G. McFiggans, U. K. Krieger, C. Marcolli, D. Topping *et al.*, *Chemical reviews*, 2015, **115**, 4115–4156.
- 32 *Pentaerythritol*, <https://pubchem.ncbi.nlm.nih.gov/compound/pentaerythritol#section=Vapor-Pressure>, Accessed: 1-7-2022.
- 33 *Cyclohexanone*, ["https://webwisser.nlm.nih.gov/WebWISER/substance?substanceId=513&identifier=Cyclohexanone&identifierType=name&catId=48"](https://webwisser.nlm.nih.gov/WebWISER/substance?substanceId=513&identifier=Cyclohexanone&identifierType=name&catId=48), Accessed: 1-7-2022.
- 34 *Dibutyl phthalate*, ["https://pubchem.ncbi.nlm.nih.gov/compound/Dibutyl-phthalate#section=InChI"](https://pubchem.ncbi.nlm.nih.gov/compound/Dibutyl-phthalate#section=InChI), Accessed: 1-7-2022.
- 35 *Phenol*, <https://pubchem.ncbi.nlm.nih.gov/compound/phenol#section=Vapor-Pressure>, Accessed: 1-7-2022.
- 36 J. Brandner, N. M. Junk, J. Lawrence and J. Robins, *Journal of Chemical and Engineering Data*, 1962, **7**, 227–228.
- 37 K. R. M. F. Vojtěch Štejfa, Kateřina Kadlecová, *Chemical Thermodynamics and Thermal Analysis*, 2021, **3**, 100020.
- 38 *Toluene*, <https://pubchem.ncbi.nlm.nih.gov/compound/toluene#section=Vapor-Pressure>, Accessed: 6-7-2022.
- 39 *Urea*, <https://pubchem.ncbi.nlm.nih.gov/compound/urea#section=Vapor-Pressure>, Accessed: 6-7-2022.
- 40 *Glycerin*, <https://pubchem.ncbi.nlm.nih.gov/compound/glycerol#section=Vapor-Density>, Accessed: 6-7-2022.
- 41 *Gesim*, <https://gesim-bioinstruments-microfluidics.com/microarray-printer>, Accessed: 16-9-2022.
- 42 *polypropylene boxes*, <https://www.ikea.com/dk/da/p/samla-boks-transparent-80102976>, Accessed: 16-9-2022.
- 43 R. O. Duda, P. E. Hart and D. G. Stork, *Pattern classification*

Phase	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	AN	0.00±0.00	100.00±0.00	0.26±0.52	99.74±0.52	-
	COfb	22.50±20.00	77.50±20.00	1.41±2.18	98.59±2.18	-
	MA	0.00±0.00	100.00±0.00	0.00±0.00	100.00±0.00	-
	DMNB	24.00±15.94	76.00±15.94	2.13±4.27	97.87±4.27	-
	EPH	90.00±20.00	10.00±20.00	0.26±0.52	99.74±0.52	90.48±13.13
	ER	0.00±0.00	100.00±0.00	0.00±0.00	100.00±0.00	-
	HEX	70.00±40.00	30.00±40.00	0.00±0.00	100.00±0.00	-
	HMTD	75.60±30.11	24.40±30.11	9.05±3.61	90.95±3.61	65.06±17.78
	PER	0.00±0.00	100.00±0.00	0.00±0.00	100.00±0.00	-
	TATP	48.86±14.55	51.14±14.55	15.95±6.05	84.05±6.05	43.17±5.44
Liquid	Ur	0.00±0.00	100.00±0.00	0.26±0.53	99.74±0.53	-
	Ac	10.00±20.00	90.00±20.00	1.15±2.31	98.85±2.31	-
	BMK	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	Cy	0.00±0.00	100.00±0.00	0.19±0.38	99.81±0.38	-
	DP	0.00±0.00	100.00±0.00	0.00±0.00	100.00±0.00	-
	EG	84.27±15.51	15.73±15.51	2.93±2.26	97.07±2.26	86.26±11.32
	Gly	0.00±0.00	100.00±0.00	0.20±0.39	99.80±0.39	-
	H <sub>2</sub> O <sub>2</sub>	95.84±6.68	4.16±6.68	1.55±1.91	98.45±1.91	95.79±5.83
	H <sub>2</sub> O	80.00±40.00	20.00±40.00	0.00±0.00	100.00±0.00	-
	MEK	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	MEKP	82.38±10.61	17.62±10.61	0.80±0.75	99.20±0.75	84.80±10.30
	Ph	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	SA	26.67±8.58	73.33±8.58	2.23±2.44	97.77±2.44	34.77±10.16
To	6.67±13.33	93.33±13.33	0.00±0.00	100.00±0.00	-	

Table 19 Classification results for the combined dataset consisting of all analytes obtained by using the RF classifier. The best classifier's results for solids and liquid analytes are shown separately.

	MA	HMTD	AA	TATP	PER	ER	EPH	COfb	DMNB	AN	Ur	HEX
MA	0	0	0	0	0	0	0	0	0	0	0	0
HMTD	0	64	1	27	0	0	2	0	0	2	2	4
AA	24	2	63	17	23	20	0	30	16	5	12	0
TATP	0	15	14	43	1	2	0	11	4	9	9	0
PER	0	0	0	0	0	0	0	0	0	0	0	0
ER	0	0	0	0	0	0	0	0	0	0	0	0
EPH	0	0	0	0	0	0	18	0	0	0	0	1
COfb	0	0	5	0	0	0	0	9	0	0	1	0
DMNB	0	0	8	0	0	0	0	0	6	0	0	0
AN	0	0	0	0	0	0	0	1	0	0	0	0
Ur	0	0	1	0	0	0	0	0	0	0	0	0
HEX	0	0	0	0	0	0	0	0	0	0	0	6

Table 20 Confusion matrix of solid analytes classification of the combined dataset with RF classifier.

and scene analysis, Wiley New York, 1973, vol. 3.  
44 S. H. Lim, J. W. Kemling, L. Feng and K. S. Suslick, *Analyst*, 2009, **134**, 2453–2457.  
45 L. Van der Maaten and G. Hinton, *Journal of machine learning*

research, 2008, **9**, 2579–2605.  
46 G. E. Hinton and S. Roweis, *Advances in neural information processing systems*, 2002, **15**, 857 – 864.

	Analyte name	TPR	FNR	FPR	TNR	F1 SCORE
Solid	AN	0.00±0.00	100.00±0.00	0.26±0.52	99.74±0.52	-
	COfb	37.22±28.75	62.78±28.75	2.25±1.69	97.75±1.69	-
	DMNB	77.00±20.40	23.00±20.40	2.13±3.64	97.87±3.64	74.48±16.41
	EPH	93.33±13.33	6.67±13.33	0.00±0.00	100.00±0.00	96.00±8.00
	ER	0.00±0.00	100.00±0.00	0.00±0.00	100.00±0.00	-
	HEX	80.00±40.00	20.00±40.00	0.26±0.51	99.74±0.51	-
	HMTD	70.66±22.50	29.34±22.50	9.07±3.92	90.93±3.92	63.83±13.38
	MA	0.00±0.00	100.00±0.00	1.32±2.04	98.68±2.04	-
	PER	0.00±0.00	100.00±0.00	0.26±0.53	99.74±0.53	-
	TATP	47.14±10.90	52.86±10.90	17.51±4.02	82.49±4.02	41.31±5.98
	Ur	0.00±0.00	100.00±0.00	0.00±0.00	100.00±0.00	-
Liquid	Ac	70.00±24.49	30.00±24.49	4.01±5.92	95.99±5.92	53.89±31.99
	BMK	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	Cy	26.67±38.87	73.33±38.87	0.58±0.77	99.42±0.77	-
	DP	0.00±0.00	100.00±0.00	0.00±0.00	100.00±0.00	-
	EG	77.57±24.42	22.43±24.42	3.66±2.18	96.34±2.18	79.74±15.98
	Gly	0.00±0.00	100.00±0.00	0.19±0.38	99.81±0.38	-
	H <sub>2</sub> O <sub>2</sub>	86.87±16.07	13.13±16.07	1.03±1.51	98.97±1.51	90.80±10.31
	H <sub>2</sub> O	90.00±20.00	10.00±20.00	0.19±0.38	99.81±0.38	89.33±13.73
	MEK	100.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	100.00±0.00
	MEKP	60.48±14.88	39.52±14.88	3.63±3.42	96.37±3.42	58.51±11.88
	Ph	90.00±20.00	10.00±20.00	0.00±0.00	100.00±0.00	93.33±13.33
	SA	14.44±8.81	85.56±8.81	3.48±2.40	96.52±2.40	-
	Tol	0.00±0.00	100.00±0.00	0.19±0.39	99.81±0.39	-

Table 21 Time series classification results for combined dataset obtained by using the CNN classifier. The best classifier's results for solids and liquid analytes are shown separately.

- 47 E. Fix and J. L. Hodges, *International Statistical Review/Revue Internationale de Statistique*, 1989, **57**, 238–247.
- 48 L. Breiman, *Machine learning*, 2001, **45**, 5–32.
- 49 C. Cortes and V. Vapnik, *Machine learning*, 1995, **20**, 273–297.
- 50 D. Pregibon, *The annals of statistics*, 1981, **9**, 705–724.
- 51 M. Yuan and Y. Lin, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2006, **68**, 49–67.
- 52 N. Simon, J. Friedman, T. Hastie and R. Tibshirani, *Journal of computational and graphical statistics*, 2013, **22**, 231–245.
- 53 M. Vincent and N. R. Hansen, *Computational Statistics & Data Analysis*, 2014, **71**, 771–786.
- 54 C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, MIT press Cambridge, MA, 2006, vol. 2.
- 55 Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, *Neural computation*, 1989, **1**, 541–551.
- 56 Z. Wang, W. Yan and T. Oates, 2017 International joint conference on neural networks (IJCNN), 2017, pp. 1578–1585.
- 57 C. Lea, M. D. Flynn, R. Vidal, A. Reiter and G. D. Hager, proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 156–165.
- 58 S. Hochreiter and J. Schmidhuber, *Neural computation*, 1997, **9**, 1735–1780.
- 59 D. J. Berndt and J. Clifford, KDD workshop, 1994, pp. 359–370.
- 60 M. Cuturi, J.-P. Vert, O. Birkenes and T. Matsui, 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, 2007, pp. II–413.
- 61 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *Journal of Machine Learning Research*, 2011, **12**, 2825–2830.
- 62 Y. M. Moe, *Efficient Group Lasso in Python*, 2019, <https://>

Analyte	Varying dyes
Ac	4, 5, 8, 10, 13, 14, 15, 16, 17, 19, 20, 22, 25, 26
AN	10, 13, 14, 16, 18, 20, 21
BMK	4, 5, 8, 10, 13, 14, 15, 16, 17, 19, 20, 22, 24
COfb	9, 13, 16, 18, 20, 21
Cy	13, 14, 16, 18, 20
DMNB	6, 13, 14, 16, 17, 18, 19, 20, 22, 25
DP	13, 14, 16, 18, 20, 21
EG	4, 5, 6, 8, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27
EPH	6, 7, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 27
ER	13, 16
Gly	13, 14, 16, 18, 21
H <sub>2</sub> O	4, 5, 14, 15, 16, 17, 18, 19, 21, 22, 25
H <sub>2</sub> O <sub>2</sub>	4, 5, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27
HEX	5, 6, 8, 14, 15, 16, 17, 18, 19, 20, 21, 22, 27
HMTD	4, 5, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 26
MA	13, 14, 16
MEK	4, 5, 7, 13, 16, 20, 25
MEKP	2, 4, 5, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22
PER	13, 14, 16
Ph	9, 11, 13, 14, 16, 18, 20, 21, 23, 25, 26, 28
SA	13, 16, 18, 21
TATP	4, 5, 8, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26
To	10, 16, 18, 20, 21
Ur	14, 16, 20, 21

Table 22 List of dyes that vary when exposed to an analyte. A simple linear regression model is fitted to the data and the dyes for which we get a fit with  $R^2 \geq 0.8$  are shown.

Analyte	Important Dyes	F1 SCORE
Ac	15, 16	100
AN	20, 21, 3	83.33
BMK	15, 16	100
COfb	20, 21	100
Cy	13, 15, 20	100
DMNB	20, 15	100
DP	3, 20	100
EPH	15	87.50
EG	21	97.22
Gly	12, 15, 17	91.67
H <sub>2</sub> O	7	100
H <sub>2</sub> O <sub>2</sub>	20	100
HEX	15	100
HMTD	13	71.11
MA	20	100
MEK	12	100
MEKP	4, 8, 12, 13	86.36
PER	20	100
Ph	12	100
SA	14, 20, 21	100
TATP	3, 6, 7, 16, 17	77.14
Tol	13, 15, 20	100
Ur	all dyes	76.92

Table 23 The results of feature selection using group lasso for the binary classification task of each analyte vs ambient air.

group-lasso.readthedocs.io/en/latest/  
63 M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga,

S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, *TensorFlow: Large-Scale*

	AA	E
AA	4	14
E	32	144

	AA	E	EP
AA	23	6	8
E	11	34	3
EP	2	0	67

Table 24 Confusion matrix for the Ex1 dataset in a condensed form. Left: Solid analytes, Right: Liquid analytes.

	AA	E	EP
AA	0	1	0
E	10	10	0
EP	0	0	11

	AA	EP
AA	7	0
EP	0	9

Table 25 Confusion matrix for the Ex2 dataset in a condensed form. Left: Solid analytes, Right: Liquid analytes.

	AA	D	DP	E	EA	EP
AA	63	30	0	24	16	59
D	5	9	0	0	0	1
DP	0	0	18	0	0	1
E	15	12	2	160	4	16
EA	8	0	0	0	6	0
EP	1	0	0	0	0	6

	AA	DP	E	EP
AA	78	35	4	39
DP	7	24	1	1
E	0	0	33	1
EP	6	2	0	199

Table 26 Confusion matrix for the combined diff dataset in a condensed form. Left: Solid analytes, Right: Liquid analytes.

	AA	D	DP	E	EA	EP
AA	60	28	0	26	5	49
D	3	15	0	4	0	2
DP	0	0	18	0	0	0
E	22	7	0	152	0	23
EA	6	0	0	1	19	0
EP	1	1	2	1	2	9

	AA	DP	E	EP
AA	73	33	4	25
DP	6	20	4	6
E	4	4	23	7
EP	6	2	0	202

Table 27 Confusion matrix for the combined time series dataset in a condensed form. Left: Solid analytes, Right: Liquid analytes.

- Machine Learning on Heterogeneous Systems*, 2015, <https://www.tensorflow.org/>, Software available from tensorflow.org.
- 64 R. Tavenard, J. Faouzi, G. Vandewiele, F. Divo, G. Androz, C. Holtz, M. Payne, R. Yurchak, M. Rußwurm, K. Kolar and E. Woods, *Journal of Machine Learning Research*, 2020, **21**, 1–6.
- 65 I. Guyon and A. Elisseeff, *Journal of machine learning research*, 2003, **3**, 1157–1182.
- 66 J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang and H. Liu, *ACM computing surveys (CSUR)*, 2017, **50**, 1–45.
- 67 L. Meier, S. Van De Geer and P. Bühlmann, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2008, **70**, 53–71.