DATA REPRESENTATION AND FEATURE SELECTION FOR COLORIMETRIC SENSOR ARRAYS USED AS EXPLOSIVES DETECTORS

Tommy S. Alstrøm[a]*, Jan Larsen*[a]*, Natalie V. Kostesha*[b]*, Mogens H. Jakobsen*[b] *and Anja Boisen*[b]

[a]Dept. of Informatics and Mathematical Modeling, Technical University of Denmark
Richard Petersens Plads 321, 2800 Kgs. Lyngby, Denmark
{tsal,jl}@imm.dtu.dk

[b]Dept. of Micro- and Nanotechnology, Technical University of Denmark
Ørsteds Plads 345 East, DK-2800, Kgs. Lyngby, Denmark
{natalie.kostesha,mogens.jakobsen,anja.boisen}@nanotech.dtu.dk

ABSTRACT

Within the framework of the strategic research project *Xsense* at the Technical University of Denmark, we are developing a colorimetric sensor array which can be useful for detection of explosives like DNT, TNT, HMX, RDX and TATP and identification of volatile organic compounds in the presence of water vapor in air. In order to analyze colorimetric sensors with statistical methods, the sensory output must be put into numerical form suitable for analysis. We present new ways of extracting features from a colorimetric sensor and determine the quality and robustness of these features using machine learning classifiers. Sensors, and in particular explosive sensors, must not only be able to classify explosives, they must also be able to measure the certainty of the classifier regarding the decision it has made. This means there is a need for classifiers that not only give a decision, but also give a posterior probability about the decision. We will compare K–nearest neighbor, artificial neural networks and sparse logistic regression for colorimetric sensor data analysis. Using the sparse solutions we perform feature selection and feature ranking and compare to Gram–Schmidt orthogonalization.

Index Terms— artificial neural networks (ANN), chemo–selective compounds, classification, colorimetric sensor array, DNT, explosives detection, feature ranking, Gram–Schmidt orthogonalization, K–nearest neighbor (KNN), sparse logistic regression (SLR), TNT

1 Introduction

Over the past decade, explosives have been a preferred tool for terrorists, yet there is no satisfactory mobile and portable solution to detect explosives. To detect a variety of military and industrial explosives easily, new technologies must be developed. There are several application areas for explosives sensors, such as anti-terrorism (screening luggage and mail packages, checking suspects and mass transit systems), demining and environmental monitoring of hazardous compounds.

Sensors must not only easily detect a variety of hidden explosives but they must also be able to detect illegal chemicals and products of the explosives industry. A further requirement is that the sensing device should be portable, rapid, highly sensitive, specific (minimize false alarms), and inexpensive [1].

Over the past years a number of detecting methods have been developed and successfully applied in explosives detectors. These include, but are not limited to, gas chromatography, Raman spectrometry, mass spectrometry, ion mobility spectrometry and colorimetric sensors. Suslick *et al.* described the application of the colorimetric sensor array for detecting volatile organic compounds in the gas phase [2, 3] as well as for identifying different organic compounds in the liquid phase [4, 5]. In our project we develop a colorimetric sensor array that can be useful in detecting and identifying explosives like TNT, DNT, HMX, RDX and TATP [6, 7]. The colorimetric sensor is a fascinating technique for detecting different chemical compounds belonging to various classes, like amines, cyanides, alcohols, arenes, ketones, aldehydes and acids in the parts-per-million (ppm) and parts-per-billion (ppb) ranges [3, 8, 9]. In our research we use a completely different class of chemo–selective compound which has already shown excellent results for detecting TNT. This type of colorimetric sensor could be successfully applied in homeland security and defense [10, 11].

A colorimetric sensor array consists of a number of chemo–selective compounds of various colors that will undergo a color change when subjected to an environment or a target substance, hereafter denoted an *analyte*. These chemo–selective compounds, which are typically called *dyes* are

We acknowledge the support from the Danish Agency for Science and Technology's, Program Commission on Nanoscience Biotechnology and IT (NABIIT). Case number: 2106-07-0031 - Miniaturized sensors for explosives detection in air.

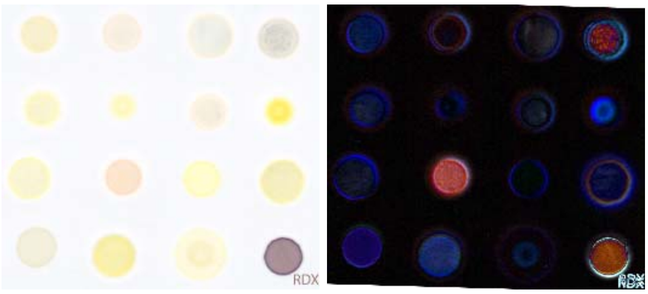


Fig. 1. An example of a colorimetric sensor exposed to the explosive analyte RDX. A: sensor before exposure. B: enhanced difference image.

digitalized using a flatbed scanner. One dye consists of several hundred pixels, but a dye is considered to have only one color which is typically found by calculating the mean pixel value. We investigate how feature extraction using the mean pixel value compares to alternatives.

Having acquired the sensor response digitally enables the application of signal processing and statistic methods such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) [12–14]. In the domains where colorimetric sensors have been investigated, HCA shows high accuracy and low false alarm rate. The closely related K-Nearest Neighbor (KNN) classifier [15] with $K = 1$ set to one has evolved as the de–facto method.

Our requirements for a classifier go beyond what KNN offers. As we are detecting hazardous compounds, we require the classifier to offer posterior probabilities and not only decisions. Further we seek to qualify which compounds in the colorimetric sensor are important, and which are less important. This knowledge enables the ability to either reduce the size of the sensor or replace less sensitive and unimportant compounds with more selective and responsive compounds. Various feature selection strategies can be employed to select the dyes but so far none have been applied to colorimetric sensor arrays. Our preference would be to use a sparse classifier thus making the feature selection an inherent part of the classification. Our main goal is not to find the classifier that has the best classification rate but to identify a classifier with these attributes and at the same time delivers comparable or better performance compared to KNN. We will consider two classifiers: sparse logistic regression (SLR) [16] which is a linear classifier that is extended to model posterior probabilities and implement sparsity, and artificial neural networks (ANN) [17] which is a proven non–linear classifier. Our primary motivation to include a non–linear classifier is to investigate if non–linear models are better suited for colorimetric sensor arrays.

Chemical family	A	B	C
Acids	32		45
Alcohols	26		27
Aldehydes	6		
Amines			42
Arenes	10		13
Drugs		6	
Environment	12	8	28
Explosives	20	8	56
Inorganic Explosives			14
Ketones	7		13
Thiols			14
Total measurements:	129	22	253

Table 1. The different chemical families and how many measurements were applied to the sensors, i.e. sensor B has been measured 22 times in total, where 6 of the measurements were drugs. Each family comprises several compounds.

2 Colorimetric sensors

We have operated with three different sensors: sensor A comprised 15 dyes [6]; sensor B equal to sensor A with one added dye [7]; sensor C which is a further extension adding 15 dyes to sensor B (unpublished results). Fig. 1A shows sensor B when exposed to the explosive analyte RDX. The sensors have been exposed to analytes belonging to the various chemical families (Table 1). For a more elaborate description of sensor fabrication, dyes and target analytes we refer to our earlier published work [6, 7].

2.1 Data acquisition

Images of the sensor were scanned using an ordinary flatbed scanner immediately after immobilization of dyes and then again after exposure of target analytes. The images were encoded in a lossless format using the red-green-blue (RGB) color scheme with 8 bits per color. The images were used to generate color difference images by pixel subtraction (Fig. 1). To align the two images a cost function that measures the L1 norm error per pixel is minimized:

$$\min_{\boldsymbol{t},\theta} \|\mathbf{I}_{\text{before}} - g(\mathbf{I}_{\text{after}}, \boldsymbol{t}, \theta)\|_1 / N_{pixels} \tag{1}$$

where I is an image matrix, $\mathbf{t}$ is an (x, y) translation, θ is a rotation and $g(\cdot)$ is the function that implements the transformation: first rotating the images using nearest neighbor interpolation, then translating (x, y)–pixels according to $\boldsymbol{t}$. The parameters $(\boldsymbol{t}, \theta)$ are initialized to zero.

The L1 norm error can be interpreted as the amount of *color* per pixel where a fully black pixel is fixed at a value of zero. At perfect image alignment, all the background pixels will be black and all dye pixels will have the weakest possible color, hence, blackness per pixel should be maximized. Dye

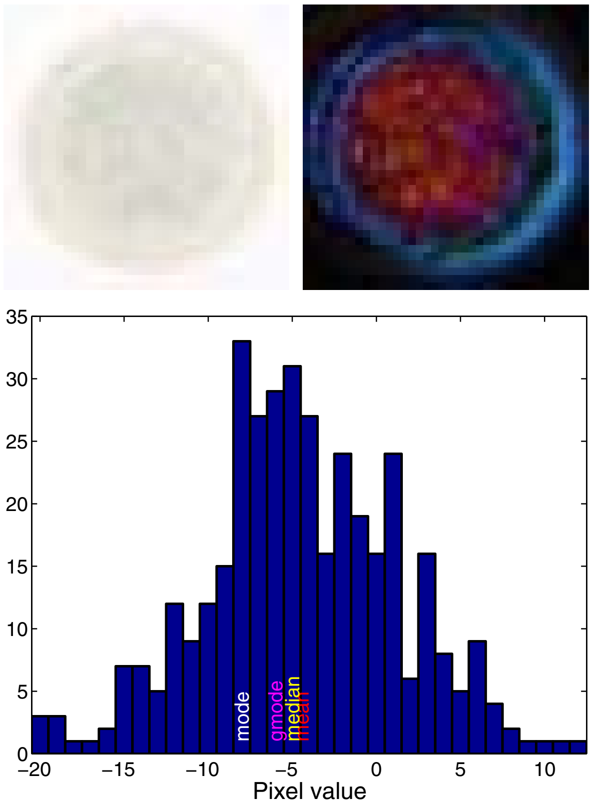


Fig. 2. An example of colorimetric sensor B exposed to the explosive analyte RDX. A: the sensor before exposure. B: the enhanced difference image. C: histogram of the green value and the value of the statistics.

localization and generation of colorimetric difference maps are described in detail in [18].

2.2 Data extraction

Once the images are digitalized feature extraction is employed typically using the mean pixel value. In order for the mean to be a robust measure of color change, the pixels of a dye have to be normally distributed (or at least have a symmetric distribution with one mode) and relatively free from outliers. As can be seen in Fig. 2 this is may not always be the case. Chemically we know that a dye should only have one color, as the dye is homogeneous and exposed to a homogeneous vapor. However, noise is induced from: the scanner, the drying process of the dye, external light, and roughness of surface. Some of these effects can be handled easily. The drying of the dyes often results in a ring near the perimeter (the coffee stain effect) and this area of the dye is unreliable. To accommodate for this effect, a smaller area of a dye is used for feature extraction, corresponding to 2/3 of the dye radius. To handle the other noise effects that cause pixel outliers, the median or mode can be used as both statistics are more robust to outliers. Just as the case with the mean, the median requires a symmetric distribution with one mode. On the other hand the mode does not require the distribution to be symmetric and could potentially be more robust. However, these three statistics have the weakness that they consider the RGB colors as independent since an RGB color value is a 3 dimension vector. The multinomial mode does not have this weakness and will find the most occurring color in the dye hence we expect the multinomial mode to be the best representation. The entire data acquisition and extraction pipeline is described in detail in [18].

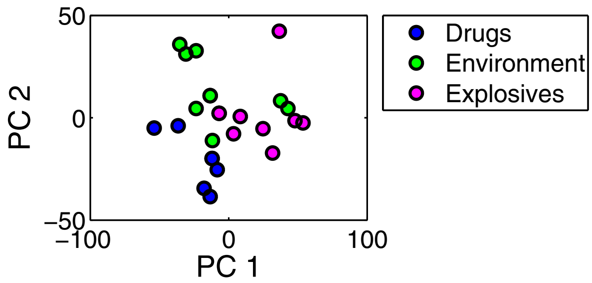


Fig. 3. PCA plot from sensor B. The PCA plot implies that the sensor should be able to separate well although there is an explosive outlier.

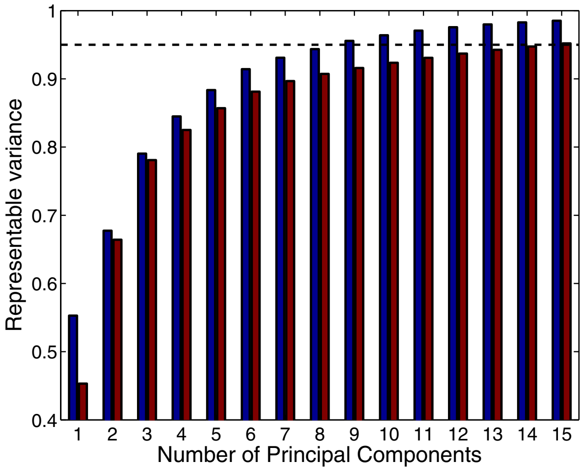


Fig. 4. The colorimetric sensor is a high dimensional sensor. For sensor A (blue) it requires 9 dimensions to represent 95% of the variance, whereas for sensor C (red) it requires 15 dimensions.

2.3 Data visualization

Data from colorimetric sensors can be visualized using principal component analysis (PCA) with a certain degree of success. However, once the sensor has been applied to many different classes, this kind of data visualization is of limited value. Fig. 3 shows how PCA can be used to plot data for the sensor B case. However for sensors A and C the data collected is too high dimensional for a PCA plot to show the entire structure of the data [6]. Fig. 4 shows how the variance in the data is distributed among the different dimensions. While the figure implies that sensor C is a higher dimensional sensor, the plot does not merit conclusions about the true dimensionality of the sensor. The dyes are highly correlated, especially the red-green-blue dimensions within each dye, but even more so there may be a lot of uncorrelated noise. PCA requires one dimension per uncorrelated signal channel and if the noise channels are sufficiently strong more dimensions will be needed to represent the data accurately. Observe that for three dimensions the sensors are almost equal, even though sensor C has twice as many dyes as sensor A.

3 Methods

Each classification method is evaluated using 10–fold double cross validation (CV) partitioning used is a stratified approach. The partitioning of the outer fold that is used to estimate the test error remains fixed, while the partitions that are used for model selection are regenerated in each run.

3.1 K–nearest neighbor

Despite its simplicity the K–nearest neighbor is an effective classification technique [15] which works as follows: when testing an unknown data point, the Euclidean distances for all known points are calculated. The classes of the closest K points are then identified and the unknown point is classified using majority voting of these known points. In the event of a tie, the algorithm uses the nearest neighbor among the tied classes to break the tie selecting the closest point as the class. All possible values of K are probed during model selection.

3.2 Sparse multinomial logistic regression

The multinomial logistic regression model offers a posterior probability of a class given a measurement. The model is written as

$$p(\mathcal{C}_k|\mathbf{x}) = \frac{\exp(\mathbf{w}_k^{\mathrm{T}}\mathbf{x})}{\sum_j \exp(\mathbf{w}_j^{\mathrm{T}}\mathbf{x})} \tag{2}$$

where $p(\mathcal{C}_k|\mathbf{x})$ is the probability of class $\mathcal{C}_k$ given a data point $\mathbf{x}$. To promote a sparse solution and to handle over-fit, we use L1 regularization. This is achieved by adding the term $\lambda\|\mathbf{w}\|_1$ to the cost function where λ is the model selection parameter. The cost function is minimized using the *Projection L1* method described by Schmidt *et al.* [16]. To further promote sparse solutions, the weights are initialized to zero. The optimal λ is searched for in the interval [0; 10].

3.3 Artificial neural networks

The artificial neural network classifier used is a two–layered feed-forward with the hidden units using tangent hyperbolic sigmoidal function as the transfer function. A quadratic cost function augmented with outlier detection and weight decay is used. $S(\mathbf{w})$ has two hyper–parameters; the regularization parameter α and outlier probability β:

$$S(\mathbf{w}) = E_D(\mathbf{w}, \beta) + \alpha E_W(\mathbf{w}) \tag{3}$$

where $E_D(\mathbf{w}, \beta)$ is the cross-entropy error function and $E_W(\mathbf{w})$ is a regularization term. The hyper-parameter α is initialized to the number of inputs, i.e., the more inputs, the stronger regularization needed. The hyper-parameter β is initialized to zero as a priori there are no known outliers. The network weights are initialized using a zero mean Gaussian with variance equal to $1/\alpha$. Each network training cycle is repeated ten times and the network with the lowest training error $E_D(\mathbf{w}, \beta)$ is selected for classification. The outputs are converted to probabilities using the soft–max function. The network training and cost function are described in detail by Sigurdsson *et al.* [17].

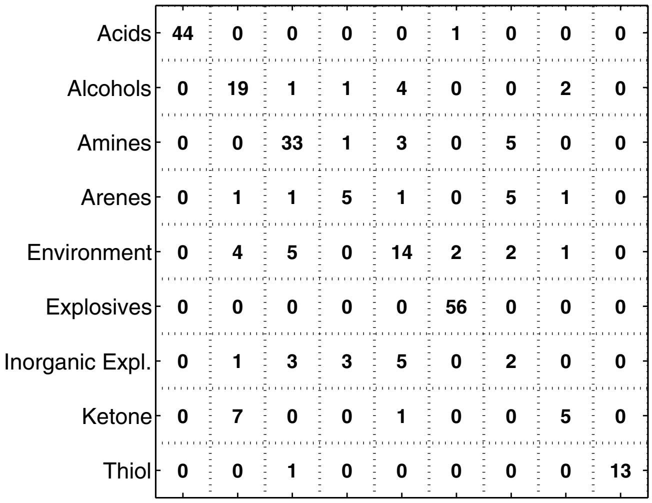
Method	Median	Mode	MMode	Mean	Comb.
A-1NN	5.40	**5.83**	5.27	5.33	5.40
A-KNN	5.77	**6.20**	5.40	5.15	5.52
A-SLR	5.71	5.52	**6.08**	5.71	5.71
A-ANN	5.52	5.40	5.52	**5.71**	5.64
B-1NN	1.77	1.77	**2.18**	1.91	1.77
B-KNN	1.64	1.50	**1.91**	1.64	1.64
B-SLR	1.36	1.36	**1.91**	1.36	1.36
B-ANN	1.36	1.09	**1.50**	1.36	1.36
C-1NN	6.05	6.12	**6.15**	6.05	6.05
C-KNN	5.91	6.05	5.94	**6.12**	6.05
C-SLR	**6.79**	6.69	6.65	6.72	**6.79**
C-ANN	6.40	6.37	6.19	**6.47**	NaN
Total	1	2	**6**	3	1

Table 2. Summary of how well the features and classifiers perform compared to random guessing, with the best performers highlighted.

4 Results and discussion

Since the sensors have a different amount of classes, we find it better to assess the quality of the classifiers by using the classification rate relative to random guessing as the classification rate alone can be misleading. For example a classification rate of 0.33% for sensor B would only be as good as random guessing while for sensor A and C it would be better than random guessing. The ratio is calculated as $(N_{correct} \cdot N_C)/N$, where N_C is the number of points correctly classified, N is the total number of points and N_C is the number of classes.

Table 2 shows the best statistic is the multivariate mode which is the best performer in six out of twelve cases. However, all the other features are also represented at least once as the top performer, so the results indicate that in order to build the most accurate classifier one must extract all of the



Acids	44	0	0	0	0	1	0	0	0
Alcohols	0	19	1	1	4	0	0	2	0
Amines	0	0	33	1	3	0	5	0	0
Arenes	0	1	1	5	1	0	5	1	0
Environment	0	4	5	0	14	2	2	1	0
Explosives	0	0	0	0	0	56	0	0	0
Inorganic Expl.	0	1	3	3	5	0	2	0	0
Ketone	0	7	0	0	1	0	0	5	0
Thiol	0	0	1	0	0	0	0	0	13

Fig. 5. Confusion matrix for C-SLR-Median where rows indicate the true class, columns indicate the predicted class and the number indicate the counts [19]. The sensor has a high classification rate for explosives with zero false negatives (all 56 measurements are classified correctly) but three false positives. .

proposed features from the colorimetric sensor and then let the feature selection be part of the model selection process.

The considered classification methods all perform similarly. As expected the simplest method, 1NN, is the best performer on the smallest dataset B. For sensor A, the KNN method is the best performer and on C the SLR is the best performer.

Sensor C is the richest sensor, with respect to data points, dyes and classes, and we will now discuss the results from this sensor in more detail. The best performing classifier for sensor C is SLR using median as feature extraction statistic. Fig. 5 shows the confusion matrix for this situation and clearly shows that the sensor is very good for detecting especially acids and explosives, although some chemicals, such as arenes, pose a greater challenge. Having established that SLR is the classifier with the highest accuracy, we will now investigate how well SLR detects informative dyes by scrutinizing the sparse solutions.

A colorimetric sensor array is likely to have redundant features, as some dyes will react similarly, and it is not clear for the experimenter which dyes should be used in the next iteration of the sensor. We want to explore how well SLR identifies important dyes compared to the simple forward selection method based on the Gram–Schmidt orthogonalization [20].

For the feature selection process we convert the classification problem into a binary problem; explosives versus non-explosives (inorganic explosives will be part of the non-explosives group, as these explosives are not part of our detection focus). For each of the four feature extraction statistics, we will train ten models using the same data partitioning that was used in the multinomial classification case, hence forty models are trained in total. The sparsity parameter λ will start from zero and be incremented slowly, removing one feature at a time from the model. Each dye consist of three values (the RGB values) so in order to remove a dye all three features belonging to a dye have to be eliminated from the model. A dye is considered as completely removed once the weights for all the values are below a threshold which we set to 10^{-3}. Fig. 6 shows how the dyes for sensor C rank according to their mean ranking.

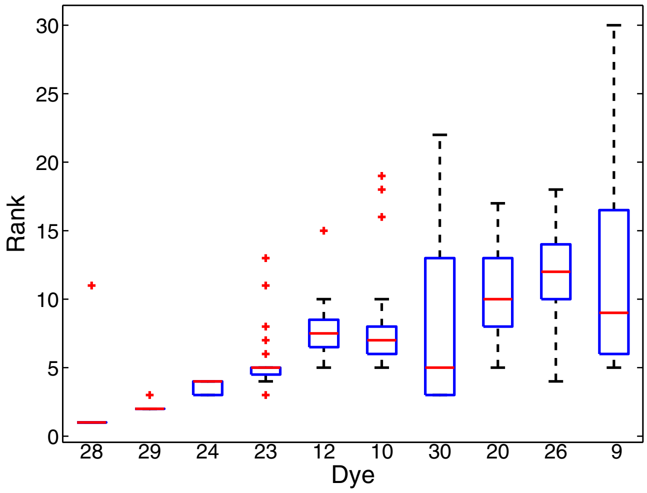


Fig. 6. Dyes ranked according to their mean rank for sensor C. The dye significance is listed from left to right where the more significant dye is the left.

The top ten ranked dyes using the median as the feature extraction statistic are used to detect explosives. Fig. 7 shows how well suited the dyes are for classification. Using three dyes the classification error is below 0.01% and after this adding more dyes only decreases the error marginally. Comparing the dyes identified by SLR to GS, the SLR dyes are clearly more suited, however if whitening is used GS show similar performance as SLR.

5 Conclusion

We have tried five different feature sets of four different classifiers on three different colorimetric sensor arrays.The logistic regression method demonstrated equal classification ability compared to KNN and due to the added perks in terms of sparsity and probabilistic decisions, SLR is preferable to KNN. The results do not merit the use of a non–linear method such ANN. This is likely because there is not enough training data and since the experimental process of colorimetric sensors is time consuming, methods that work with fewer points are more appealing.

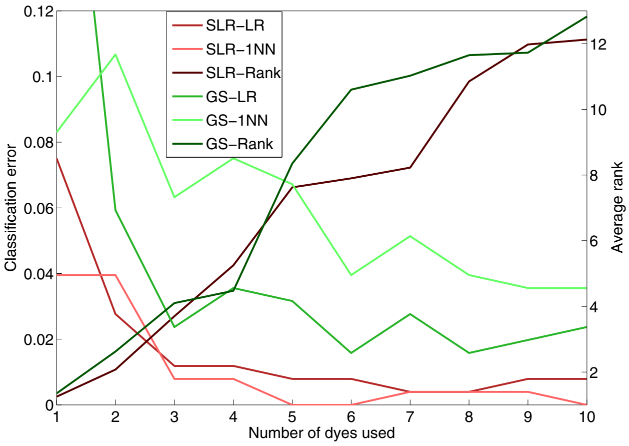


Fig. 7. Classification error with respect to number of dyes used, based on sensor C–Median. The classification error decreases to a certain point as dyes are added. SLR-LR is logistic regression classifier applied on the dyes identified by SLR. SLR-1NN used 1NN as classifier instead. SLR-Rank indicates the mean feature rank of dye i as shown on Fig. 6. Similarly GS-X uses the features found by the GS based feature selection method.

The classification results allow us to make a recommendation about which statistic to use for feature extraction. The de–facto approach is to use the mean but as the results show, this statistic only gives the best results on three instances, and never when combined with SLR. When using SLR the multivariate mode proved to be the best statistic. For sensor A and B the multivariate mode scores highest although for sensor C the median scored 2% higher than the multivariate mode.

SLR showed remarkable ability to identify a subset of dyes that could accurately identify explosives from non–explosives. SLR did not reliably estimate the same features as important, but using the average rankings proved to be an adequate solution. Using the top 5 identified compounds we were are able to train a classifier that identified explosives without any false negatives (Fig. 7).

Other popular classification methods such as linear discriminant analysis (LDA) and support vector machines (SVM) was not considered since they do not *both* model posterior probabilities and implement sparsity although one could consider using sparse relevance vector machines. Further, the dyes contribute with three features each (the RGB values) so one could also consider using group lasso instead of traditional regularization. This will be subject for future study.

References

[1] M. S. Schmidt, N. Kostesha, F. Bosco, J. K. Olsen, C. Johnsen, K. A. Nielsen, J. O. Jeppesen, T. S. Alstrøm, J. Larsen, T. Thundat, M. H. Jakobsen, and A. Boisen, “Xsense - a miniaturised multi-sensor platform for explosives detection,” in *Proceedings of SPIE*, 2011, p. In press.

[2] K. S. Suslick, N. A. Rakow, and A. Sen, “Colorimetric sensor arrays for molecular recognition,” *Tetrahedron*, vol. 60, no. 49, pp. 11133–11138, Nov. 2004.

[3] N. Rakow, A. Sen, M. C. Janzen, J. B. Ponder, and K. S. Suslick, “Molecular recognition and discrimination of amines with a colorimetric array.,” *Angewandte Chemie (International ed. in English)*, vol. 44, no. 29, pp. 4528–32, July 2005.

[4] C. Zhang and K. S. Suslick, “A colorimetric sensor array for organics in water.,” *Journal of the American Chemical Society*, vol. 127, no. 33, pp. 11548–9, Aug. 2005.

[5] C. Zhang, D. P. Bailey, and K. S. Suslick, “Colorimetric sensor arrays for the analysis of beers: a feasibility study.,” *Journal of agricultural and food chemistry*, vol. 54, no. 14, pp. 4925–31, July 2006.

[6] N. V. Kostesha, T. S. Alstrøm, C. Johnsen, K. A. Nilesen, J. O. Jeppesen, J. Larsen, M. H. Jakobsen, and A. Boisen, “Development of the colorimetric sensor array for detection of explosives and volatile organic compounds in air,” in *Proceedings of SPIE*, Apr. 2010, vol. 7673, pp. 76730I–76730I–9.

[7] N. V. Kostesha, T. S. Alstrøm, C. Johnsen, K. A. Nielsen, J. O. Jeppesen, J. Larsen, A. Boisen, and M. H. Jakobsen, “Multi-colorimetric sensor array for detection of explosives in gas and liquid phase,” in *Proceedings of SPIE*, 2011, p. In Press.

[8] S. H. Lim, L. Feng, J. W. Kemling, C. J. Musto, and K. S. Suslick, “An Optoelectronic Nose for Detection of Toxic Gases.,” *Nature chemistry*, vol. 1, pp. 562–567, Sept. 2009.

[9] C. Zhang and K. S. Suslick, “Colorimetric sensor array for soft drink analysis.,” *Journal of agricultural and food chemistry*, vol. 55, no. 2, pp. 237–42, Jan. 2007.

[10] D. Kim, V. M. Lynch, K. Nielsen, C. Johnsen, J. O. Jeppesen, and J. L. Sessler, “A chloride-anion insensitive colorimetric chemosensor for trinitrobenzene and picric acid.,” *Analytical and bioanalytical chemistry*, vol. 395, no. 2, pp. 393–400, Sept. 2009.

[11] J. S. Park, F. Le Derf, C. M. Bejger, V. M. Lynch, J. L. Sessler, K. Nielsen, C. Johnsen, and J. O. Jeppesen, “Positive Homotropic Allosteric Receptors for Neutral Guests,” *A European Journal*, vol. 16, no. 3, pp. 848–54, Jan. 2010.

[12] M. C. Janzen, J. B. Ponder, D. P. Bailey, C. K. Ingison, and K. S. Suslick, “Colorimetric sensor arrays for volatile organic compounds.,” *Analytical chemistry*, vol. 78, no. 11, pp. 3591–600, June 2006.

[13] B. A. Suslick, L. Feng, and K. S. Suslick, “Discrimination of complex mixtures by a colorimetric sensor array: coffee aromas.,” *Analytical chemistry*, vol. 82, no. 5, pp. 2067–73, Mar. 2010.

[14] X. Luo, P. Liu, C. Hou, D. Huo, J. Dong, H. Fa, and M. Yang, “A novel chemical detector using colorimetric sensor array and pattern recognition methods for the concentration analysis of NH3.,” *The Review of scientific instruments*, vol. 81, no. 10, pp. 105113, Oct. 2010.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[16] M. Schmidt, G. Fung, and R. Rosales, “Fast Optimization Methods for L1 Regularization: A Comparative Study and Two New Approaches,” in *Machine Learning: ECML 2007*, Joost Kok, Jacek Koronacki, Raomon Mantaras, Stan Matwin, Dunja Mladenic, and Andrzej Skowron, Eds. 2007, vol. 4701 of *Lecture Notes in Computer Science*, pp. 286–297, Springer Berlin / Heidelberg.

[17] S. Sigurdsson, J. Larsen, L. K. Hansen, P. A. Philipsen, and H. C. Wulf, “Outlier estimation and detection application to skin lesion classification,” in *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP ’02). IEEE International Conference on*, 2002, vol. 1, pp. I–1049 – I–1052 vol.1.

[18] T. S. Alstrøm and J. Larsen, “Feature Extraction and Signal Representation for Colorimetric Sensor Arrays,” Tech. Rep., DTU Informatics, 2011.

[19] I. T. Nabney, “`http://www1.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/`,”.

[20] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, Mar. 2003.