Proceedings Volume 11870, Artificial Intelligence and Machine Learning in Defense Applications III,
13-18 September 2021, Virtual Event
DOI:10.1117/12.2600182

Classification of colorimetric sensor data using time series

Deena P. Francis[a], Milan Laustsen[b], Hamid Babamoradi[c], Jesper Mogensen[d], Eleftheria Dossi[e], Mogens H. Jakobsen[f], and Tommy S. Alstrøm[a]

[a]DTU Compute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[b]CrimTrack ApS, Denmark
[c]DTU Nanotech, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[d]Danish Emergency Management Agency
[e]Centre for Defence Chemistry, Cranfield University, Defence Academy of United Kingdom, Shrivenham, SN6 8LA, UK
[f]DTU Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

ABSTRACT

Colorimetric sensors are widely used as pH indicators, medical diagnostic devices and detection devices. The colorimetric sensor captures the color changes of a chromic chemical (dye) or array of chromic chemicals when exposed to a target substance (analyte). Sensing is typically carried out using the difference in dye color before and after exposure. This approach neglects the kinetic response, that is, the temporal evolution of the dye, which potentially contains additional information. We investigate the importance of the kinetic response by collecting a sequence of images over time. We applied end-to-end learning using three different convolution neural networks (CNN) and a recurrent network. We compared the performance to logistic regression, k-nearest-neighbor and random forest, where these methods only use the difference color from start to end as feature vector. We found that the CNNs were able to extract features from the kinetic response profiles that significantly improves the accuracy of the sensor. Thus, we conclude that the kinetic responses indeed improves the accuracy, which paves the way for new and better chemical sensors based on colorimetric responses.

Keywords: colorimetric sensor, deep learning, kinetic response, time series classification, convolutional neural network

1. INTRODUCTION

Colorimetric sensors have been used successfully in the detection of several organic and inorganic substances, ranging from explosives to microbes. The data generated from such sensors are used in numerous applications such as organic molecule detection,[1] detection of explosives,[2] detection of toxins[3] and so on. This work focuses on the application of machine learning, namely, time series classification algorithms on such dataset, with application to explosives detection. Previous work on colorimetric sensor data classification[4] have worked with the color change of the sensor dyes after a set amount of time, denoted $Data_{diff}$ using classifiers such as k-Nearest Neighbor (KNN), Random Forest (RF) and Logistic Regression (LR). In that particular work, it was demonstrated that RF was able to detect explosives and their precursors with very good accuracy.[5] used LR to classify different species of pathogenic bacteria, while[6] used KNN and Linear Discriminant Analysis for identifying acidic and basic samples. In all these works, the $Data_{diff}$ has been used as input to the classifiers like KNN, RF and LR. This type of analysis of the colorimetric sensor data is considered the state-of-the-art. In contrast, our work focuses on the kinetic aspect of the data generated through the entire experiment of exposing a colorimetric sensor chip with an analyte. The time series produced through the experiments are used for performing time series classification using neural networks. We aim to enhance the understanding of time series aspects of the colorimetric sensor data, and to exploit this aspect for achieving good detection performance. This is important in improving existing detection capabilities as well paving the way for future developments in detection.

Further author information: (Send correspondence to D. P. F.)
D.P.F.: E-mail: dfra@dtu.dk

Published by SPIE. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial License (CC:BY:NC 4.0).
The final published version (version of record) is available online at DOI:10.1117/12.2600182. Please refer to any applicable publisher terms of use.

1.1 Aim and contributions

In this work, deep learning architectures that have been successful in other domains have been utilized to perform accurate classification of colorimetric sensor time series data.

Aim: *To use time series/kinetic nature of the colorimetric sensor data that encodes the characteristics of the chemical reaction for the accurate detection of explosives, their precursors and other compounds.*

The contributions of this work are as follows.

- The novel characterisation of colorimetric sensor data as multidimensional time series.
- Development of time series classifiers that perform better than the state-of-the-art in colorimetric data classification.

The results are expected from a machine learning perspective, so it should be stressed that the novelty is on the colorimetric sensing side. They provide empirical evidence that it is worth-while to collect the kinetic responses of colorimetric sensors, and not doing so will result in loss of information. This work focuses on the kinetic aspect of the data generated through the entire experiment of exposing a colorimetric sensor chip with an analyte. The time series produced through the experiments are used for performing time series classification using neural networks. We aim to enhance the understanding of time series aspects of the colorimetric sensor data, and to exploit this aspect for achieving better detection performance when compared to the state-of-the-art.

2. PROPOSED METHODS

We use the colorimetric data obtained through the sensor chips for automatically detecting the presence of Home-Made Explosives (HMEs) and their precursors.

The benefits of taking a machine learning approach on HME detection can be summarized as follows.

- *Improved detection capability*: Manually observing the colorimetric sensor array after exposing it to the analytes and making subsequent inferences is not only error-prone, but also difficult, owing to the complexity of the sensors used today. An algorithmic solution to observing and detecting the presence of analytes is more convenient as well as less-error prone.
- *Improvement of sensor chip design*: By analyzing the data obtained from a colorimetric sensor, meaningful insights about the usefulness of its components (dyes) can be obtained. For instance, the usefulness of certain dyes towards detection of certain target analytes can be quantified and subsequently, the sensor can be made more effective by weeding out unimportant dyes.

Input data: The data obtained from the sensor are in the form of images that have been captured by a digital camera, taken at a regular interval of 'm' and for a fixed duration of T, depending on the analyte. Typical values for m and T 15 seconds and 5 minutes respectively. The data processing pipeline[7] that consists of a localization of dyes, pixel normalization and background removal is employed. This pipeline extracts the pixel values of the spots within the chips and performs normalization to account for illumination effects. The resulting data is in the form of several measurements of the values of the dyes taken over T. Subsequently, we obtain data corresponding to each dye for different time points. There are 29 unique chemicals, called dyes, on a chip, and since an RGB image is captured by the camera, there are 3 channels for each dye. We consider two kinds of data in this work: (1) Data of the dyes extracted by the data processing pipeline, each example naively treated as independent observations (2) Time series data that includes the measurements (obtained after passing through the data processing pipeline) as a single example. The time series consists of the number of dyes times the number of channels generated multiple times with different target analytes (chemical compounds of interest).

The traditional method of classifying colorimetric sensor data is to use what we refer to as the delta data. In this approach, the difference between the last image and first image captured by the camera is computed, and this difference serves as the input to the classifier.[8] This approach has been shown to be very effective in various tasks. In this work we also try alternative methods of generating the input to the classifier, such as time series

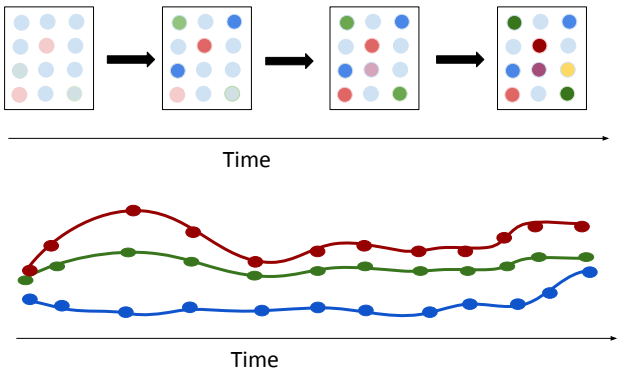


Figure 1. Method of generating the time series data.**Top** image of the colorimetric chip is captured at various time intervals using a digital camera. **Bottom** The R, G, B channel variation for a dye.

data because it captures the kinetic nature of the chemical reaction taking place when the chip is exposed to the analyte.

Classifiers: In order to classify the two kinds of data, two different categories of classifiers are needed. In the first category, we have traditional classifiers such as Random Forest (RF), k-Nearest Neighbors (KNN) and Logistic Regression (LR) that have been proven to be good for classifying sensor data. In the second category, we consider classifiers that have been successful in classifying time series data such as Convolutional Neural Networks (CNN), its variants and Long Short-Term Memory (LSTM) Networks.

2.1 Traditional classifiers

2.1.1 RF

This classifier belongs to the family of ensemble learning algorithms. The ensemble in this context refers to a collection of Decision Tree classifiers. The prediction made by the random forest is the prediction that is output by majority of the trees in it. This method has been shown to have the best performance in classification of colorimetric sensor data.[4]

2.1.2 KNN

The assignment of a class to a datapoint occur by means of a distance metric. This metric is used to assess if a data point is close to another point (belonging to a particular class). This algorithm is particularly susceptible to noise or irrelevant features, which can affect its accuracy.

2.1.3 LR

This classifier uses the logistic function to model the probability of a datapoint belonging to a particular class. In order to do this the logistic functions parameters, need to be learned. This method can be thought of as a single layer of a neural network.

2.2 Time series classifiers

The detection of explosives from the time series extracted from colorimetric sensor data is done by means of the following deep learning architectures: CNN, Fully Convolutional Neural Network (FCN), Temporal Convolutional Network with an Encoder-Decoder architecture (TCN-ED) and a Long Short Term Memory network (LSTM). We emphasize that the approaches used in this work are not an exhaustive list of time series classifiers, but rather the methods that were chosen are suitable for the type of datasets used in this work. In general, CNN and its variants, namely, FCN and TCN-ED have been chosen because of the ability of CNNs to extract features from time series data, and have proven success in other tasks and domains.[9] The LSTM architecture was included because of its success in various applications[10],[11] and its ability to exploit the dependencies that exists in time series. Figure 1 shows the method of generating time series data and the classifier architectures.

2.2.1 Convolutional Neural Network

Convolution is the key operation of a CNN[12] and is in neural networks described as $y = f(\mathbf{w} * X + b)$. Here $\mathbf{w}$ is a weight vector of length m that constitutes a filter (or sometimes called a convolution kernel), b is a bias term, X is the multidimensional time series of length T, and f is is typically a non-linear function like a Rectified Linear Unit (ReLU). When X is a multivariate time series, the filter $\mathbf{w}$ takes the same dimensionality of the input. For our use-case, we consider a CNN with one dimensional (1D) convolution, which computes y along a single dimension, namely, time. Thus the result of a single filter is another time series y. The application of several filters on the input time series results in a multivariate time series with dimensionality equal to the number of filters used. In our experiments we observed that a simple CNN with just one convolution layer gave good results. In fact, it was observed that increasing the number of layers, significantly reduced the generalisation performance.

2.2.2 Fully Convolutional Neural Network

The main difference between a CNN and an FCN[13] is the way in which pooling is done. In the former, local pooling layers are used after each convolution layer, and in the latter, a single global pooling layer replaces the local pooling layers. This architecture has the advantage of drastically reducing the number of parameters in the network.

2.2.3 Temporal Convolutional Network

A TCN uses multiple layers of temporal convolution (convolution along time) followed by pooling layer and the discriminatory layer. In our case, we used an encoder-decoder architecture of TCN made popular for the task of action segmentation and detection.[14]

2.2.4 Long Short Term Memory network

LSTMs[15] have a fundamentally different architecture compared to CNNs and its variants. They use a recurrent architecture where the nodes form a directed graph with feedback connections. In addition, temporal dependencies within the data are captured using special gate and memory cell mechanisms. LSTMs have been shown to be effective in learning the long-range dependencies that exist in data. These networks have been used for language translation,[16] time series classification,[17] forecasting[18] among other things.

3. EXPERIMENTS AND RESULTS

This section details the results obtained when performing multi-class classification for the two datasets used in this work. Initially, the datasets are split into train and test sets, with the training set further undergoing a 10-fold cross validation for model selection. The best model obtained during cross-validation is used to classify the test set. The above procedure is repeated 10 times where each run uses randomized data splits, and then the aggregated results are reported. The aggregation is done by computing the sum of the resulting confusion matrices, and deriving the following performance metrics from it.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

,

$$F1\text{-}score = \frac{2 * Precision}{Precision\ +\ Recall}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Here TP refers to the number of True Positives, FP refers to the number of False Positives, TN refers to the number of True Negatives and FN refers to the number of False Negatives. Since our goal is the detection of certain analytes, we are interested in measuring the above metrics for these individual analytes. In the

results section (Section 3.3), there are two kinds of results: (a) per analyte metrics (Table 1) - the values of the performance metrics are reported for each class in the dataset. This tells us how well the classifier detects individual analytes, (b) average values of metrics: a single value corresponding to the average value of each metric over all classes in the dataset. This summarizes the overall performance of the classifier on the dataset.

3.1 Baseline

The state-of-the-art in colorimetric sensor analysis consists of the following method. An image of the colorimetric chip prior to exposing it to the target analyte is captured, and another image is captured after exposing the chip to the analyte for a specified time period (often in the 5 minute range). Subsequently, the difference between the second and first image is computed. This difference, which is called the difference image in the literature,[8] serves as the input to the classification algorithm. We refer to this type of data as $Data_{diff}$. Typical classifiers used in the literature include, RF and KNN. We use this data construction method, RF, KNN and LR as baseline for our experiments.

3.2 Dataset description

Data acquisition: The measurements were carried out using a prototype called the CRIM-TRACK sniffer. This apparatus has an exchangeable flow chamber that houses the colorimetric chip, into which airflow is pumped in using an air sampling system. The experimental campaigns are referred to as box experiments because the sample of the target analyte is kept inside a closed box, and the resulting target analyte air sample is pumped into the exchangeable chamber.

- HMTD dataset: This dataset was obtained from an experimental campaign of box experiments with various samples of HMTD, H_2O_2, Hexamine and Ambient air. It contains the following percentage of instances 57.5%, 12.5%, 15% and 15% respectively.

We demonstrate good detection capability of these two explosives using the methods described in Section 2. In addition, we compare the results obtained with the results of the baseline method (Section 3.1).

3.3 Results

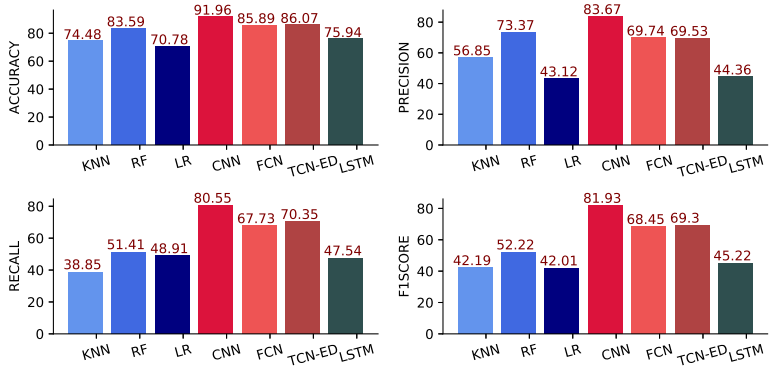


Figure 2. Plots of average values of metrics for the HMTD dataset.

The first comparison of performance is done with respect to average values of the performance metrics, which helps to gain an overall picture of the classifier performance. The results are shown in Figure 2. The metrics shown in the figure correspond to those of the main analyte that we are interested in, namely HMTD. For the HMTD dataset (Figure 2), the RF classifier achieves the best performance, with the highest F1-score. Although the TNR of LR appears to be higher than that of the others, its recall and hence F1-score is low.

The second comparison of performance is done between the per-class performance metrics of the classifiers for the dataset. The results are shown in Table 1. For the HMTD dataset, CNN obtains the best performance in terms of F1-score, Accuracy and Precision. Although RF has a very high recall, its precision is much lower than that of CNN. In the case of other analytes, CNN achieves much better results when compared to the others.

Analyte	Method	Precision	Recall	F1-score	Accuracy
HMTD	KNN	50.8	69.1	58.6	51.0
	RF	60.9	**98.8**	75.4	67.7
	LR	60.6	26.9	37.2	54.7
	CNN	**85.5**	88.4	**86.9**	**84.8**
	FCN	76.9	77.2	77.1	73.8
	TCN-ED	80.9	74.5	77.6	75.4
	LSTM	70.4	57.7	63.4	61.9
Hexamine	KNN	60.9	17.5	27.2	84.4
	RF	**100**	25	40	87.5
	LR	38.2	65.0	48.2	76.7
	CNN	90.4	**82.5**	**86.3**	**96.3**
	FCN	64.9	60	62.3	89.6
	TCN-ED	62.6	71.3	66.7	89.8
	LSTM	39.7	38.8	39.2	82.9
H_2O_2	KNN	94.7	44.4	60.4	90.3
	RF	**99.2**	78.8	87.8	96.4
	LR	41.7	37.5	39.5	80.8
	CNN	97.1	85.0	**90.7**	**97.5**
	FCN	89.7	76.3	82.4	95.4
	TCN-ED	86.9	70.6	77.9	94.3
	LSTM	39.8	58.8	47.5	81.4
Ambient air	KNN	21.1	24.4	22.6	72.2
	RF	33.3	3.1	5.7	82.8
	LR	32.0	66.3	43.2	70.9
	CNN	**61.6**	**66.3**	**63.9**	**89.3**
	FCN	47.4	57.5	51.9	84.8
	TCN-ED	47.7	65.0	55.0	84.8
	LSTM	27.5	35.0	30.8	77.5

Table 1. Results for HMTD dataset

4. DISCUSSION AND CONCLUSION

The results in Section 3 indicate that the time series representation of the colorimetric sensor data combined with a simple CNN architecture is capable of detecting the explosives much better than the state-of-the-art. There are a couple of notable observations:

- Superior classification results were obtained when time series classifiers were used to classify the entire time series obtained from the sensor. In particular, the CNN classifier appears to have the best overall performance on the dataset. This implies that the time series data (the kinetic response) contains important information about the chemical reaction, which is crucial to accurately detecting the analytes.
- LSTM does not perform well in comparison with others. In our opinion, this indicates that the convolution (of CNN) is more suitable to the dataset that we used than the gating and memory mechanism of an LSTM
- The state-of-the-art in colorimetric sensor data classification were much less effective in detecting the explosives when compared to the methods proposed in this work. This implies that the time series data (the kinetic response) contains important information about the chemical reaction, which is crucial to accurately detecting the analytes.

This work differs from the previous works in colorimetric sensor data classification in two main aspects: (1) the type of data being used, and (2) the kind of classifiers that are used. The experiments indicate that the time series data combined with time series classifiers is indeed useful for detecting analytes more effectively than the state-of-the-art.

ACKNOWLEDGMENTS

We acknowledge the support from the Danish Defense, contract number 4600005161, for the PORDEX project and EU FP7 grant agreement, contract number 313202, for the CrimTrack project. The authors would also like to thank Caroline O'Keefe, Cranfield University, Mads Skak Jensen, DEMA, Ida Thygesen, DTU Nanotech, Paul D. Mines, DTU Nanotech, Lasse Mølgaard, DTU Compute for their help in making the data available.

REFERENCES

[1] Suslick, K. S., Rakow, N. A., and Sen, A., "Colorimetric sensor arrays for molecular recognition," *Tetrahedron* **60**(49), 11133–11138 (2004).

[2] Kostesha, N. V., Alstrøm, T. S., Johnsen, C., Nilesen, K. A., Jeppesen, J. O., Larsen, J., Jakobsen, M. H., and Boisen, A., "Development of the colorimetric sensor array for detection of explosives and volatile organic compounds in air," in [*Advanced environmental, chemical, and biological sensing technologies VII*], **7673**, 76730I, International Society for Optics and Photonics (2010).

[3] Feng, L., Musto, C. J., Kemling, J. W., Lim, S. H., Zhong, W., and Suslick, K. S., "Colorimetric sensor array for determination and identification of toxic industrial chemicals," *Analytical chemistry* **82**(22), 9433–9440 (2010).

[4] Mølgaard, L. L., Buus, O. T., Larsen, J., Babamoradi, H., Thygesen, I. L., Laustsen, M., Munk, J. K., Dossi, E., O'Keeffe, C., Lässig, L., Tatlow, S., Sandström, and L., Jakobsen, M. H., "Improved detection of chemical substances from colorimetric sensor data using probabilistic machine learning," in [*Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XVIII*], **10183**, 1018307, International Society for Optics and Photonics (2017).

[5] Lonsdale, C. L., Taba, B., Queralto, N., Lukaszewski, R. A., Martino, R. A., Rhodes, P. A., and Lim, S. H., "The use of colorimetric sensor arrays to discriminate between pathogenic bacteria," *PLoS One* **8**(5), e62726 (2013).

[6] Kangas, M. J., Burks, R. M., Atwater, J., Lukowicz, R. M., Garver, B., and Holmes, A. E., "Comparative chemometric analysis for classification of acids and bases via a colorimetric sensor array," *Journal of chemometrics* **32**(2), e2961 (2018).

[7] Alstrøm, T. S., "Assessing miniaturized sensor performance using supervised learning, with application to drug and explosive detection," (2013).

[8] Li, Z., Askim, J. R., and Suslick, K. S., "The optoelectronic nose: colorimetric and fluorometric sensor arrays," *Chemical reviews* **119**(1), 231–292 (2018).

[9] Dhillon, A. and Verma, G. K., "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence* **9**(2), 85–112 (2020).

[10] Yoo, H., "Deep convolution neural networks in computer vision: a review," *IEIE Transactions on Smart Processing & Computing* **4**(1), 35–43 (2015).

[11] Wehrmann, J., Becker, W., Cagnini, H. E. L., and Barros, R. C., "A character-based convolutional neural network for language-agnostic twitter sentiment analysis," in [*2017 International Joint Conference on Neural Networks (IJCNN)*], 2384–2391, IEEE (2017).

[12] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D., "Backpropagation applied to handwritten zip code recognition," *Neural computation* **1**(4), 541–551 (1989).

[13] Wang, Z., Yan, W., and Oates, T., "Time series classification from scratch with deep neural networks: A strong baseline," in [*2017 International joint conference on neural networks (IJCNN)*], 1578–1585, IEEE (2017).

[14] Lea, C., Flynn, M. D., Vidal, R., Reiter, A., and Hager, G. D., "Temporal convolutional networks for action segmentation and detection," in [*proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 156–165 (2017).

[15] Hochreiter, S. and Schmidhuber, J., "Long short-term memory," *Neural computation* **9**(8), 1735–1780 (1997).

[16] Guo, D., Zhou, W., Li, H., and Wang, M., "Hierarchical lstm for sign language translation," in [*AAAI*], (2018).

[17] Tang, Y., Xu, J., Matsumoto, K., and Ono, C., "Sequence-to-sequence model with attention for time series classification," in [*2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*], 503–510, IEEE (2016).

[18] Gao, M., Li, J., Hong, F., and Long, D., "Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using lstm," *Energy* **187**, 115838 (2019).



https://dspace.org/

Classification of colorimetric sensor data using time series

Francis, Deena P.; Laustsen, Milan; Babamoradi, Hamid; Mogensen, Jesper; Dossi, Eleftheria ...

2021-09-12
Society of Photo-Optical Instrumentation Engineers (SPIE)
https://doi.org/10.1117/12.2600182
Francis DP, Laustsen M, Babamoradi H, et al., (2021) Classification of colorimetric sensor data using time series. In: Artificial Intelligence and Machine Learning in Defense Applications III, 13-18 September 2021, Virtual Event

Downloaded from a DSpace repository. https://dspace.org/