https://doi.org/10.2352/ISSN.2470-1173.2021.16.COLOR-342


# Spectral Imaging to Differentiate Contaminant Levels

***Min Zhao, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA***
***Qiyue Liang, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA***
***Susana Diaz-Amaya, School of Materials Engineering, Purdue University, West Lafayette, IN, USA***
***Amanda J. Deering, Department of Food Science, Purdue University, West Lafayette, IN, USA***
***Lia Stanciu, School of Materials Engineering, Purdue University, West Lafayette, IN, USA***
***George T.-C. Chiu, School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA***
***Jan P. Allebach, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA***

## Abstract

*Mercury (Hg) and Arsenic (As) have been recognized as chemical threats to human health. Still, the detection of lower contamination levels using traditional image analysis remains challenging due to the small number of available data samples and the insufficient utilization of the spatial information contained in the sensor pad images. To overcome this challenge, we use the spectra data of the colorimetric response pads and propose two kinds of classification models for differentiating contaminant levels with high test accuracy. In the first model, we use the SMOTE method to solve the imbalanced data problem, then apply the sequential forward selection algorithm to select optimal wavelength features in combination with the k-NN classifier to discriminate five contaminant levels. The second technique comprises principal component analysis (PCA) used as a dimensionality reduction technique combined with the random forest (RF) classifier to classify five contaminant levels. Our proposed system is trained and evaluated on a limited dataset of 126 spectral responses of five contamination levels. Our algorithms can yield 77% and 87% average accuracy, respectively. We will present an overview of the base model, the pipelines and the comparison of our proposed two classification models, and the phone-based narrow-band spectral imaging system that can obtain the camera spectral response for accurate and precise heavy metals analyses with the aid of narrow bandpass filters in front of a cell phone's camera lens.*

## 1. Introduction

Nowadays, the safety of food has become crucial. One of the main types of threats related to food safety is heavy metals [1]. Heavy metals, including Mercury (Hg), Arsenic (As), Copper (Cu), and so on, can be enriched in living tissue through food chains and have been reported to be harmful to human health at low concentrations. The commonly used methods for detecting heavy metals are mass spectroscopy, atomic emission spectroscopy, potentiometric methods, and so on [2]. These methods are sensitive, but require expensive equipment, trained personnel, and cannot support on-site detection. Therefore, rapid and low-cost detection methods for contaminants are more and more in demand.

To detect multiple targets in one test, our group focuses on developing a novel paper-based, microfluidic biosensor for colorimetric detection of two types of heavy metals: As and Hg [3], [4]. Figure 1 shows the proposed detection mechanism of our biosensors and the test interpretation. A cell-phone integrated image analysis pipeline can determine the detection result of our biosensors. Two kinds of the aptamer-functionalized particles (ssDNA-PEI-Au-Ps) specific to $Hg^{2+}$ and $As^{3+}$ are preloaded on each of the upper, and lower two circular pads, respectively. These four pads serve as the colorimetric labels. To detect the analytes in the test samples, test samples with different concentrations of $Hg^{2+}$ or $As^{3+}$ were dropped in the inlet of the biosensors. There is a colorimetric response in the presence of the target after the test solution interacts with the colorimetric label deposited on each of the testing areas. Figure 2 shows the colorimetric signal evolution versus various concentrations of $As^{3+}$ and $Hg^{2+}$ from 0 ppm to 100 ppm. The color of the target testing areas gradually changes from light pink to deep purple as the concentration increases from 0 ppm to 30 ppm. Then, there is a drastic color change from deep gray to light gray for a higher concentration test, from 50 ppm to 100 ppm.

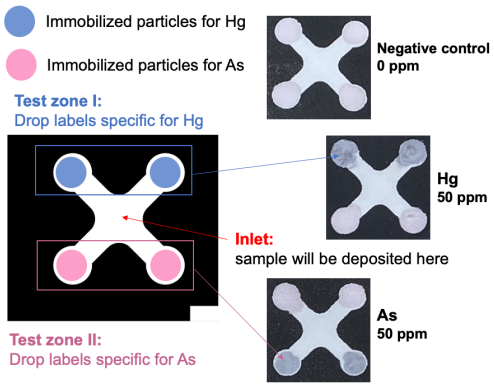



***Figure 1.*** *The detection mechanism of our biosensors and test interpretation. (To illustrate the different particles specific for Hg and As, the particles specific for Hg are labeled blue in the figure; but the actual particles are colored light pink.)*

To yield a quantitative and objective color analysis, color measuring instruments are favored. As digital technologies continue to develop, cheap, and compact image sensors are widely used in everyday electronics, like cell phones. A phone-based imaging system is promising for signal detection due to the above features emerging in different fields.

In our previous work, our proposed optical system and image analysis pipeline provides consistent data acquisition captured by a mobile phone camera, and delivers quantitative responses to correlate the colorimetric change of the biosensors to the concentration of the target substance [4]. We used the grayscale values as

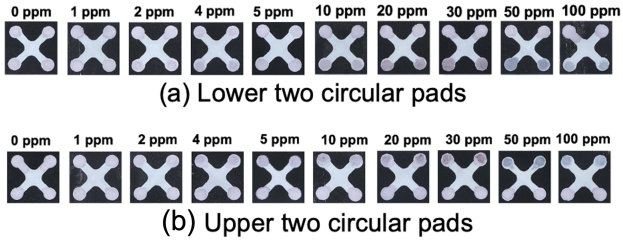



***Figure 2.*** *(a) The colorimetric signal response to $As^{3+}$, (b) The colorimetric signal response to $Hg^{2+}$.*

a metric, calculating average *CIE* $\Delta E$ [5], [6] from the white reference, to characterize the response of the paper-based devices, and to correlate the response with the concentration of the analytes. As an example, Figure 3 shows the correlations between the $\Delta E$ values and the increasing concentrations of $As^{3+}$. According to the data collected, the variable $\Delta E$ and the $As^{3+}$ concentration is found to be strongly correlated from 0 to 30 ppm, which can be fitted as $y = 0.52x + 17.12$, with $R^2$ of 0.9238. Nonetheless, the relationship is not monotonic. In particular, the repsonses to 0 ppm and 4 ppm are too high.

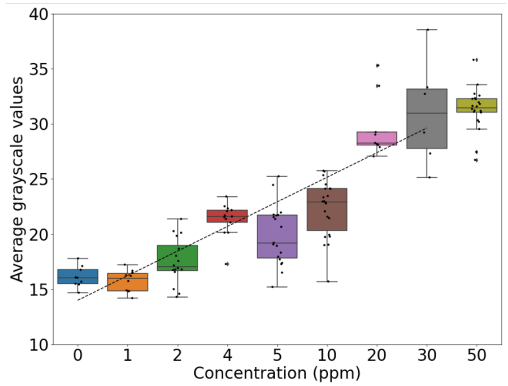



***Figure 3.*** *The correlation between $\Delta E$ values and the increasing concentrations of $As^{3+}$ for a non-spectral imaging method.*

Thus, we aim to find a prediction model with higher accuracy for our limited data set. In this study, the colorimetric responses of 5 contamination levels ($As^{3+}$) are used as the experimental data. Considering the limited number of samples for each concentration, we first rearrange the 126 phone captured images of the samples into five classes: 35 in Class 1 (0, 1, 2 ppm), 32 in Class 2 (4, 5 ppm), 22 in Class 3 (10 ppm), 15 in Class 4 (20, 30 ppm), and 22 in Class 5 (50 ppm). Then, we divide the original dataset into a training set, a validation set, and a test set according to the ratio 5: 2: 3, as shown in Table 1.

**Table 1: Overview of the small-scale dataset showing the division, respectively, into training, validation, and test sets.**

| Class | Training set | Validation set | Test set | Total |
|---|---|---|---|---|
| Class 1 | 17 | 7 | 11 | 35 |
| Class 2 | 16 | 6 | 10 | 32 |
| Class 3 | 11 | 4 | 7 | 22 |
| Class 4 | 7 | 3 | 5 | 15 |
| Class 5 | 11 | 4 | 7 | 22 |

In the following sections, we first present an overview of the base model using grayscale values and its prediction accuracy. Second, we propose two classification models for discriminating the spectral responses for the different classes; then, we compare the prediction accuracy. Finally, we present our phone-based narrow-band spectral imaging system that can obtain the camera spectral response for accurate and precise heavy metals analyses with the aid of narrow bandpass filters in front of the cell phone's camera lens.

## 2. Base Model

To evaluate the prediction accuracy of our base model, we apply the Lloyd-Max scalar quantizer method to find the optimal threshold boundaries, i.e. average deltaE values, for the five classes [7], [8]. Figure 4 shows the optimal threshold boundaries for the five classes. The prediction accuracy for the test dataset is shown in Table 2. The base model shows a relatively high prediction accuracy for Classes 1 and 5, but a relatively low accuracy for Classes 2 - 4, with an average prediction performance of 41%.

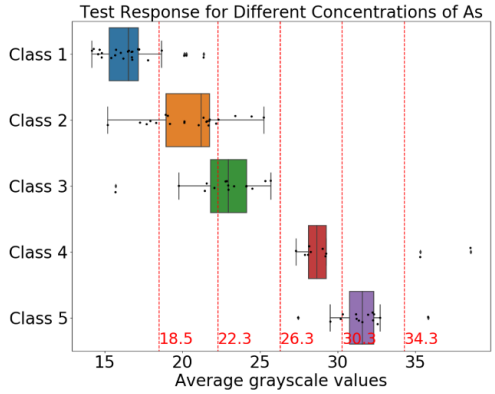



***Figure 4.*** *The optimal threshold boundaries for 5 classes detecting $As^{3+}$ for the base non-spectral imaging method.*

**Table 2: Performance of the method based on $\Delta E$ from the global background.**

| Class | C 1 | C 2 | C 3 | C 4 | C 5 |
|---|---|---|---|---|---|
| Accuracy | 91% | 60% | 43% | 20% | 86% |

The $\Delta E$ from the global background method calculates the Euclidean distance between the white reference and the response area. As can be seen in Figure 4, there exist situations where the $\Delta E$ values calculated from two different reaction colors are out of order, which causes low accuracy. The limited dataset and the insufficient utilization of the spatial information contained in the sensor pad images also restrict the effectiveness of the base model.

## 3. Spectral Imaging Classification Models

The main challenges of our project are the insufficient feature information, the limited number of samples, and the large intra-class variance of the sensor pad images. To overcome these challenges, we use the spectra data of the colorimetric response pads and propose two kinds of classification models for differentiating contaminant levels with higher test accuracy.

### 3.1 Non-contact optical measurement system

The spectral radiance in the detection zones is acquired by our non-contact optical measuring system, illustrated in Figure 5. The optical system mainly consists of a photo studio booth (FotodioX, purchased from bhphotovideo.com) for providing the controlled illumination environment, and for measuring the visible wavelength range, a spectroradiometer (PR 705, Photo Research Inc., CA, USA), and a tripod to provide a 45° configuration of the spectroradiometer. The spectral radiance of the samples and of a white reflectance standard (Spectralon white diffuse reflectance standard, model #54-302, Edmund Optics) are obtained from 380 nm to 780 nm with an interval of 2 nm. To evaluate stability of the illumination light intensity, the measurements of the spectral irradiance of the white reflectance standard are taken every 30 minutes for 3 hours. Then the spectral reflectance of the colorimetric response is calculated by dividing the spectral radiance of the object by the average reflectance radiance of the perfect reflecting standard under the same spectral conditions of measurement [9], [10].

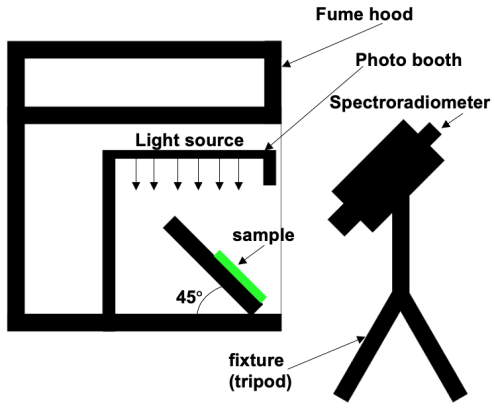



***Figure 5.*** *The optical setup for spectral data acquisition.*

The light sources used are an LED light (FotodioX, purchased from bhphotovideo.com), fluorescent light (mounted on the ceiling of the laboratory), and halogen light (Sunlite, purchased from bhphotovideo.com). Figure 6 shows the relative luminous power comparison of these three kinds of illumination used in our project. Our goal is to find the optimal illumination source for which the spectral data of the entire set of objects (responses of $As^{3+}$ and of $Hg^{2+}$) are most distant from each different concentration. The spectral radiance of each object under the three illuminations is taken and averaged with respect to wavelength. Table 3 shows that the LED source is optimal to distinguish the different concentrations of the colorimetric response of both $As^{3+}$ and $Hg^{2+}$. Therefore, LED illumination is used as an optimal light source for spectral acquisition.

### 3.2 Data description

The spectral radiances of the colorimetric responses to $As^{3+}$ are used as the experimental data. An example of the measured spectral radiance curves for all 5 contamination levels is illustrated in Figure 7. For each contamination level, the curve is the mean of all spectral radiance measurements of the training samples. Although there are visible differences at approximately 460 nm and from 480 nm to 550 nm, the spectral radiance curves have

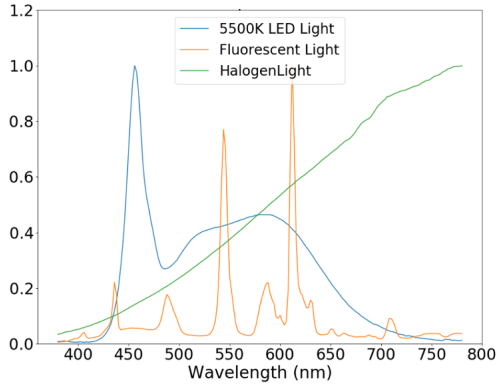



***Figure 6.*** *The relative luminous power comparison of the three kinds of illumination sources.*

**Table 3: The relationship between the averaged spectral reflectance and the contamination levels under the three different illuminations.**

| Illumination | $As^{3+}$ | $Hg^{2+}$ |
|---|---|---|
| Fluorescent Light | C1 → C2 → C3 → C4 → C5 | C2 → C1 → C3 → C4 → C5 |
| 5500K LED Light | C1 → C2 → C3 → C4 → C5 | C1 → C2 → C3 → C4 → C5 |
| Halogen Light | C1 → C2 → C3 → C4 → C5 | C2 → C1 → C3 → C4 → C5 |

quite similar shapes on the visible wavelength ranges.

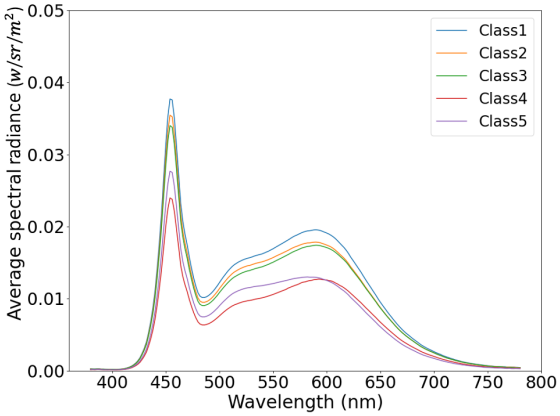



***Figure 7.*** *The mean spectral radiance measurements of the training samples for five contamination levels of $As^{3+}$.*

The data set comprising 126 spectral radiances at 5 contamination levels, are randomly divided into training, validation, and test data sets, as shown in Table 1. For each sample, the measured spectral band varies from 380 nm to 780 nm with a sampling interval of 2 nm; this leads to an original vector space of dimension 200. The number of components of the feature vector is much larger than the small number (less than 40) of samples for each class in our application. The number of training samples required for typical machine learning problems increases dramatically with the dimensionality for such high dimensionality [11]. Therefore, the first step in preprocessing is to obtain the smaller feature di-

mensions by dividing the spectral range of the wavelengths into 20 equal parts, then calculating the corresponding average. The corresponding feature names are '390 nm', '410nm', . . . , '770 nm'.

## 4. Machine Learning Algorithm

The next goal is to develop classification models for differentiating contaminant levels with high test accuracy. The majority of the existing works prove that k-nearest-neighbor and random forest classifiers have powerful classification capabilities [12], [13]. Further, sequential selection and PCA are widely used to extract a subset of features in higher dimensions to improve computational efficiency and reduce the generalization error of classification [14], [15]. Based on this, we use the spectral data of the colorimetric response pads and propose two kinds of classification models for differentiating contaminant levels with high test accuracy. The flow chart of the general machine learning pipeline for these two models is shown in Figure 8.

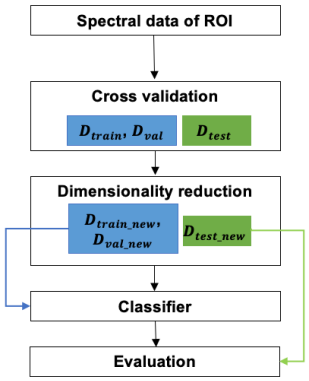



***Figure 8.*** *The flow chart of the general machine learning pipeline for two proposed classification models.*

Here, two kinds of classification models are investigated in our work to evaluate their classification accuracy and generalizability to the test data.

### *4.1 Multiclass classification model I*

In the first model, we apply the sequential forward feature selection (SFS) algorithm [14] to select or extract a subset of wavelength features in combination with the k-nearest-neighbor (k-NN) classifier to discriminate five contaminant levels.

### *A. Sequential forward feature selection*

After the first step of preprocessing, we have 20 wavelength features for each spectral radiance data. To improve the classification performance and simplicity, we can further reduce the dimensionalities. The sequential feature selection has been recognized as a crucial feature selection technique by applying an iterative procedure. The SFS method takes the following steps: (1) starts with an empty feature set, (2) generate all possible feature subsets of size 1, then choose the feature subset that leads to the best classification accuracy, (3) add another feature from the remaining available features to generate all possible feature subsets of size 2, (4) gradually add features until the size of the subsets is equal to the number of desired features. In this work, we apply the SFS based on the k-NN classifier to choose a subset of wavelength features that yields the minimum classification error.

### *B. k-NN classifier*

The k-NN classifier is one of the most widely used classification methods based on the majority vote of the neighbors of the test sample. The k-NN method calculates a distance between the test sample and all training samples to obtain its nearest neighbors, and then assigns the test sample a label according to the majority vote of the nearest neighbors [16].

We calculate the standard Euclidean distance to measure the similarity between the test sample and the training samples. As is well known, the selection of the value for $k$ is crucial for good classification performance. In this work, we obtain the appropriate value for k experimentally. With the validation data set, we evaluate the k-NN classifier with different $k$ values from 1 to 6. This procedure can be repeated each time by increasing $k$ to include one more neighbor. Two examples are illustrated in Figure 9. We note that almost all of these 6 models yield 80% validation accuracy when 4 wavelengths are selected, so the feature subset's size is taken as 4. The $k$ value is chosen to be 5 because the validation accuracy stabilizes after a certain point as the number of feature selections increases. The optimal wavelength features include 670 nm, 490 nm, 410 nm, and 430 nm.

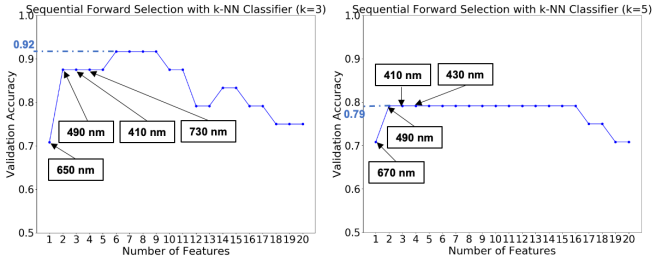



***Figure 9.*** *Validation accuracy of the k-NN classifier models with k = 3 and k = 5, respectively.*

### *C. Classification result of model I*

We perform 4-fold cross-validation to evaluate the performance of the k-NN classifier ($k$ = 5) with the SFS algorithm. The confusion matrix for the test data is reported in Table 4. The classification performance yields 84.42% average precision, 76.7% average recall, and 80.4% average F-1 score. We note that the low classification accuracy for Class 3 might be due to the fact that a fixed k-NN classifier is applied to all test samples. This leads to a low prediction rate in real applications in many existing works because the fixed classifier for all test data does not consider the distribution of the data [17].

**Table 4: Confusion matrix for the multiclass classification model I (k-NN classifier).**

| Class | C 1 | C 2 | C 3 | C 4 | C 5 | Precision | Recall |
|---|---|---|---|---|---|---|---|
| C 1 | 10 | 1 | 0 | 0 | 0 | 84.6% | 90.9% |
| C 2 | 4 | 5 | 1 | 0 | 0 | 62.5% | 50% |
| C 3 | 2 | 2 | 3 | 0 | 0 | 75 % | 42.9% |
| C 4 | 0 | 0 | 0 | 5 | 0 | 100% | 100% |
| C 5 | 0 | 0 | 0 | 0 | 7 | 100% | 100% |

### 4.2 Multiclass classification model II

The second classification model comprises principal component analysis (PCA) used as a dimensionality reduction technique in combination with a random forest (RF) classifier to classify five contaminant levels.

### A. PCA

Principle component analysis (PCA) is one of the most widely used algorithms for reducing redundant and irrelevant features. PCA uses singular value decomposition to project the high feature dimensions into an orthogonal basis set called the principal components, while preserving as much of the data's variation as possible [15], [18]. In our work, PCA is applied to the training data set to select the principal components that explain the data's maximum variance. Figure 10 shows that the first three principal components can cover the 99% of the variance of the training data set. Then we reduce the original feature vectors of the testing data to the same lower dimensional subspace as the training data set.

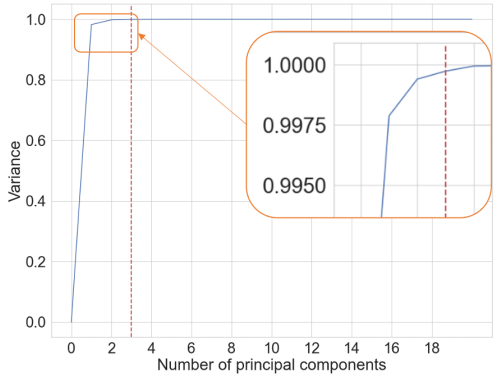



***Figure 10.*** *Variance of the first 20 principal components for the training data set.*

### B. Random forest

Random forest (RF) has been successfully applied to multiclass classification problems [19]. The RF algorithm is an ensemble method of classification based on generating multiple decision trees. The RF algorithm independently constructs each tree using bootstrap sample data. Each node in the standard tree is split using the best decision based on a randomly chosen subset of the input variables. Each tree in RF predicts its output; then, the RF makes a final prediction based on the majority vote of all the trees.

### C. Classification result of model II

The RF models involve several parameters: depth of trees, number of features randomly selected, and number of trees in the forest. We perform 4-fold cross-validation to select the optimal parameters which yield the lowest classification errors on the validation data set.

After selecting the optimal parameter values, the classification model is evaluated on the test data with the lower dimensional subspace. Table 5 shows the confusion matrix for the test data. The classification performance yields 86.82% average precision, 86.64% average recall, and 86.73% average F-1 score. It turns out that the RF model with PCA feature selection performs well in terms of accuracy compared to the k-NN classifier.

**Table 5: Confusion matrix for the multiclass classification model II (RF classifier).**

| Class | C 1 | C 2 | C 3 | C 4 | C 5 | Precision | Recall |
|---|---|---|---|---|---|---|---|
| C 1 | 9 | 1 | 1 | 0 | 0 | 90% | 81.8% |
| C 2 | 1 | 8 | 1 | 0 | 0 | 72.7% | 80% |
| C 3 | 0 | 2 | 5 | 0 | 0 | 71.4 % | 71.4% |
| C 4 | 0 | 0 | 0 | 5 | 0 | 100% | 100% |
| C 5 | 0 | 0 | 0 | 0 | 7 | 100% | 100% |

## 5. Phone-based Narrow Band Spectral Imaging

In the previous sections, we prove that the spectral data of the colorimetric response pads can improve classification performance. The challenging part is that the spectral data must be obtained using an expensive and professional optical component, like a spectroradiometer. As digital technologies continue to develop, cell phones are cheap and widely used globally, so a phone-based spectral imaging system is promising for differentiating different contaminant levels.

We propose a smartphone-based narrow-band spectral imaging system that is incorporated with a hardware plug-in module that fixes the bandpass filter in front of the smartphone's camera lens. In the optical setup, the cell phone (iPhone 8, CA, USA) with a bandpass filter (center wavelength is 620 nm, and full width at half maximum is 10 nm, Edmund Optics Inc., Barrington, NJ) replaced the spectroradiometer on the tripod, as shown in Figure 5. We also conduct a preliminary evaluation of the proposed phone-based narrow-band spectral imaging system regarding its performance and capability to replace the spectroradiometer.

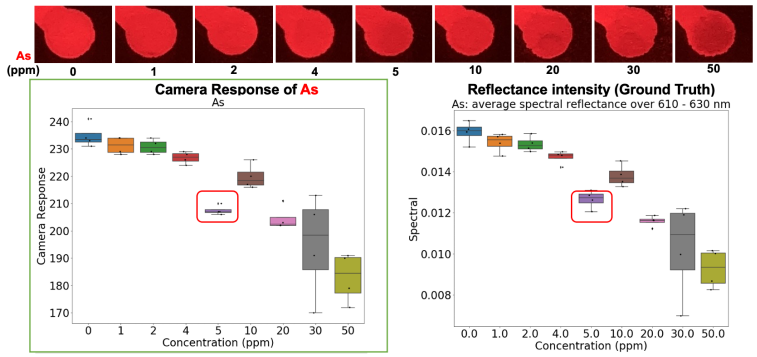



***Figure 11.*** *(Left) Box plot of the camera response with a bandpass filter (620nm) for different $As^{3+}$ concentrations. (Right) Box plot of the corresponding spectral radiance measured by using PR 705.*

We use the average R value of the colorimetric response under the bandpass filter to represent the camera response. Figure 11 shows the camera response for various concentrations of $As^{3+}$ from 0 ppm to 50 ppm. According to the data collected, it is found that the variable camera response had a very similar trend to the spectral radiance at $As^{3+}$ concentrations from 0 to 50 ppm. Our preliminary result suggests that the phone-based narrow-band spectral imaging system can replace the spectroradiometer for differentiating different contaminant levels.

## 6. Conclusion

In this paper, we investigated how to improve the accuracy of our previously developed paper-based devices for detecting and measuring heavy metal contaminants ($As^{3+}$ and $Hg^{2+}$) in food or liquids. Specifically, we considered the use of the spectral reflectance of the sensor pad, as opposed to our baseline method which simply conputes $\Delta E$ from a white background, and optimally quantizes these responses into five groups. We described a laboratory set-up for capturing the spectral reflectances of the detection devices, including an investigation of three possible types of illumination. Having chosen an LED as the best source of illuminantion, we then developed two different machine learning approaches for classifying the level of contamination by $As^{3+}$ into one of five categories: k-nearest-neighbor with sequential forward feature selection to determine the best number of features, and random forest with principal component analysis for feature reduction. We found that the latter yields the best performance. Finally, we compared the spectral responses, as a function of contaminant level, of the sensor pads within the band 610-630 nm measured with our spectroradiometer to the spectral responses captured by a mobile phone with an inexpensive narrowband filter attached to the front of the camera lens. Based on the similarity of the responses between these two capture modalities, we conclude that the mobile phone narrowband filter combination could be used as an inexpensive means of accurately measuring heavy metal contaminant levels, as indicated by the color change in our paper-based sensor device.

## Acknowledgments

This manuscript is based upon work supported by the U.S. Department of Agriculture, Agricultural Research Service, under Agreement No. 59-8072-6-001. Any opinions, findings, conclusion, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

## References

[1] G. Mance, "Pollution threat of heavy metals in aquatic environments", *Springer Science & Business Media*, December 2012.

[2] M.R. Saidur, A.A. Aziz, and W.J. Basirun, "Recent advances in DNA-based electrochemical biosensors for heavy metal ion detection: a review", *Biosensors and Bioelectronics*, 90, pp. 125-139, April 2017.

[3] S. Diaz Amaya, M. Zhao, J. Allebach, G.T.C Chiu, and L. Stanciu, "Ionic strength influences on biofunctional Au-decorated microparticles for enhanced performance in multiplexed colorimetric sensors", *ACS Applied Materials & Interfaces*, 12(29), pp. 32397-32409, July 2020.

[4] M. Zhao, S. Diaz Amaya, L.K. Lin, S.A. Jin, L.K. Lin, A.J. Deering, L. Stanciu, G.T.C Chiu, and J. Allebach, "Image Analytics for Food Safety", *Proceedings of Imaging and Multimedia Analytics in a Web and Mobile World 2020 (part of Electronic Imaging 2020)*, pp. 302(1-8), January 2020.

[5] M. Zhao, S. Diaz Amaya, S.A. Jin, L.K. Lin, A.J. Deering, L. Stanciu, G.T.C Chiu, and J. Allebach, "Inkjet printing platforms for DNA-based pathogen detection", *NIP & Digital Fabrication Conference*, 2018, pp. 107-112, September 2018.

[6] M. Zhao, R. Zhang, S. Diaz Amaya, S.A. Jin, L.K. Lin, A.J. Deering, L. Stanciu, G.T.C Chiu, and J. Allebach, "Detection, imaging, and quantification of DNA-based pathogen based on inkjet-printed test strips", *NIP & Digital Fabrication Conference*, 2019, pp. 177-181, Sepetember 2019.

[7] S. Lloyd, "Least squares quantization in PCM", *IEEE Transactions on Information Theory*, 28(2), pp. 129-137, March 1982.

[8] J. Max, "Quantizing for minimum distortion", *IRE Transactions on Information Theory*, 6(1), pp. 7-12, March 1960.

[9] A. Bierman, M.G. Figueiro, and M.S. Rea, "Measuring and predicting eyelid spectral transmittance", *Journal of Biomedical Optics*, 16(6), pp. 067011(1-8), June 2011.

[10] A.G. Wee, D.T. Lindsey, S. Kuo, and W.M. Johnston, "Color accuracy of commercial digital cameras for use in dentistry", *Dental Materials*, 2(6), pp. 553-559, June 2006.

[11] K. Fukunaga, "Introduction to statistical pattern recognition", *Elsevier*, 2013.

[12] N.S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression", *The American Statistician*, 46(3), pp. 175-185, August 1992.

[13] A. Liaw, and M. Wiener, "Classification and regression by randomForest", *R News*, 2(3), pp. 18-22, December 2002.

[14] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection", *Pattern Recognition Letters*, 15(11), pp. 1119-1125, November 1994.

[15] H. Hotelling, "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, 24(6), pp. 417-439, September 1933.

[16] M.L. Zhang, and Z.H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning", *Pattern Recognition*, 40(7), pp. 2038-2048, July 2007.

[17] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors", *IEEE Transactions on Neural Networks and Learning Systems*, 29(5), pp. 1774-1785, April 2017.

[18] A. Amjad, R. Ullah, S. Khan, M. Bilal, and A. Khan, "Raman spectroscopy based analysis of milk using random forest classification", *Vibrational Spectroscopy*, 99, pp. 124-129, November 2018.

[19] L. Breiman, "Random forests", *Machine Learning*, 45(1), pp. 5-32, October 2001.

## Author Biography

***Min Zhao*** *is a graduate student in Electrical and Computer Engineering Department at Purdue University under the supervision of Professor Jan P. Allebach. She received her B.S. degree in Mechanical Engineering in 2010 from Qingdao Technology University, Shandong, China. Her current research interests include image processing, computer vision, machine learning, and deep learning.*

***Qiyue Liang*** *is pursuing a Ph.D degree in electrical engineering at Purdue University under the supervision of Professor Jan P. Allebach. She received her B.S of science in electrical engineering degree from Purdue University (2018). Her current research interests include image processing and machine learning.*

***Susana Diaz-Amaya****, Ph.D. in materials engineering from Purdue University with a solid background in microbiology from the Pontifical Javeriana University, Colombia. Susana started her industrial career in 2011, as leader of R&D and independent consultant for business acceleration. Currently, she serves Bayer Crop Science as the head of the innovation site "Bayer at Convergence".*

***Amanda Deering*** *is a clinical associate professor of Food Science at Purdue University. Her research focuses on examining the internalization of human pathogenic bacteria in plants, and routes of contamination*

*that can contribute to plants harboring pathogenic bacteria. Amanda also works closely with industry to develop and test novel sanitization treatments that can be used for fresh produce, and works directly with Indiana fruit and vegetable growers to address food safety issues on the farm.*

***Lia Stanciu** is a professor of Materials Engineering at Purdue University. She earned her Ph.D. in Materials Science from U.C. Davis in 2003. In 2005, she joined Purdue's School of Materials Engineering as a faculty member. Her main research areas include advanced biomaterials for implantation, materials for biosensing devices and nanotechnology.*

***George T. Chiu** is a Professor of Mechanical Engineering with courtesy appointments in Electrical and Computer Engineering and Psychological Sciences at Purdue University. He received the B.S. degree from National Taiwan University and the M.S. and Ph.D. degrees from University of California at Berkeley. His research interests are mechatronics and control with applications to digital printing and imaging systems, digital fabrications and functional printing. He is a Fellow of ASME and IS&T.*

***Jan P. Allebach** is Hewlett-Packard Distinguished Professor of Electrical and Computer Engineering at Purdue University. Allebach was named Electronic Imaging Scientist of the Year by IS&T and SPIE, and was named Honorary Member of IS&T, the highest award that IS&T bestows. He has received the IEEE Daniel E. Noble Award, the IS&T/OSA Edwin Land Medal, the IS&T Johann Gutenberg Prize, is a Fellow of the National Academy of Inventors, and is a member of the National Academy of Engineering.*

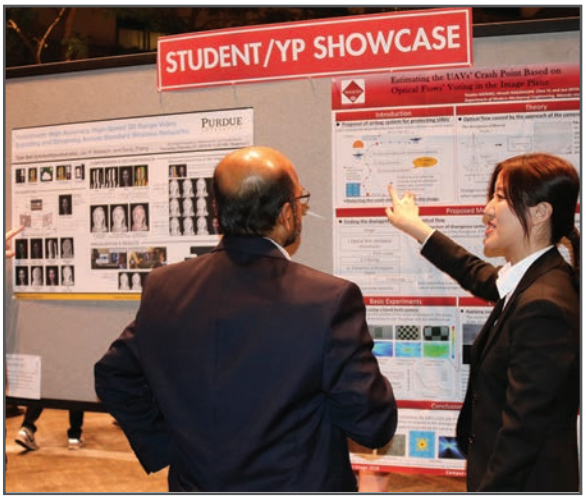
# JOIN US AT THE NEXT EI!

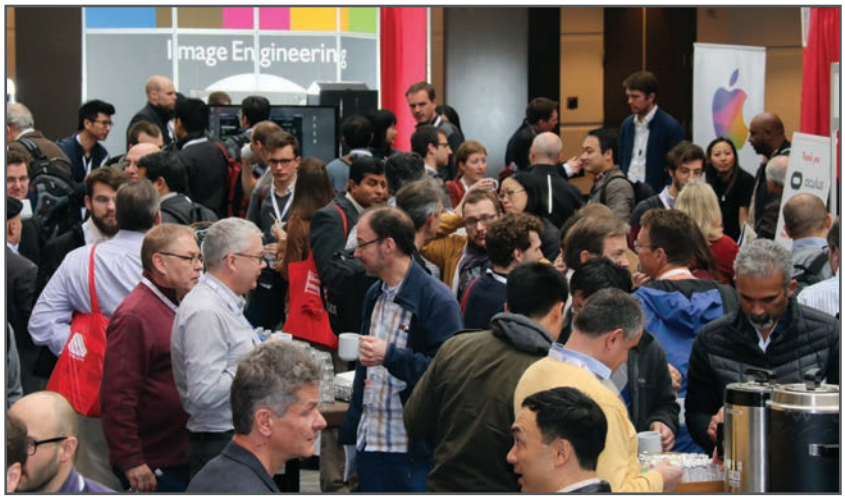
IS&T International Symposium on

# Electronic Imaging

SCIENCE AND TECHNOLOGY

*Imaging across applications . . . Where industry and academia meet!*




• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

**www.electronicimaging.org**

IS&T
imaging.org