1 **Colorimetric detection on paper analytical device using machine learning**

2

3 Bidur Khanal[1,2], Pravin Pokhrel[1], Bishesh Khanal[2*], Basant Giri[1*]

4 *1Center for Analytical Sciences, Kathmandu Institute of Applied Sciences, Kathmandu, Nepal*

5 *2NepAl Applied Mathematics and Informatics Institute for Research, Kathmandu, Nepal*

6

7 ***For correspondence**

8 Email: bgiri@kias.org.np (Dr. Basant Giri)

9 Email: bishesh.khanal@naamii.org.np (Dr. Bishesh Khanal)

10

11

12

13

14

15

16

17

18

19

20

21

Abstract

Paper-based analytical devices (PADs) employing colorimetric detection and smartphone images have gained wider acceptance in a variety of measurement applications. The PADs are primarily meant to be used in field settings where assay and imaging conditions greatly vary resulting in less accurate results. Recently, machine learning (ML) assisted models have been used in image analysis. We evaluated a combinations of four ML models - logistic regression, support vector machine, random forest, and artificial neural network, and three image color spaces - RGB, HSV, and LAB for their ability to accurately predict analyte concentrations. We used images of PADs taken at varying lighting conditions, with different cameras, and users for food color and enzyme inhibition assays to create training and test datasets. Prediction accuracy was higher for food color than enzyme inhibition assays in most of the ML model and colorspace combinations. All models better predicted coarse level classification than fine grained concentration labels. ML models using sample color along with a reference color increased the models' ability in predicting the result in which the reference color may have partially factored out the variation in ambient assay and imaging conditions. The best concentration label prediction accuracy obtained for food color was 0.966 when using ANN model and LAB colorspace. The accuracy for enzyme inhibition assay was 0.908 when using SVM model and LAB colorspace. Appropriate model and colorspace combinations can be useful to analyze large numbers of samples on PADs as a powerful low-cost quick field-testing tool.
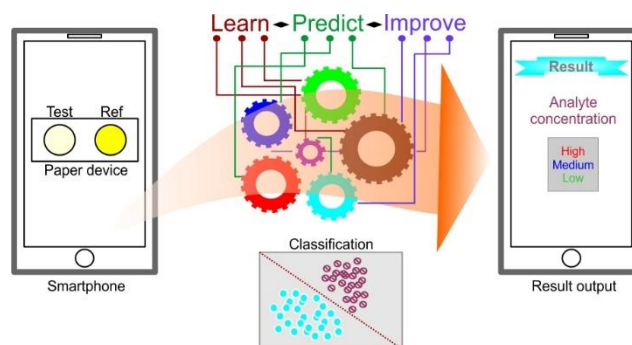
Introduction

Paper-based analytical devices (PADs) have gained wider acceptance in clinical diagnosis, environmental pollution, and food quality monitoring and pharmaceutical quality screening among many other applications. Assays involving PADs are high throughput, less costly, easy to use, and are considered as point-of-need assays.[1–5] Electrochemical and optical detection methods are primarily used to record the assay signal on the PADs.[6] Due to the proliferation of digital cameras, particularly the smartphone cameras, digital image-based colorimetric detection method is one of the widely used methods where color information encoded in the digital image is used for quantitative analysis.[3] Smartphone image-based colorimetric detection could be a cost-effective and an attractive field-based alternative to conventional techniques such as spectrophotometers, colorimeters or fluorometers.[7,8]

Digital cameras use multiple charge-coupled device (CCD) or complementary metal oxide semiconductor (CMOS) sensors to capture the light intensity signal separately from the primary colors Red (R), Green (G) and Blue (B) using a mosaic patterned filter array. The signals are then combined using demosaicing, resulting in three color values R, G and B at each pixel of the digital image.[9] The image formation process in digital cameras is non-linear. The raw signal or the intensity value at each pixel of the imaged area depends on the lighting condition, the sensor sensitivity, the distance between the object and camera, and the reflectance property of the object being imaged.[10,11] Some of these variations such as lighting condition, object-camera distance can be minimized by using a controlled environment that is only possible in lab settings.[12–15] However, the PADs are ultimately meant to be used in field settings by a minimally trained user during which criteria of controlled imaging conditions may not be achieved resulting in errors.

Attaching an extra device on smartphones is possible to maintain a constant source of

71   illumination at the cost of higher price and same device may not be applicable for all

72   smartphones having diverse shapes and sizes.[16] Another approach uses a blank or reference

73   assay along with sample assay at the same time to factor out the impact of illumination and

74   camera quality changes.[10,17] Since the raw signal captured by the camera sensors process non-

75   linearly during image acquisition before saving to the memory, above mentioned approaches

76   only partially address the problem. Furthermore, image auto-correction options such as

77   automatic exposure correction, color correction based on ambient light selection, white

78   balancing, and contrast enhancement can highly influence the overall color calibration.[16,28] Thus,

79   estimating the analyte concentration from color intensity in PADs is an inverse problem where

80   any estimation models will have explicit or implicit assumptions on the image formation

81   process. To make robust and reliable low-cost models that can work in diverse point-of-need

82   settings, one needs to account for the non-linear factors.

83   Traditional computer vision and machine learning based models have been used for

84   image processing to enhance the robustness of the PADs assay. Traditional computer vision

85   based algorithms try to be invariant to illumination, scale, and camera.[10] Most of these

86   algorithms use color spaces including the red, green and blue (RGB); hue, saturation and value

87   (HSV); hue, saturation and lightness (HSL); and CIE L*a*b*.[10,16,18] Each particular application

88   usually requires careful selection of color space model based on preliminary data and

89   experiments. Similarly, other corrections such as white balance correction, contrast transfer and

90   gamma correction have been used but these corrections are specific for individual camera.

91   These camera-specific corrections are not practical when the goal is to enable low-cost

92   colorimetry to the large variety of consumer cameras.

93   In recent years, data-driven machine learning (ML) algorithms are getting increasingly

94   common for colorimetric detection.[19,20] ML has a potential to be robust against unwanted

95   variation as one does not need to explicitly design an algorithm to extract information but let the

96 model learn from the data to work in diverse environmental settings.[21] Bao et al. trained Support

97 Vector Machines (SVM) on RGB color channels in multiple indoor settings using a single

98 camera.[22] Similarly, Solmaz et al. used Least-Squares Support-Vector Machine (LS- SVM) and

99 multi-class Random Forest classifier to predict per-oxide content on colorimetric test strips and

100 report over 90% accuracy for 6 classes with inter-phone repeatability under versatile

101 illumination.[23] Kim et al. applied Linear Discriminant Analysis (LDA), Support-Vector

102 Machine (SVM) and Artificial Neural Network (ANN) for colorimetric analysis of saliva–

103 alcohol concentrations, with average cross-validation accuracy rates of 100% and 80% for the

104 standard and enhanced concentrations.[19] Even though few papers reported the use of ML in

105 image-based colorimetric detection, they lack insights on the relative efficacy of various ML

106 algorithms and their generalization capabilities. Generalization is an important issue in ML,

107 where a model's high performance in training and validation data degrades in test data with

108 different distribution. As reported by Morbioli et al.[24], most of the proposed ML models for

109 colorimetric detection with PADs do not release their source code and data used, making it

110 difficult to reproduce results and perform benchmark comparisons.

111 In this work, we designed a set of comprehensive experiments to analyze the

112 performance and utility of ML for colorimetric image analysis of PADs using two different data

113 sets – food color and enzyme inhibition assay for pesticide residue determination. We assessed

114 four different ML models - logistic regression (LR), support vector machine (SVM), random

115 forest (RF), and artificial neural network (ANN) - and three image color spaces - RGB, HSV,

116 and LAB - to predict the target analyte concentration using images of PADs taken at varying

117 lighting conditions, with different cameras, and users. The colorimetric assays involved in our

118 approach included both target analyte and a reference assay zone on the paper device in contrast

119 to most of the previous works that captured only the target sample image when estimating its

120 concentration. In this setting, we obtained the concentration label prediction accuracy of 0.966

121   and 0.908 for food color and pesticide residue analysis datasets, respectively in separate real-

122   life test datasets. We also highlight the limitations of all ML models when there is domain shift

123   in test data, and their inability to predict with the same accuracy at all concentration levels. We

124   have made our source code and data publicly available to contribute to the reproducibility issues

125   in this rapidly progressing field.

Experimental

Fabrication of paper analytic device

128   We designed a layout of circular patterns in computer and printed on Whatman No 1

129   grade filter paper using Xerox ColorQube 8580 solid wax printer.[25] The wax printed paper was

130   heated from backside by pressing with a dry clothing iron on its surface. The backside of the

131   PADs was laminated to prevent liquid from spilling through the other side of the paper. Finally,

132   the paper sheet was cut in such a way that each PADs contained two circular assay regions as

133   reference and sample zones (*see* Fig. 1A).

Classification datasets

135   We prepared datasets for two different assays – first one using food color and second

136   one using pesticide assay. Each dataset used four different smartphones (Huawei SCC-U21,

137   iPhone 6, Honor 8C, and Samsung Galaxy J7 Max) for images acquisition at different lighting

138   conditions, camera to PAD distance, and capture angle. The lighting conditions included

139   outdoor sunlight, indoor daylight, fluorescent light, incandescent light, and combination of

140   them. A general procedure of assay on a paper device is given in Fig. 1B.

141   The food color data sets were obtained by loading food color (Foster Clark Product Ltd.)

142   solutions of ten different dilutions onto the PADs. We captured 2400 images in total under

143   various conditions. Images that were unclear or blurred were removed and 2353 images were

144   used for training ML models. The images were labeled from 1 to 10 - 1 for highest concentration

6

145 and 10 for lowest concentration of food color. The dataset contained approximately equal

146 number of images per label. New set of 600 images of the same food color concentrations were

147 obtained and were used as test data set. These images were taken in a different day - varying

148 the illumination, randomly changing the camera, and camera-PAD distance – to create different

149 test datasets from training datasets. Representative images are shown in Fig. 1C.
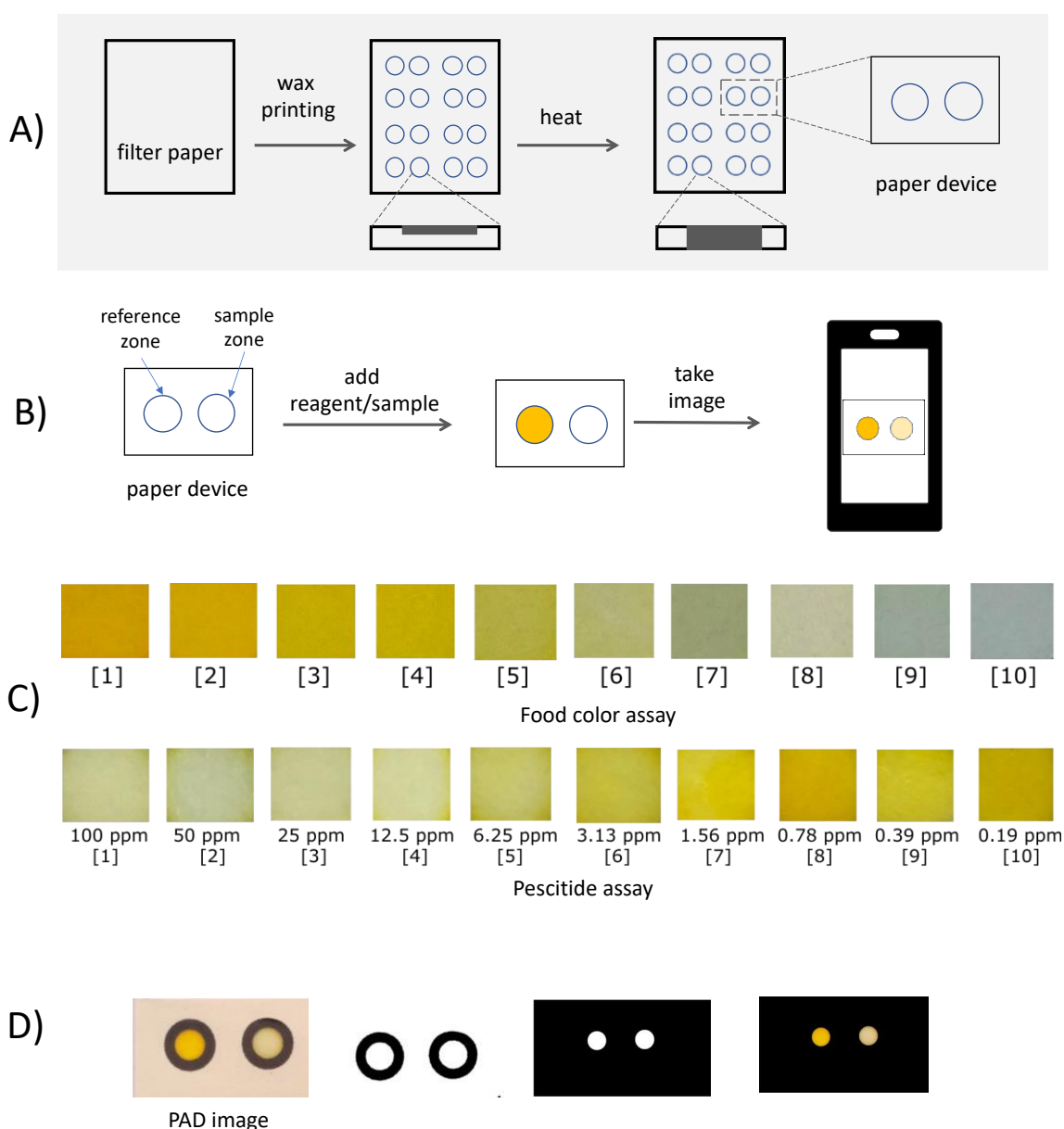
150 *Figure 1: (A) Fabrication of paper device. Solid wax was printed on Whatman filter paper*

151 *which penetrated through the paper after heating (see, insert illustrations) (B) General*

152 *procedure for paper-based pesticide assay. (C) Representative images of paper device test*

153 *zones after assays were performed. Ten different dilutions of food color and ten different*

154 *concentrations of pesticide were used. Since pesticide assay followed enzyme inhibition*

155 *reaction, the concentrations and color intensity are inversely corelated. (D) Automatic color*

156 *pixel extraction procedure from the PADs: left to right represent binary thresholding, mask*

157 *generation and masked region of interest (ROI).*

158   The second dataset prepared was for a more realistic application which included enzyme

159 inhibition assay for pesticide residue measurement.[26] In this assay, the acetylcholinesterase

160 enzyme (Sigma-Aldrich) breaks down the acetylthiocholine chloride (AtCh) substrate (Sigma-

161 Aldrich) into thiocholine and acetic acid. The thiocholine molecules react with Ellman's reagent

162 (Dithiobisnitrobenzoic acid - DTNB) (Sigma-Aldrich) to give a yellow-colored product of

163 thionitrobenzoic acid.[26] In the presence of organophosphate and carbamate group of pesticides,

164 the enzyme activity is inhibited. Based on the extent of inhibition, amount of pesticide on the

165 sample is estimated. The intensity of yellow color produced is indirectly proportional to the

166 concentration of pesticide in the sample. Since no pesticide is added to the reference assay zone,

167 it has high color intensity.[1,27] Images of PADs were captured at 10 minutes of the enzyme

168 reaction using a smartphone for further analysis (*see* Fig. 1B for general outline). We collected

169 1872 images for training data sets by repeating the same experiment in multiple days and

170 different light conditions using 4 different smartphones. Each image was given a label from 1

171 to 10 similar to food color assay images. Label 1 represents a pesticide (Paraoxon, Sigma-

172 Aldrich) concentration of 100 ppm. Same pesticide solution was serially diluted half in the

173 remaining assays. New experiments at different lighting conditions were performed to obtain

174 601 new images as test datasets. See Fig. 1C for representative images of pesticide assay.

175

176

Extraction of pixel values from assay images

The pre-processing of a PAD image is outlined in Fig. 1D. The leftmost image in panel 1D is a typical image of the PAD captured using a camera. The two regions encircled by black rings are the regions of interest (ROIs) that contain the color information of the reference and target samples. We developed an automatic threshold-based segmentation algorithm to extract all the pixels lying in these two regions. RGB images were converted to grayscale images which were then converted to binary image by applying a threshold $T = 0.8 \cdot I_m + 30$, where I_m is the mean intensity of an image converted to grayscale, 0.8 and 30 are empirically chosen values after visual inspection across multiple images. The binary images provide us the masks which were used to extract the pixel values lying in the two ROIs of the corresponding original images as shown in the rightmost image of Fig. 1D.

Evaluation of multi-class classification models

We used Average Classification Accuracy (ACA), which is a commonly used metric for multiclass classification problems. ACA is a ratio of the total number of correct predictions to the total number of predictions $\text{ACA} = \frac{Total\ no\ correct\ predictions}{Total\ no\ of\ predictions}$. In addition, we visualized the results using confusion matrices which provides information on how many samples of a particular class are misclassified to another class. We used a 5-fold cross validation where the training dataset was randomly split into 5 subsets (folds), and the model was trained five times such that each time a unique fold was selected for validation and the remaining four for training. The mean ACA and its standard deviation were reported for the cross-validation experiments. In order to evaluate the robustness of ML models and its generalization ability, we also evaluated the models with a separate test set under different conditions trying to emulate actual real-life testing scenarios.

200   We used various approaches to extract features and feed them as input to the ML

201   models. We compare these approaches using the following two experimental setups: (a) Color

202   spaces: RGB vs. HSV vs. LAB color representation; (b) Mean pixel value of ROIs for each

203   channel vs. using all the pixel values of the down sampled ROIs. Similarly, we assessed the

204   impact of reference test region by using the third setup (c): Using only target sample vs. using

205   both the reference and target sample.

206   After extracting the mean color intensity of the sample and reference assay zones, we

207   obtained $2 \times 3 = 6$ unique values for each assay (two circular zones and the three-color

208   channels). This can be fed as a 6-dimensional feature vector to ML classification models. The

209   mean values from ROI do not capture the variation of pixel values within the ROI. However,

210   using all the pixels of the ROIs as input features to train ML models dramatically increases the

211   feature dimension which computationally affects the training of some ML algorithms such as

212   SVM. As most of the PADs colorimetry images only have color information without texture

213   and shape information, as a compromise, we downsampled the cropped image into a size of 16

214   x 16 and converted it to a 1D Vector of dimension $16 \times 16 \times 3 = 768$ (for three color channels)

215   for each reference and targe ROIs, resulting in two 1D vectors.

216   We compared following four most widely used supervised ML models for colorimetry

217   with PADs: logistic regression[28], support vector machine (SVM)[29], random forest[30] and

218   artificial neural network (ANN)[31]. We used python-based ML library, scikit-learn[32] to build

219   LR, SVM and RF models. For ANN, we adopted a popular python-based framework Keras[33].

220   Details of implementation configuration is given in Table 1.

221

Table 1: Implementation details of machine learning models

Model type	Library or framework	Implementation details of model
Logistic Regression	Scikit-learn	Multi-nominal + L2 penalty, L-BFGS solver, maximum iteration = 10000
Support Vector Machine	Scikit-learn	Squared L2 penalty, Linear Kernel
Random Forrest	Scikit-learn	no. of trees = 1000, split criterion = gini
Artificial Neural Network (Fully connected)	Keras	3-Dense Layers, sigmoid + softmax activation, adam optimizer, categorical cross-entropy loss
Artificial Neural Network (Convolutional + Fully connected)	Keras	2*(2-Conv Layer + Max pooling) Layers, 3-Dense Layers, relu + softmax activation, adam optimizer, categorical cross-entropy loss

223

Results and discussion

225   We established a baseline result by using sample color and using a mean color intensity as

226 input feature to ML models. Cross validation results for various color spaces and ML models

227 using mean pixel values of the ROIs as input features to classify into 10 concentration labels

228 are shown in Fig. 2. At first, we tested the models using test zones of samples only. In general,

229 RF model yielded higher accuracy in all three-color spaces in case of food color experiments.

230   In this case, HSV color space had higher (0.691) value than LAB (0.669) and RGB (0.588).

231   The LR and SVM models gave lower accuracy compared to RF model. In both LR and SVM

232   models, all three-color spaces had similar values compared to RF. We also looked at the ANN

233   model's ability to predict correct assay values. It gave comparable values with LR and SVM

234   models but the values for three different color spaces had some variation. In case of pesticide

235   assay, the accuracy was lower in all combinations of models and color spaces. However, the

236   trend was similar to food color assay.

237   **Combining sample and reference assay color**

238   To improve the accuracy of prediction, we took into consideration of both sample and

239   reference assay color. In these experiments, we looked into if imaging both reference or control

240   assay and sample assay at the same time could improve the prediction accuracy by ML models.

241   The later approach showed an improvement in the prediction accuracy of all the ML models

242   when the mean color intensity from the reference region was included as the feature (*see* Fig.2).

243   We believe that the ML model is able to use information from the reference sample region to

244   partially factor out the variations due to ambient conditions and camera parameters, improving

245   the result when using the reference sample compared to using only the target sample.

246   The cross-validation accuracy using RGB color space for food color assay was found to

247   be higher in LR, SVM and ANN models while it was lower in RF model when compared to LAB

248   and HSV color spaces. The HSV color space in RF and RGB color space in ANN models

249   showed higher accuracy. However, we did not find any specific trends among color spaces and

250   models used. Similar results were observed for pesticide assay dataset. It is interesting to note

251   that the HSV color space using Random Forest showed highest average accuracy in comparison

252   to other color models (i.e., 0.804 in food color and 0.596 in pesticide). In LR, the highest average

253   accuracy was observed in RGB color model (0.684 in food color and 0.406 in pesticide assay).

254   Similarly, in SVM, the highest average accuracy was in LAB with 0.68 for food color and 0.42

255   for pesticide. For ANN, the highest average CV accuracy was in RGB, i.e., 0.721 for food color

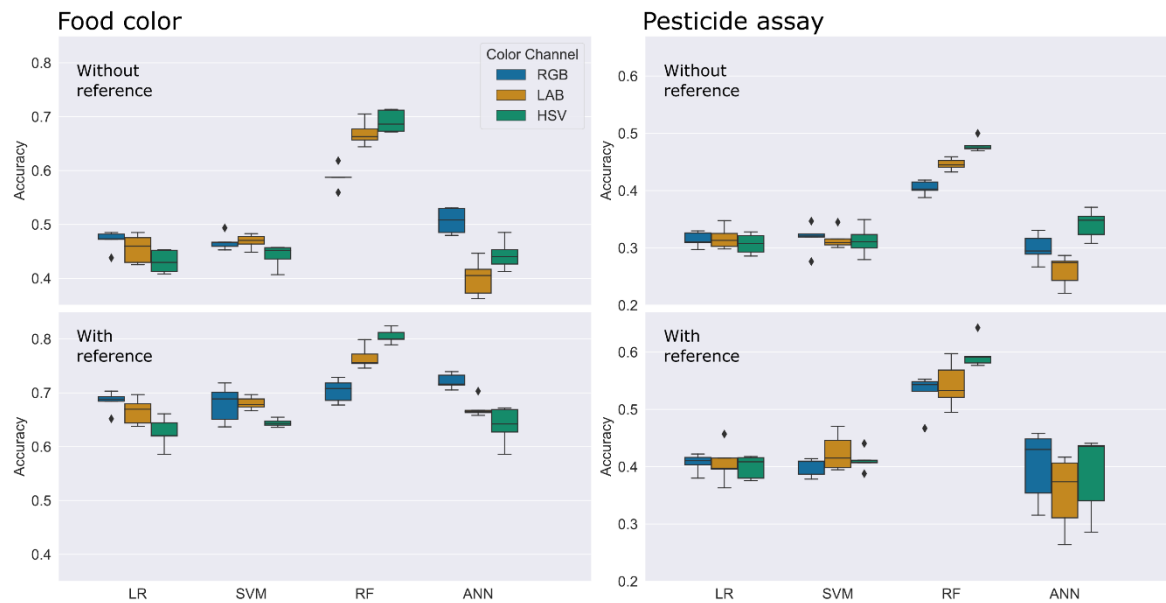256   and 0.401 for pesticide.

257

258   *Figure 2: Cross validation accuracy results obtained for both food color prediction (left*

259   *panel) and pesticide assay (right panel). Each box describes the full range of variation*

260   *(whisker's height), the likely range of variation (box height), and median (horizontal line*

261   *within the box) in the classification accuracy score of five cross-validation folds. All results in*

262   *this figure were obtained using mean of each color channel values of sample assays without*

263   *(top) and with (bottom) reference assay test zones.*

264   Along with cross-validation experiments, we also evaluated all the models and color

265   spaces in separate test dataset. The highest test accuracy was 0.67 with HSV color space in

266   SVM for food color and 0.34 with HSV color space in ANN for pesticide (in ANN with HSV).

267   The results, as expected, showed that test results did not necessarily agree with the cross-

268   validation accuracy. Here, HSV color model showed good results in both sets of assays. In

269   addition, a large drop in test accuracy was observed across all the models compared to the cross-

270   validation results. This highlights that reporting only cross-validation scores or scores in test

271   data that are very similar to training set can overestimate ML model's performance. ML

272   models' performance can severely degrade when the statistical distribution of the test data is

273   different from that of the training set. This is closer to the real-world scenario we wanted to

274   emulate in order to further assess the efficacy and applicability of ML systems when test images

275   are captured in different field settings. When comparing the results across food color and

276   pesticide, we observed that the accuracy in pesticide assay is relatively lower than that of food

277   color for all the models.

278   The confusion matrix in Fig. 3 shows food color has its values clustered near the diagonal

279   line expect for labels 8, 9 and 10. These labels correspond to low concentration values with faint

280   colors (best viewed in color image). For pesticide assay, the values in the initial labels (1,2,3,4)

281   and final labels (8,9,10) are misclassified in large numbers. The matrix fields in the middle,

282   though not accurate, show a diagonal pattern. This result seems to follow a typical *S-shaped*
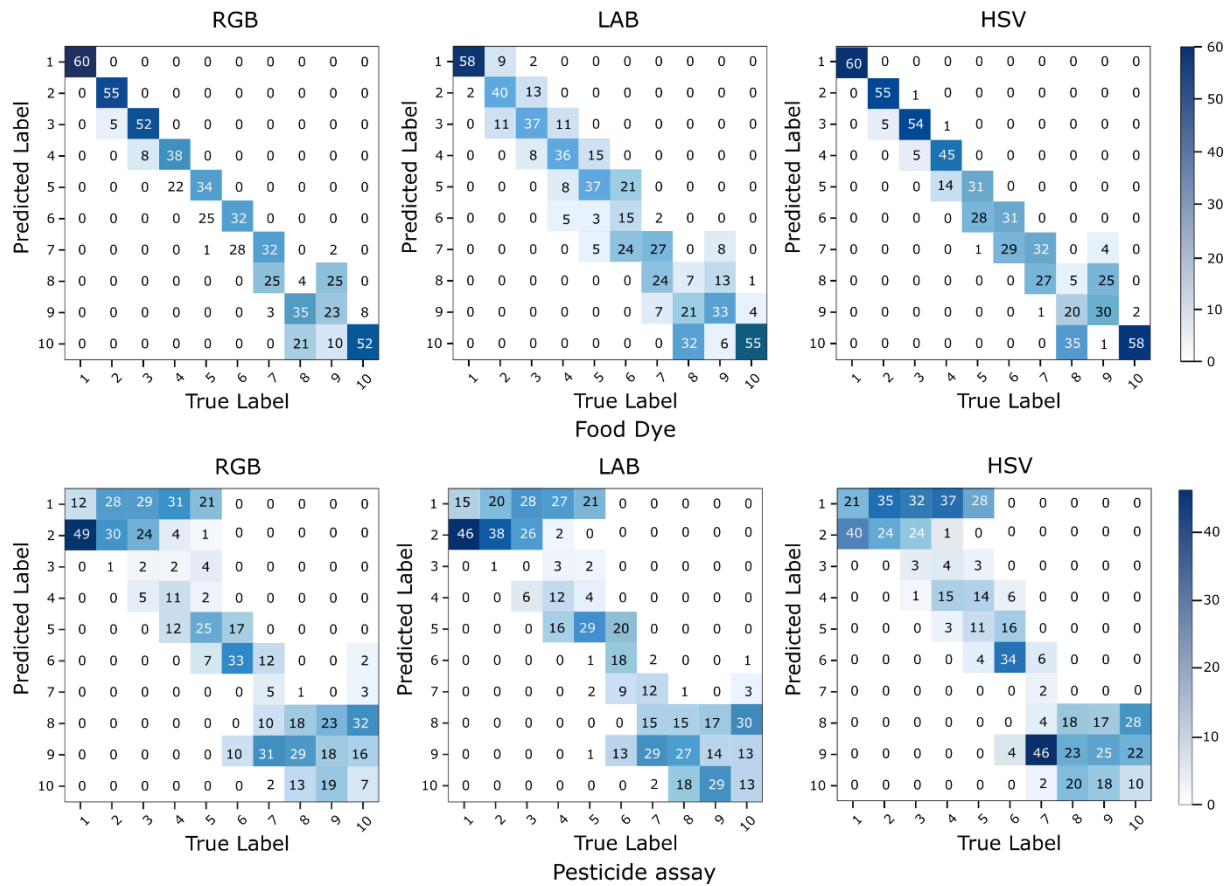
283   enzyme assay curve.[34]

Figure 3: Confusion Matrix (of test accuracy of Logistic Regression) for Food Color and Pesticide Assay when RGB, LAB, HSV color models are used respectively. The Confusion Matrix of other models also show similar patterns.

Input feature vectors from downsampled image

Fig. 4 shows the CV accuracy using all the pixels of downsampled ROIs as a feature vector for ML models. Since images containing both sample and reference colors provided better accuracy than using sample color only, we used the former approach in this experiment. CV accuracy is generally higher when using all pixels from a downsampled image as input features compared to when using color channel means. The improvement in the accuracy was observed in most of the models and color spaces. However, the extent of improvement varied with the models and color spaces tested. Similar trend was observed both in food color assay and pesticide assay. LR and SVM models with RGB and LAB color spaces resulted in the

15

297 highest accuracies. For LR, CV accuracy was 0.975 and 0.977 when using RGB and LAB

298 respectively. Likewise, for SVM, CV accuracy was 0.971 when using RGB or LAB. When the

299 test dataset was evaluated, the test accuracy was lower in both food color and pesticide.

300 However, the gap between cross-validation and test accuracy in pesticide is higher for pesticide

301 in all the color channels and models. (*see* Table SI1).

302
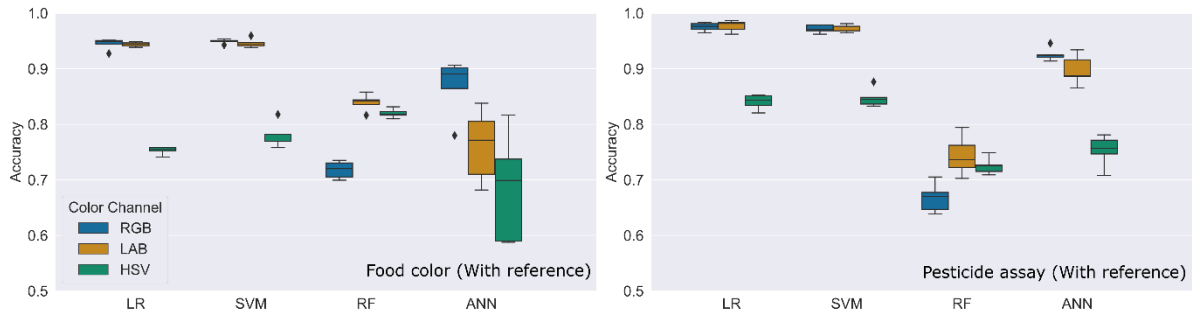
303 *Figure 4: Cross-validation accuracy using all the pixel values from 16*16 downsampled*

304 *images of reference and sample as input features. Each box describes the full range of variation*

305 *(whisker's height), the likely range of variation (box height), and Median (horizontal line*

306 *within box) in the accuracy score of 5 cross-validation folds. LR, SVM, RF and ANN are*

307 *implemented in 3 color models: RGB, LAB, and HSV.*

308 **Classification into high, medium, and low**

309   The results in the previous section show that accurately and robustly estimating fine-

310 grained concentration labels is difficult even with powerful ML models using 10 concentration

311 labels. Therefore, we merged 10 concentration classes into 3 distinct labels: high, medium, and

312 low for semi-quantitative prediction of both the food dye and pesticide samples. Labels 1, 2,

313 and 3 shown in Fig.1C were merged into high, 4, 5, 6, and 7 were merged into medium and 8,

314 9, and 10 were merged into low category in case of food color assay. In case of pesticide assay,

315 labels 1, 2, 3, 4 in Fig. 1C were merged into high, 5 and 6 into medium and 7, 8, 9, 10 into low
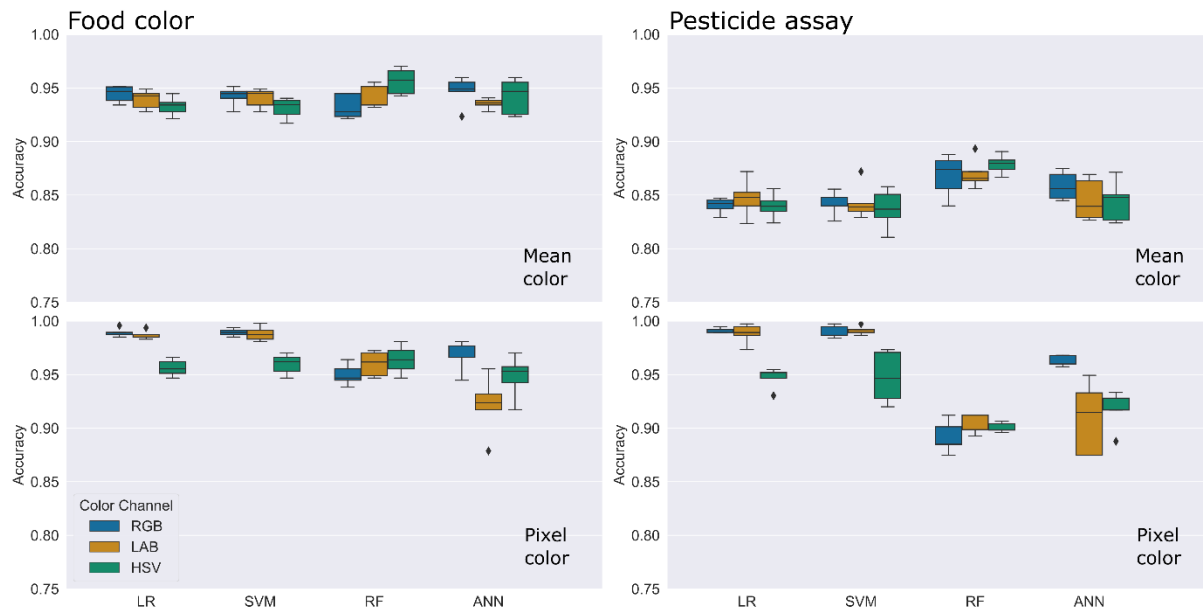
316 categories.

Figure 5: Cross-validation accuracy for three reduced classes of high, medium, low. Results

of both mean color and 16*16 downscaled RGB Pixels are presented for food color (left

panel) and pesticide assay (right panel).

Fig. 5 shows the overall CV accuracy in food dye and pesticide datasets when they were

reduced into three classes of high, medium, and low. We found that the food color dataset when

individual means of 3-color channels were used as input features showed similar accuracy values

with all four models and all three-color spaces. CV accuracy values for pesticide assay with mean input

features produced similar results as with food color but slightly lower values. Most of the models and

color spaces in both food color and pesticide assays produced better accuracies when using all

the pixels from the downsampled image as input features.

To emulate a realistic setting and test generalization capability, we evaluated all the

models using test dataset. Table 2 shows the average test accuracies, where we see that in

general, the food color dye dataset exhibited higher classification accuracy compared to the

pesticide dataset. For food color, we observe the highest accuracy of 0.966 in ANN with LAB

color space and input feature from all color pixels of the downsampled image. However, the

models for pesticide concentration prediction did not benefit much by using all the pixels as input

334 features. For pesticide, the highest accuracy of 0.908 was obtained in SVM with LAB color space

335 and input feature from individual means of each color channel. The obtained results show that

336 accurate semi-quantitative prediction of concentration from PAD images is possible even in

337 uncontrolled setup with enough data and suitable ML model.

338 *Table 2: Average accuracy using 3 concentration labels (high, medium, low). A) all color*

339 *pixels from 3-channels of downsampled 16*16 image as input feature and B) individual mean*

340 *of each color channels (3 means) as input feature. All images included sample and reference*

341 *assays.*

A)

	Food Dye			Pesticide		
	RGB	HSV	LAB	RGB	HSV	LAB
LR	0.940	0.921	0.946	0.753	0.637	0.743
SVM	0.936	0.926	0.938	0.738	0.674	0.774
RF	0.901	0.933	0.940	0.865	0.875	0.870
ANN	0.960	0.946	0.966	0.851	0.8336	0.778

B)

	Food Dye			Pesticide		
	RGB	HSV	LAB	RGB	HSV	LAB
LR	0.933	0.936	0.928	0.900	0.873	0.895
SVM	0.936	0.941	0.936	0.903	0.883	0.908
RF	0.898	0.915	0.928	0.837	0.855	0.852
ANN	0.893	0.868	0.926	0.866	0.868	0.901

342

Conclusions

We evaluated four ML models for their ability to accurately predict the concentration of target analytes on paper device platform. We found that the ML models that used sample color along with a reference color increased the models' ability in predicting the result. In such cases, the ML models can utilize information from the reference region to partially factor out the variations of ambient lighting conditions and image acquisition setup and learn to estimate the concentration level by looking at the differences in the signal color of the two regions. Using a printed reference color instead of a reference assay performed on the same paper device may not correct the variation resulting from the assay procedure. The relationship between reference and sample color could be a simple difference or a n-degree polynomial. The reference assay color may provide a one-point calibration to estimate or predict the concentration of analyte in the given sample.

In general, we found accuracy for food color assays were higher than accuracy for pesticide assay in most of the combinations. Unlike food color, the final color in the pesticide assay is obtained by an enzyme inhibition reaction. The enzyme reaction varies with ambient environmental conditions. Such variation in assay temperature and moisture can result in inconsistent color development on the surface of paper devices. Our results show that the ML models may provide only limited accuracy when using fine grained estimation of concentration labels but provide high accuracy when using them for coarse level classification such as low, medium and high. Ability of ML models to accurately classify pesticide concentration to such three classes even in difficult real-life test images show the potential of using ML powered PADs as a low-cost quick field testing method.

Smartphone cameras allow post-processing even before we save or see images. Since it is very hard to understand and identify these individual pre-processing steps, letting the ML

models learn from the data instead of trying to build inverse models to revert the camera post-processing is a more promising approach. Convolutional Neural Networks (CNNs) have seen tremendous success in the last few years in the computer vision field. Since the colorimetric assays on paper devices do not provide variation in texture and shape, the neural networks have limited to no benefit compared to other ML models such as RF. We might be able to leverage the power of CNNs and build more accurate analyte concentration estimation methods if we can develop novel PADs that express shape and texture variation depending on the target concentration analyte.

Finally, robust ML models can be useful in analyzing large numbers of samples in applications such as environmental monitoring, clinical diagnosis during emergencies for assays involving colorimetric paper devices. Appropriate ML models integrated in smartphones that can read assay results performed on PADs platform by taking images, process or analyze the signal to accurately predict assay results, and report or store the results locally or on cloud could be powerful tools in several measurement applications.

Acknowledgement

References

(1) Fuyal, M.; Giri, B. A Combined System of Paper Device and Portable Spectrometer for the Detection of Pesticide Residues. *Food Anal. Methods* **2020**, *13* (7), 1492–1502. https://doi.org/10.1007/s12161-020-01770-y.

(2) Spica, N.; Green, M.; Lown, L.; Duwal, R.; Fuyal, M.; Giri, S.; Giri, B.; Paetkau, D.;

Lamichhane-Khadka, R. Development of a Microbiological Paper-Based Analytical Device to Detect Fecal Contamination of Water in Resource-Limited Settings. *Water. Air. Soil Pollut.* **2021**, *232* (5), 1–12.

(3) Ozer, T.; McMahon, C.; Henry, C. S. Advances in Paper-Based Analytical Devices. *Annu. Rev. Anal. Chem.* **2020**, *13* (1), 85–109. https://doi.org/10.1146/annurev-anchem-061318-114845.

(4) Sharma, N.; Barstis, T.; Giri, B. Advances in Paper-Analytical Methods for Pharmaceutical Analysis. *Eur. J. Pharm. Sci.* **2018**, *111*, 46–56. https://doi.org/10.1016/j.ejps.2017.09.031.

(5) Pokhrel, P.; Jha, S.; Giri, B. Selection of Appropriate Protein Assay Method for a Paper Microfluidics Platform. *Pract. Lab. Med.* **2020**, *21*, e00166.

(6) Fu, L.-M.; Wang, Y.-N. Detection Methods and Applications of Microfluidic Paper-Based Analytical Devices. *TrAC Trends Anal. Chem.* **2018**, *107*, 196–211. https://doi.org/10.1016/j.trac.2018.08.018.

(7) Rezazadeh, M.; Seidi, S.; Lid, M.; Pedersen-Bjergaard, S.; Yamini, Y. The Modern Role of Smartphones in Analytical Chemistry. *TrAC Trends Anal. Chem.* **2019**, *118*, 548–555. https://doi.org/10.1016/j.trac.2019.06.019.

(8) Zhao, W.; Tian, S.; Huang, L.; Liu, K.; Dong, L.; Guo, J. A Smartphone-Based Biomedical Sensory System. *Analyst* **2020**, *145* (8), 2873–2891. https://doi.org/10.1039/C9AN02294E.

(9) Li, X.; Gunturk, B.; Zhang, L. Image Demosaicing: A Systematic Survey. In *Visual Communications and Image Processing 2008*; International Society for Optics and Photonics, 2008; Vol. 6822, p 68221J. https://doi.org/10.1117/12.766768.

(10) Komatsu, T.; Mohammadi, S.; Busa, L. S. A.; Maeki, M.; Ishida, A.; Tani, H.; Tokeshi, M. Image Analysis for a Microfluidic Paper-Based Analytical Device Using the CIE L*a*b* Color System. *Analyst* **2016**, *141* (24), 6507–6509. https://doi.org/10.1039/C6AN01409G.

(11) Byrne, L.; Barker, J.; Pennarun-Thomas, G.; Diamond, D.; Edwards, S. Digital Imaging as a Detector for Generic Analytical Measurements. *TrAC Trends Anal. Chem.* **2000**, *19* (8), 517–522. https://doi.org/10.1016/S0165-9936(00)00019-4.

(12) Digital Image Analysis for the Colorimetric Determination of Aluminum, Total Iron, Nitrite and Soluble Phosphorus in Waters: Analytical Letters: Vol 50, No 2 https://www.tandfonline.com/doi/abs/10.1080/00032719.2016.1182542 (accessed 2021 - 02 -09).

(13) Masawat, P.; Harfield, A.; Namwong, A. An IPhone-Based Digital Image Colorimeter for Detecting Tetracycline in Milk. *Food Chem.* **2015**, *184*, 23–29. https://doi.org/10.1016/j.foodchem.2015.03.089.

(14) Lima, M. B.; Andrade, S. I. E.; Barreto, I. S.; Almeida, L. F.; Araújo, M. C. U. A Digital Image-Based Micro-Flow-Batch Analyzer. *Microchem. J.* **2013**, *106*, 238–243. https://doi.org/10.1016/j.microc.2012.07.010.

(15) Digital image-based titrations - ScienceDirect https://www.sciencedirect.com/science/article/abs/pii/S0003267006008634 (accessed 2021 -02 -09).

(16) Geng, Z.; Zhang, X.; Fan, Z.; Lv, X.; Su, Y.; Chen, H. Recent Progress in Optical Biosensors Based on Smartphone Platforms. *Sensors* **2017**, *17* (11). https://doi.org/10.3390/s17112449.

(17) Ra, M.; Muhammad, M. S.; Lim, C.; Han, S.; Jung, C.; Kim, W.-Y. Smartphone-Based Point-of-Care Urinalysis Under Variable Illumination. *IEEE J. Transl. Eng. Health Med.* **2018**, *6*, 1–11. https://doi.org/10.1109/JTEHM.2017.2765631.

(18) Evaluation of color mapping algorithms in different color spaces

https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9971/1/Evaluation-of-color-mapping-algorithms-in-different-color-spaces/10.1117/12.2238435.short?SSO=1 (accessed 2021 -02 -10).

(19) Kim, H.; Awofeso, O.; Choi, S.; Jung, Y.; Bae, E. Colorimetric Analysis of Saliva–Alcohol Test Strips by Smartphone-Based Instruments Using Machine-Learning Algorithms. *Appl. Opt.* **2017**, *56* (1), 84–92. https://doi.org/10.1364/AO.56.000084.

(20) Mutlu, A. Y.; Kılıç, V.; Özdemir, G. K.; Bayram, A.; Horzum, N.; Solmaz, M. E. Smartphone-Based Colorimetric Detection via Machine Learning. *Analyst* **2017**, *142* (13), 2434–2441. https://doi.org/10.1039/C7AN00741H.

(21) Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.-D. Machine Learning in Manufacturing: Advantages, Challenges, and Applications. *Prod. Manuf. Res.* **2016**, *4* (1), 23–45. https://doi.org/10.1080/21693277.2016.1192517.

(22) Bao, X.; Jiang, S.; Wang, Y.; Yu, M.; Han, J. A Remote Computing Based Point-of-Care Colorimetric Detection System with a Smartphone under Complex Ambient Light Conditions. *Analyst* **2018**, *143* (6), 1387–1395. https://doi.org/10.1039/C7AN01685A.

(23) Solmaz, M. E.; Mutlu, A. Y.; Alankus, G.; Kılıç, V.; Bayram, A.; Horzum, N. Quantifying Colorimetric Tests Using a Smartphone App Based on Machine Learning Classifiers. *Sens. Actuators B Chem.* **2018**, *255*, 1967–1973. https://doi.org/10.1016/j.snb.2017.08.220.

(24) Morbioli, G. G.; Mazzu-Nascimento, T.; Stockton, A. M.; Carrilho, E. Technical Aspects and Challenges of Colorimetric Detection with Microfluidic Paper-Based Analytical Devices (MPADs) - A Review. *Anal. Chim. Acta* **2017**, *970*, 1–22. https://doi.org/10.1016/j.aca.2017.03.037.

(25) Carrilho, E.; Martinez, A. W.; Whitesides, G. M. Understanding Wax Printing: A Simple Micropatterning Process for Paper-Based Microfluidics. *Anal. Chem.* **2009**, *81* (16), 7091–7095. https://doi.org/10.1021/ac901071p.

(26) Ellman, G. L.; Courtney, K. D.; Andres, V.; Featherstone, R. M. A New and Rapid Colorimetric Determination of Acetylcholinesterase Activity. *Biochem. Pharmacol.* **1961**, *7* (2), 88–95. https://doi.org/10.1016/0006-2952(61)90145-9.

(27) Rajangam, B.; Daniel, D. K.; Krastanov, A. I. Progress in Enzyme Inhibition Based Detection of Pesticides. *Eng. Life Sci.* **2018**, *18* (1), 4–19.

(28) *Encyclopedia of Machine Learning and Data Mining*, 2nd ed.; Sammut, C., Webb, G. I., Eds.; Springer US, 2017.

(29) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20* (3), 273–297. https://doi.org/10.1007/BF00994018.

(30) Random Forests | SpringerLink https://link.springer.com/article/10.1023/A:1010933404324 (accessed 2021 -02 -14).

(31) Dixon, M. F.; Halperin, I.; Bilokon, P. Feedforward Neural Networks. In *Machine Learning in Finance: From Theory to Practice*; Dixon, M. F., Halperin, I., Bilokon, P., Eds.; Springer International Publishing: Cham, 2020; pp 111–166. https://doi.org/10.1007/978-3-030-41068-1_4.

(32) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *Mach. Learn. PYTHON* 6.

(33) Keras: the Python deep learning API https://keras.io/ (accessed 2021 -02 -14).

(34) Bisswanger, H. Enzyme Assays. *Perspect. Sci.* **2014**, *1* (1), 41–55. https://doi.org/10.1016/j.pisc.2014.02.005.

491