

Machine-Learning-Assisted Analysis of Colorimetric Assays on Paper Analytical Devices

Bidur Khanal, Pravin Pokhrel, Bishesh Khanal,* and Basant Giri*



Cite This: *ACS Omega* 2021, 6, 33837–33845



Read Online

ACCESS |



Metrics & More

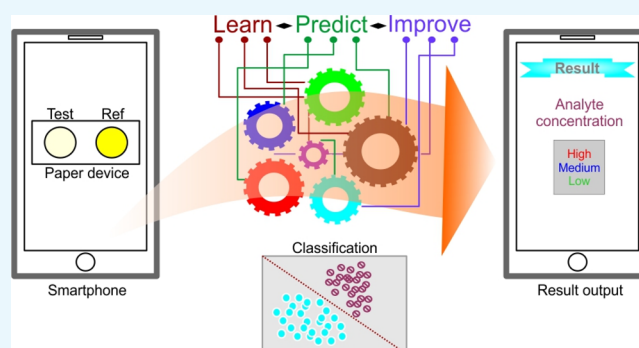


Article Recommendations



Supporting Information

ABSTRACT: Paper-based analytical devices (PADs) employing colorimetric detection and smartphone images have gained wider acceptance in a variety of measurement applications. 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 combination of four ML models—logistic regression, support vector machine (SVM), random forest, and artificial neural network (ANN)—as well as 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. The prediction accuracy was higher for food color than enzyme inhibition assays in most of the ML models and color space combinations. All models better predicted coarse-level classifications than fine-grained concentration classes. ML models using the sample color along with a reference color increased the models' ability to predict the result in which the reference color may have partially factored out the variation in ambient assay and imaging conditions. The best concentration class prediction accuracy obtained for food color was 0.966 when using the ANN model and LAB color space. The accuracy for enzyme inhibition assay was 0.908 when using the SVM model and LAB color space. Appropriate models and color space 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, food quality monitoring, and pharmaceutical quality screening among many other applications. Assays involving PADs are 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 PADs.⁶ Because of the proliferation of digital cameras, particularly smartphone cameras, the digital image-based colorimetric detection method is one of the widely used methods where color information encoded in the digital image is used for the analysis of assay results.³ Smartphone image-based colorimetric detection has been considered as a cost-effective and an attractive field-based alternative to conventional techniques such as spectrophotometry, colorimetry, and fluorometry.^{7,8}

Digital cameras use multiple charge-coupled devices (CCD) or complementary metal-oxide semiconductor (CMOS) sensors to capture the light intensity signal separately from 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 nonlinear. The raw signal or the intensity value at each pixel

of the imaged area depends on the lighting condition, sensor sensitivity, distance between the object and camera, and reflectance property of the object being imaged.^{10,11} Some of these variations such as the lighting condition and object-camera distance can be minimized using a controlled environment that is only possible in laboratory settings.^{12–15} However, PADs are ultimately meant to be used in field settings by a minimally trained user, during which the criteria of controlled assay and imaging conditions may not be achieved, resulting in less accurate results.

Constant illumination may be possible to maintain while imaging the assay signal by attaching an extra device on smartphones with added cost. The extra device may not be applicable in all types of smartphones having diverse shapes and sizes.¹⁶ Another approach uses a blank or reference assay along with the sample assay simultaneously to factor out the

Received: September 14, 2021

Accepted: November 19, 2021

Published: December 2, 2021



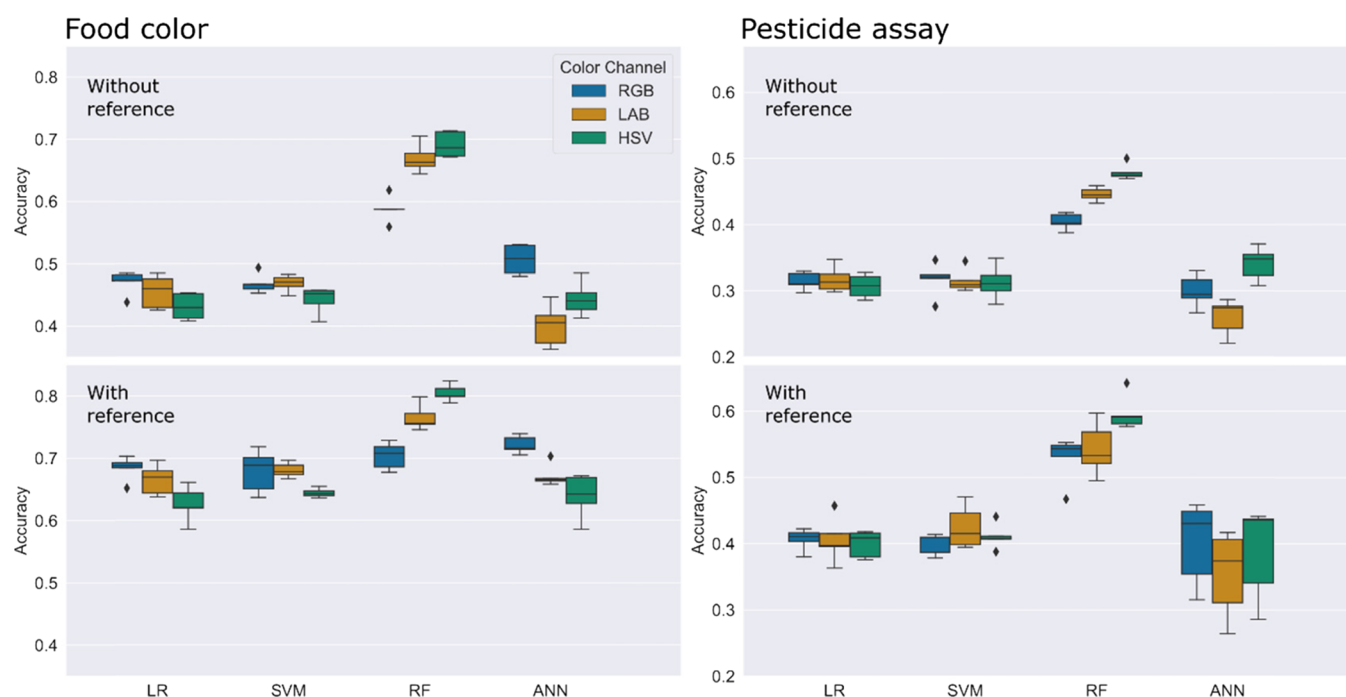


Figure 1. Cross-validation accuracy results obtained for both food color prediction (left panel) and pesticide assay (right panel). Each box describes the full range of variation (whisker's height), the likely range of variation (box height), and the median (horizontal line within the box) in the classification accuracy score of five cross-validation folds. All results in this figure were obtained using the mean of each color channel value of sample assays without (top) and with (bottom) reference assay test zones.

impact of illumination and camera quality changes.^{10,17} Because the raw signals captured by the camera sensors process nonlinearly during image acquisition before saving to the memory, the abovementioned approaches only partially address the problem. Furthermore, image autocorrection options such as automatic exposure correction, color correction based on ambient light selection, white balancing, and contrast enhancement can highly influence the overall color calibration.^{16,18} Thus, the estimation of the analyte concentration from the color intensity in PADs is an inverse problem where any estimation models will have explicit or implicit assumptions on the image formation process. The nonlinear factors need to be accounted for to make a robust and reliable low-cost model applicable in diverse point-of-need settings.

Traditional computer vision and ML models have been used for image processing to enhance the robustness of the PAD assay. Traditional computer vision-based algorithms try to be invariant to illumination, scale, and camera.¹⁰ Most of these algorithms use color spaces including the red, green, and blue (RGB); hue, saturation, and value (HSV); hue, saturation, and lightness (HSL); and luminance, a (green to red), and b (blue to yellow) (CIE $L^*a^*b^*$ or LAB).^{10,16,18} Each particular application usually requires careful selection of the color space model based on preliminary data and experiments. Similarly, other corrections such as white balance correction, contrast transfer, and gamma correction have been used. These corrections are specific for individual cameras. The camera-specific corrections are not practical when the goal is to enable low-cost colorimetry to the large variety of consumer cameras.

In recent years, data-driven ML algorithms are becoming increasingly common for colorimetric detection.^{19,20} ML has the potential to be robust against unwanted variations. It does not need an explicitly designed algorithm to extract

information. The model learns from data to work in diverse environmental settings.²¹ Bao et al. trained support vector machines (SVM) on RGB color channels in multiple indoor settings using a single camera.²² Similarly, Solmaz et al. used least-squares SVM (LS-SVM) and a multiclass random forest (RF) classifier to predict the peroxide content on colorimetric test strips and reported over 90% accuracy for six classes with interphone repeatability under variable illumination.²³ Kim et al. applied linear discriminant analysis (LDA), SVM, and artificial neural network (ANN) for the colorimetric analysis of saliva–alcohol concentrations, with average cross-validation accuracy rates of 100 and 80% for five standard and nine increased concentrations.¹⁹ Even though a few studies have reported the use of ML in image-based colorimetric detection, they lack insights into the relative efficacy of ML algorithms and their generalization capabilities. Generalization is an important issue in ML, where a model's high performance in training and validation data degrades in test data with different distributions. As reported by Morbioli et al.,²⁴ source codes and data used in most of the proposed ML models for colorimetric detection with PADs are not publicly available, making it difficult to reproduce results and perform benchmark comparisons.

In this work, we designed a set of comprehensive experiments to analyze the performance and utility of ML for colorimetric image analysis of PADs using two different data sets, food color and enzyme inhibition assays, for pesticide residue determination. We assessed four different ML models—logistic regression (LR), SVM, RF, and ANN—and three image color spaces—RGB, HSV, and LAB—to predict the target analyte concentration using images of PADs taken at varying lighting conditions, with different cameras and users. The colorimetric assays involved in our approach included both samples and reference assay zones on the paper device in

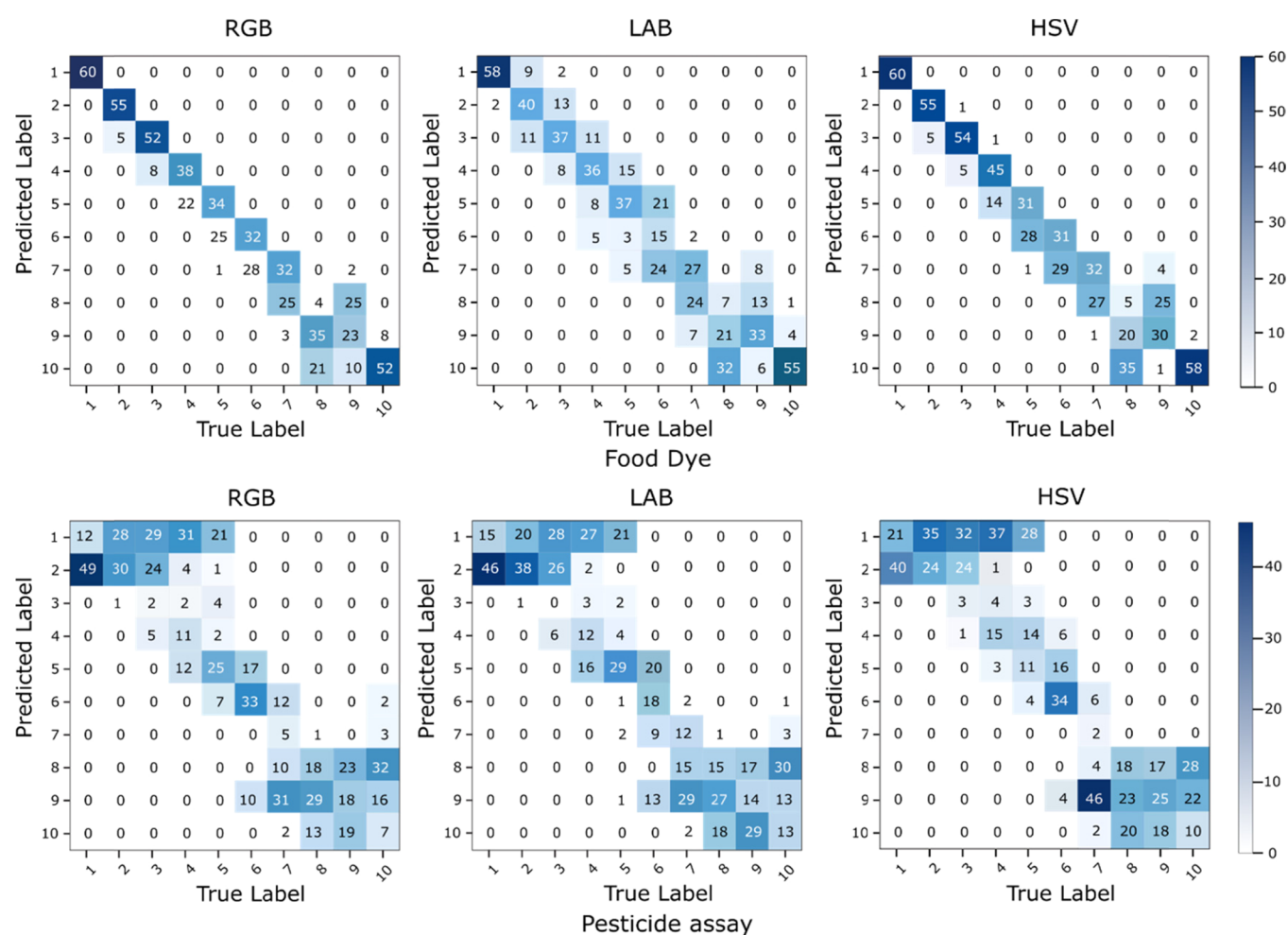


Figure 2. Confusion matrix (of the test accuracy of LR) for the food color and pesticide assay when RGB, LAB, and HSV color models are used. The confusion matrix of other models also shows similar patterns.

contrast to most of the previous studies that captured only the target sample assay zone. In this setting, we obtained the concentration class prediction accuracy of 0.966 and 0.908 for food color and pesticide residue analysis datasets, respectively, in independently created test datasets. We also highlight the limitations of all ML models when there is a domain shift in test data and their inability to predict with the same accuracy at all concentration levels. We have made our source code and data publicly available to contribute to the reproducibility issues in this rapidly progressing field.

RESULTS AND DISCUSSION

We established a baseline result using the mean color intensity as the input feature to ML models. Cross-validation results for various color spaces and ML models using mean pixel values of the ROIs as input features to classify into 10 concentration classes are shown in Figure 1. At first, we tested the models using test zones of samples only. In general, the RF model yielded higher accuracy in all three-color spaces in the case of food color experiments. In this case, the HSV color space had a higher (0.691) value than LAB (0.669) and RGB (0.588). The LR and SVM models exhibited lower accuracy compared to the RF model. In both LR and SVM models, all three-color spaces had similar values compared to RF. We also looked at the ANN model's ability to predict correct assay values. It gave comparable values with the LR and SVM models, but the

values for three different color spaces exhibited some variations. In the case of pesticide assay, the accuracy was less than that seen in food color experiments for all combinations of models and color spaces, as shown in Figure 1. However, the trend of accuracy results was similar to that of food color assay.

Combining the Sample and Reference Assay Color.

To improve the prediction accuracy, we took into consideration both the sample and reference assay colors. In these experiments, we investigated if imaging both the reference or control assay and sample assay at the same time could improve the prediction accuracy by ML models. The latter approach showed an improvement in the prediction accuracy of all the ML models when the mean color intensity from the reference region was included as the feature (Figure 2). Figure 2 suggests that the ML model was able to use information from the reference sample region to partially factor out the variations due to ambient lighting conditions and camera parameters and the image acquisition setup, improving the concentration prediction results when using the reference and sample compared to using only the target sample. 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 the reference and sample color could be a simple difference or an n -degree polynomial.

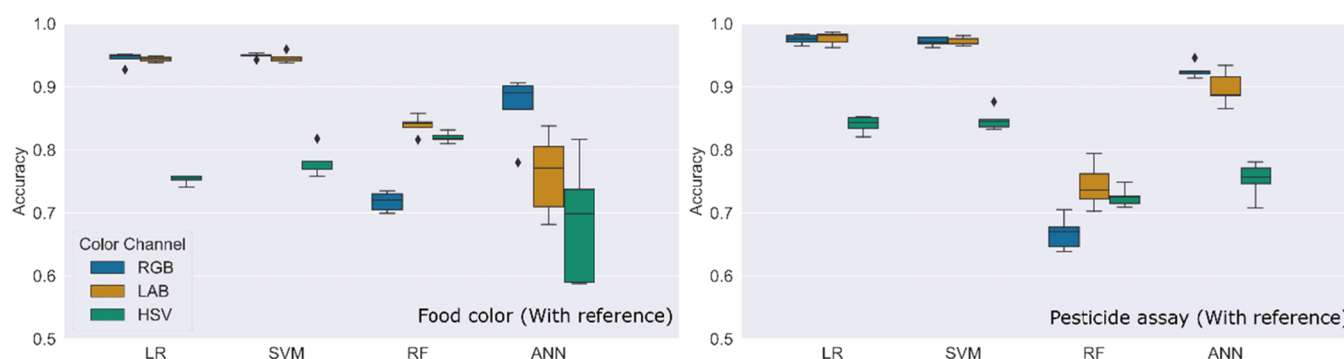


Figure 3. Cross-validation accuracy using all the pixel values from 16×16 downsampled images of reference and sample as the input features. Each box describes the full range of variation (whisker's height), the likely range of variation (box height), and the median (horizontal line within the box) in the accuracy score of five cross-validation folds. LR, SVM, RF, and ANN are implemented in three color models: RGB, LAB, and HSV.

The cross-validation accuracy using the RGB color space for the food color assay was found to be higher in LR, SVM, and ANN models, while it was lower in the RF model when compared to the LAB and HSV color spaces. The HSV color space in RF and the RGB color space in ANN models showed higher accuracy. However, we did not find any specific trends among color spaces and models used. Similar results were observed for the pesticide assay dataset. It is interesting to note that the HSV color space using RF showed the highest average accuracy in comparison to other color models (i.e., 0.804 in the food color assay and 0.596 in the pesticide assay). In LR, the highest average accuracy was observed in the RGB color model (0.684 in the food color assay and 0.406 in the pesticide assay). Similarly, in SVM, the highest average accuracy was in LAB with 0.68 for the food color assay and 0.42 for the pesticide assay. For ANN, the highest average cross validation (CV) accuracy was in RGB, that is, 0.721 for the food color assay and 0.401 for the pesticide assay.

Along with cross-validation experiments, we also evaluated all the models and color spaces in a separate test dataset, the details of which are provided in Table S1. As shown in Table S1, the highest test accuracy was 0.67 with the HSV color space in SVM for the food color assay and 0.34 with the HSV color space in ANN for the pesticide assay (in ANN with HSV). The results, as expected, showed that test results did not necessarily agree with the cross-validation accuracy. A large drop in test accuracy was observed across all the models compared to the cross-validation results (Tables S1 and S2). Here, the HSV color model showed good results in both sets of assays. Our results highlight that reporting only cross-validation scores or scores in test data that are very similar to the training set can overestimate ML model's performance. ML models' performance can severely degrade when the statistical distribution of the test data is different from that of the training set. This is closer to the real-world scenario we wanted to emulate to further assess the efficacy and applicability of ML systems when test images are captured in different field settings. When comparing the results across food color and pesticide assays, we observed that the accuracy in the pesticide assay is relatively lower than that of the food color assay for all the models. We attribute this observation to the chemistry of the pesticide assay. Unlike the food color assay, 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.

The confusion matrix shown in Figure 2 shows that the food color assay has its values clustered near the diagonal line, except for classes 8, 9, and 10. These classes correspond to low concentration values with faint colors (best viewed in the color image). For the pesticide assay, the values in the initial classes (1, 2, 3, and 4) and final classes (8, 9, and 10) are misclassified in large numbers. The matrix fields in the middle, although not accurate, show a diagonal pattern. This result seems to follow a typical S-shaped enzyme assay curve.²⁵

In such cases, the ML models can utilize information from the reference region to partially factor out the variations of ambient lighting conditions and the 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 the reference and sample color could be a simple difference or an n -degree polynomial.

Input Feature Vectors from the Downsampled Image. Figure 3 shows the CV accuracy using all the pixels of downsampled ROIs as a feature vector for ML models. Because images containing both sample and reference colors provided better accuracy than using the 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. A similar trend was observed both in the food color assay and pesticide assay. LR and SVM models with RGB and LAB color spaces resulted in the highest accuracies. For LR, CV accuracies were 0.975 and 0.977 when using RGB and LAB, respectively. Likewise, for SVM, the CV accuracy was 0.971 when using RGB or LAB. When the test dataset was evaluated, the test accuracy was lower in both food color and pesticide assays. However, the gap between cross-validation and test accuracy in pesticide is higher for the pesticide assay in all the color channels and models (Table S1).

Classification into High, Medium, and Low. The results in the previous section show that accurately and robustly estimating fine-grained concentration classes is difficult even with powerful ML models using 10 concentration classes.

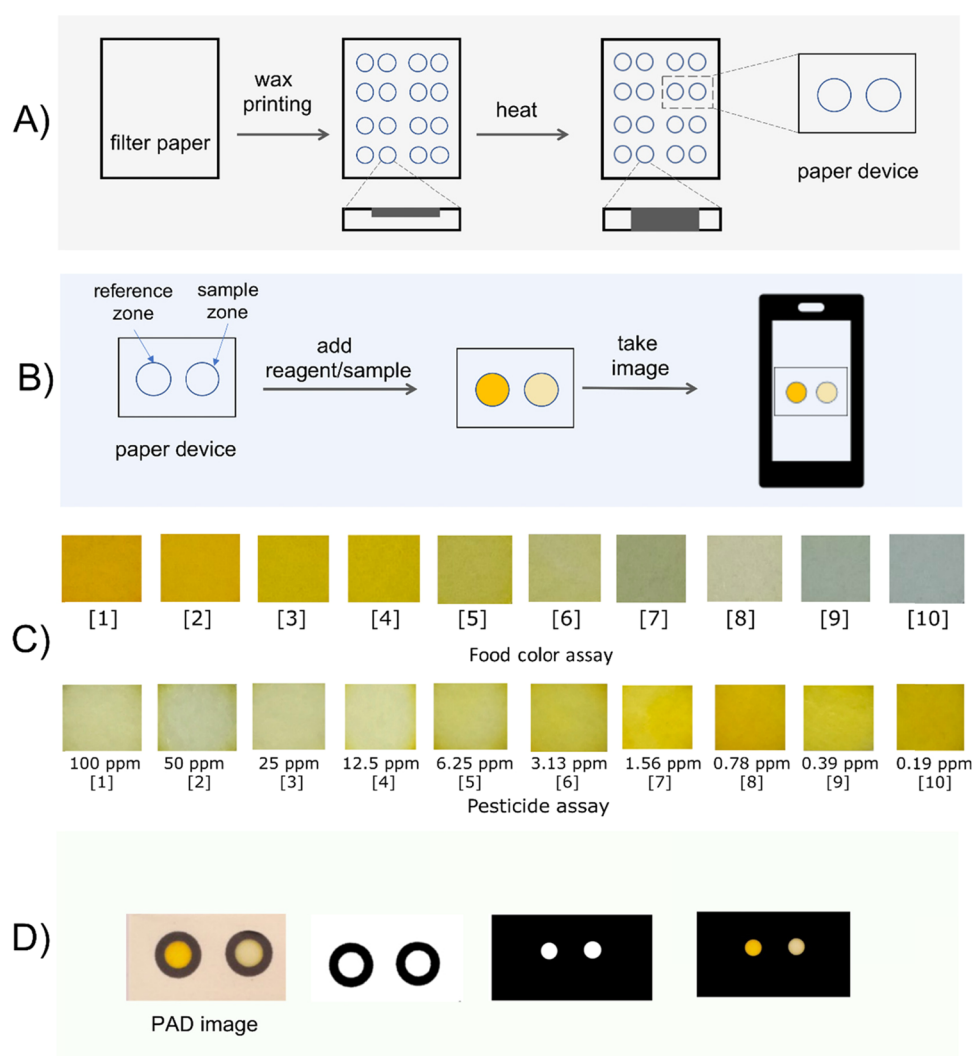


Figure 4. (A) Fabrication of the paper device. Solid wax was printed on Whatman filter paper, which penetrated through the paper after heating (see inset illustrations). (B) General procedure for paper-based pesticide assay. (C) Representative images of paper device test zones after assays were performed. Ten different dilutions of the food color and 10 different concentrations of the pesticide were used. Because the pesticide assay followed enzyme inhibition reaction, the concentrations and color intensity are inversely correlated. (D) Automatic color pixel extraction procedure from the PADs: left to right represent binary thresholding, mask generation, and masked region of interest (ROI).

Therefore, we merged 10 concentration classes into three distinct classes: high, medium, and low for semiquantitative prediction of both the food dye and pesticide samples. Classes 1, 2, and 3 shown in Figure 4C were merged into the high category, 4, 5, 6, and 7 were merged into the medium category, and 8, 9, and 10 were merged into the low category in the case of the food color assay. In the case of the pesticide assay, classes 1, 2, 3, and 4 in Figure 4C were merged into high, 5 and 6 into medium, and 7, 8, 9, and 10 into low categories.

Figure 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 three-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 the food color assay 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 the test dataset. We reported the test accuracies for the case of three classes, as shown in Table 1. Similar to the CV accuracies for 10 classes, the test accuracies for 10 classes are poor compared to three classes (see Table S1). Table 1 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 the food color assay, we observe the highest accuracy of 0.966 in ANN with the 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 features. For the pesticide assay, the highest accuracy of 0.908 was obtained in SVM with the LAB color space and the input feature from the individual means of each color channel. The obtained results show that accurate semiquantitative prediction of concentration from PAD images is possible even in an uncontrolled setup with enough data and a suitable ML model.

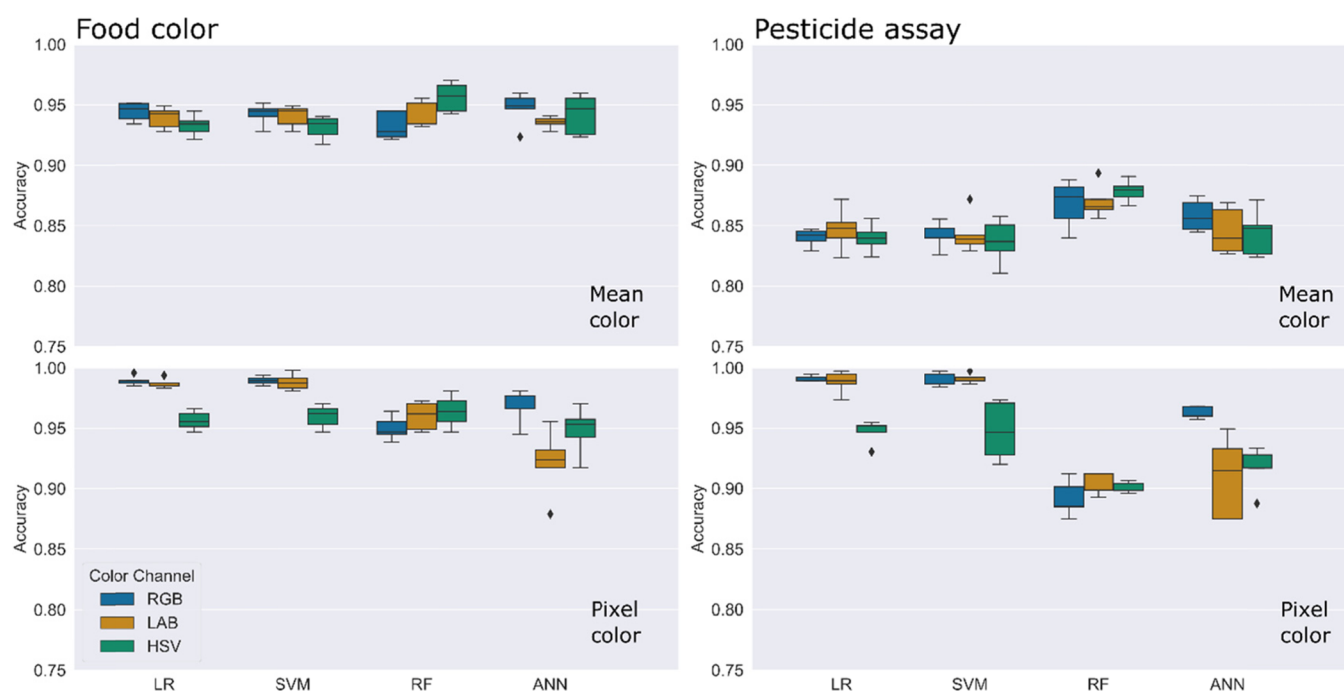


Figure 5. Cross-validation accuracy for three reduced classes of high, medium, low. Results of both the mean color and 16×16 downscaled RGB pixels are presented for food color (left panel) and pesticide assays (right panel).

Table 1. Average Test Accuracy Using Three Concentration Classes (High, Medium, and Low)^a

(A) 16×16 image as the input feature						
	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) means of individual R, G, and B channels as the input feature						
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

^a(A) All color pixels from the three channels of downsampled 16×16 image as the input feature and (B) individual mean of each color channel (3 means) as the input feature. All images included sample and reference assays.

CONCLUSIONS

We evaluated four ML models for their ability to accurately predict the concentration of target analytes on the paper device platform. We found that the ML models that used the sample color along with a reference color increased the models' ability in predicting the result. The reference assay color may provide a one-point calibration to estimate or predict the concentration of analytes in the given sample. In general, we found that the accuracy of the food color assay was higher than the accuracy of the pesticide assay in most of the combinations. Our results show that ML models may provide only limited accuracy when using fine-grained estimation of concentration classes but provide high accuracy when using them for a coarse-level classification such as low, medium, and high. The ability of ML models to accurately classify the pesticide concentration to such three classes even in difficult real-life test images shows

the potential of using ML-powered PADs as a low-cost quick field-testing method.

Smartphone cameras allow for postprocessing even before we save or see images. Because it is very hard to understand and identify these individual preprocessing 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. Because the colorimetric assays on paper devices do not provide variations 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 and clinical diagnosis during emergencies for assays involving colorimetric paper devices. Appropriate ML models integrated in smartphones that can read assay results performed on the PAD platform by taking images process or analyze the signal to accurately predict assay results and report or store the results locally or on cloud, which could be powerful tools in several measurement applications. Our future work will investigate the images from real samples that have a complex matrix and mixture of different colors instead of single color to understand how interferents play roles.

EXPERIMENTAL SECTION

Fabrication of PADs. We designed a layout of circular patterns in a computer and printed on Whatman No 1 grade filter paper using a Xerox ColorQube 8580 solid wax printer.²⁶ The wax printed paper was heated from the backside by pressing with a dry clothing iron on its surface. The backside of

the PADs was laminated to prevent the leakage of reagents to the other side of the paper. Finally, the paper sheet was cut in such a way that each PAD contained two circular assay regions as reference and sample zones (Figure 4A).

Classification Datasets. We prepared datasets for two different assays: the first one using the food color assay and the second one using the pesticide assay. Each dataset used four different smartphones (Huawei SCC-U21, iPhone 6, Honor 8C, and Samsung Galaxy J7 Max) for image acquisition at different lighting conditions, camera to PAD distances, and capture angles. The lighting conditions included outdoor sunlight, indoor daylight, fluorescent light, incandescent light, and a combination of them. A general procedure of the assay on a paper device is given in Figure 4B.

The food color data sets were obtained by loading 10 different dilutions of yellow food color (Foster Clark Product Ltd.) onto the PADs. We chose yellow dye to closely match the color produced in pesticide experiments so that our analysis remains consistent. We captured 2400 images in total under various conditions. Images that were unclear or blurred were removed, and 2353 images were used for training ML models. The images were labeled from class 1 to 10. One for the highest concentration and 10 for the lowest concentration of the food color. The dataset contained an approximately equal number of images per class. A new set of 600 images of the same food color concentrations was obtained and was used as the test data set. These images were taken on a different day—varying the illumination, randomly changing the camera, and camera–PAD distance—to create different test datasets from the training datasets. Representative images are shown in Figure 4C.

The second dataset was prepared for a more realistic application, which included enzyme inhibition assay for the pesticide residue measurement based on the Ellman method.²⁷ This assay is based on the inhibition of acetylcholinesterase enzyme activity in the presence of pesticides. In this assay, the acetylcholinesterase enzyme (Sigma-Aldrich) breaks down the acetylthiocholine chloride (AtCh) substrate (Sigma-Aldrich) into thiocholine and acetic acid. The thiocholine molecules react with Ellman's reagent (dithiobisnitrobenzoic acid; DTNB) (Sigma-Aldrich) to give a yellow product of thionitrobenzoic acid.²⁷ To run the assay, we first loaded the test and reference zones on PADs with the acetylcholinesterase enzyme. Then, the sample containing the pesticide was added to the test zone, while a blank solution with no pesticides was added in the reference zone. After 5 min, an enzyme substrate acetylthiocholine chloride was added to both zones. In the reference zone, the enzyme remains active. In the test zone, its activity is compromised due to the presence of organophosphate and the carbamate group of pesticides, which inhibit the enzyme activity by binding to it. Based on the extent of inhibition, the amount of pesticide on the sample is estimated. Finally, Ellman's reagent (DTNB) was added, which reacted with the thiocholine molecules produced by the enzyme reaction to give a characteristic yellow product of thionitrobenzoic acid. In the reference zone, a very vibrant yellow color develops because of the retention of full enzyme activity, whereas the color intensity is inversely proportional to the concentration of the pesticide present in the sample in the test zone.^{1,28} The images of PADs were captured at 10 min of the enzyme reaction using a smartphone for further analysis (see Figure 4B for a general outline). Considering the short stability

of the enzyme and enzyme substrate, we used a freshly prepared enzyme and enzyme substrate during the assay.

We collected 1872 images for training data sets by repeating the same experiment in multiple days and different lighting conditions using four different smartphones. Each image was categorized into a class from 1 to 10 like food color assay images. Class 1 represents a pesticide (Paraoxon, Sigma-Aldrich) concentration of 100 ppm. The same pesticide solution was serially diluted half in the remaining assays. New experiments at different lighting conditions were performed to obtain 601 new images as test datasets. See Figure 4C for the representative images of the pesticide assay.

Extraction of Pixel Values from Assay Images. The preprocessing of a PAD image is outlined in Figure 4D. The leftmost image in panel 4D is a typical image of the PAD captured using a camera. The two regions encircled by black rings are the 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 a 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, and 0.8 and 30 are empirically chosen values after visual inspection across multiple images. The binary images provide us with the masks that were used to extract the pixel values lying in the two ROIs of the corresponding original images, as shown in the rightmost image of Figure 4D.

Evaluation of Multiclass 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 $ACA = \frac{\text{total no correct predictions}}{\text{total no of predictions}}$. In addition, we visualized the results using confusion matrices that provide information on how many samples of a particular class are misclassified to another class. We used a fivefold cross-validation where the training dataset was randomly split into five 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. To evaluate the robustness of ML models and their generalization ability, we also evaluated the models with a separate test set under different conditions trying to emulate actual real-life testing scenarios.

We used various approaches to extract features and feed them as input to the ML models. We compare these approaches using the following two experimental setups: (a) color spaces: RGB vs HSV vs LAB color representation; (b) mean pixel value of ROIs for each channel vs using all the pixel values of the downsampled ROIs. Similarly, we assessed the impact of the reference test region using the third setup; and (c) using only target sample vs using both the reference and target sample.

After extracting the mean color intensity of the sample and reference assay zones, we obtained $2 \times 3 = 6$ unique values for each assay (two circular zones and three-color channels). This can be fed as a six-dimensional feature vector to ML classification models. The mean values from the ROI do not capture the variation of pixel values within the ROI. However, using all the pixels of the ROIs as the input features to train ML models dramatically increases the feature dimension,

Table 2. Implementation Details of ML Models

model type	library or framework	implementation details of models
logistic regression	Scikit-learn	multinomial + L2 penalty, L-BFGS solver, maximum iteration = 10,000
support vector machine	Scikit-learn	squared L2 penalty, linear kernel
random forest	Scikit-learn	no. of trees = 1000, split criterion = gini
artificial neural network (fully connected)	Keras	3-dense layers, sigmoid + softmax activation, adam optimizer, and 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

which computationally affects the training of some ML algorithms such as SVM. As most of the PAD colorimetry images only have color information without texture and shape information, as a compromise, we downsampled the cropped image into a size of 16×16 and converted it to a one-dimensional (1D) vector of dimension $16 \times 16 \times 3 = 768$ (for three-color channels) for each reference and target ROIs, resulting in two 1D vectors.

We compared the following four most widely used supervised ML models for colorimetry with PADs: LR,²⁹ SVM,³⁰ RF,³¹ and ANN.³² We used Python-based ML library, Scikit-learn,³³ to build LR, SVM, and RF models. For ANN, we adopted a popular Python-based framework, Keras.³⁴ The details of the implementation configuration are given in Table 2.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.1c05086>.

Average accuracy across food color and pesticide test sets in various settings and cross-validation accuracy (in the mean and standard deviation) across food color and pesticides in various settings (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Bishesh Khanal – Nepal Applied Mathematics and Informatics Institute for Research, Kathmandu 44600, Nepal; Email: bishesh.khanal@naamii.org.np

Basant Giri – Center for Analytical Sciences, Kathmandu Institute of Applied Sciences, Kathmandu 44600, Nepal; orcid.org/0000-0003-4798-3414; Email: bgiri@kias.org.np

Authors

Bidur Khanal – Center for Analytical Sciences, Kathmandu Institute of Applied Sciences, Kathmandu 44600, Nepal; Nepal Applied Mathematics and Informatics Institute for Research, Kathmandu 44600, Nepal

Pravin Pokhrel – Center for Analytical Sciences, Kathmandu Institute of Applied Sciences, Kathmandu 44600, Nepal

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acsomega.1c05086>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the National Academy of Sciences (NAS) and United States Agency for International Development (USAID) for funding this work through Partnerships for

Enhanced Engagement in Research (PEER) program (award no: AID-OAA-A-11-00012 to BG). The opinions, findings, conclusions, or recommendations expressed in this article are those of the authors alone and do not necessarily reflect the views of NAS or USAID. Krishna Pokharel and Retina Shrestha helped during image acquisition experiments. Article processing charge was supported by a The World Academy of Sciences (TWAS) to Dr. Basant Giri (Grant # 19-259 RG/CHE/AS_G).

■ 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*, 1492–1502.
- (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*, 1–12.
- (3) Ozer, T.; McMahan, C.; Henry, C. S. Advances in Paper-Based Analytical Devices. *Annu. Rev. Anal. Chem.* **2020**, *13*, 85–109.
- (4) Sharma, N.; Barstis, T.; Giri, B. Advances in Paper-Analytical Methods for Pharmaceutical Analysis. *Eur. J. Pharm. Sci.* **2018**, *111*, 46–56.
- (5) Pokhrel, P.; Jha, S.; Giri, B. Selection of Appropriate Protein Assay Method for a Paper Microfluidics Platform. *Pract. Lab. Med.* **2020**, *21*, No. 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.
- (7) Rezaadeh, 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.
- (8) Zhao, W.; Tian, S.; Huang, L.; Liu, K.; Dong, L.; Guo, J. A Smartphone-Based Biomedical Sensory System. *Analyst* **2020**, *145*, 2873–2891.
- (9) Li, X.; Gunturk, B.; Zhang, L. Image Demosaicing: A Systematic Survey. In *Visual Communications and Image Processing*; International Society for Optics and Photonics, 2008; Vol. 6822, p 68221J.
- (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*, 6507–6509.
- (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*, 517–522.
- (12) Barros, J. A. V. A.; de Oliveira, F. M.; Santos, G. D. O.; Wisniewski, C.; Luccas, P. O. Digital Image Analysis for the Colorimetric Determination of Aluminum, Total Iron, Nitrite and Soluble Phosphorus in Waters. *Anal. Lett.* **2017**, *50*, 414–430.
- (13) Masawat, P.; Harfield, A.; Namwong, A. An iPhone-Based Digital Image Colorimeter for Detecting Tetracycline in Milk. *Food Chem.* **2015**, *184*, 23–29.
- (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.

- (15) da Nobrega Gaiao, E.; Martins, V. L.; da Silva Lyra, W.; de Almeida, L. F.; da Silva, E. C.; Araújo, M. C. U. Digital Image-Based Titrations. *Anal. Chim. Acta* **2006**, *570*, 283–290.
- (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*, 2449.
- (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.
- (18) Bronner, T.-F.; Boitard, R.; Pourazad, M. T.; Nasiopoulos, P.; Ebrahimi, T. Evaluation of Color Mapping Algorithms in Different Color Spaces. In *Applications of Digital Image Processing XXXIX*; International Society for Optics and Photonics, 2016; Vol. 9971, p 99710H.
- (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*, 84–92.
- (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*, 2434–2441.
- (21) Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.-D. Machine Learning in Manufacturing: Advantages, Challenges, and Applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45.
- (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*, 1387–1395.
- (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* **2018**, *255*, 1967–1973.
- (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.
- (25) Bisswanger, H. Enzyme Assays. *Perspect. Sci.* **2014**, *1*, 41–55.
- (26) Carrilho, E.; Martinez, A. W.; Whitesides, G. M. Understanding Wax Printing: A Simple Micropatterning Process for Paper-Based Microfluidics. *Anal. Chem.* **2009**, *81*, 7091–7095.
- (27) Ellman, G. L.; Courtney, K. D.; Andres, V.; Featherstone, R. M. A New and Rapid Colorimetric Determination of Acetylcholinesterase Activity. *Biochem. Pharmacol.* **1961**, *7*, 88–95.
- (28) Rajangam, B.; Daniel, D. K.; Krastanov, A. I. Progress in Enzyme Inhibition Based Detection of Pesticides. *Eng. Life Sci.* **2018**, *18*, 4–19.
- (29) *Encyclopedia of Machine Learning and Data Mining*, 2nd ed.; Sammut, C.; Webb, G. I., Eds.; Springer US, 2017.
- (30) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
- (31) Random Forests|Springer Link. <https://link.springer.com/article/10.1023/A:1010933404324> (accessed February 14, 2021).
- (32) 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.
- (33) 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**; 2012, arXiv:1201.0490.
- (34) Keras: the Python deep learning API <https://keras.io/> (accessed February 14, 2021).