(Received 5 December 2024, Accepted 19 December 2024, Scheduled 21 December 2024)

Smartphone-Integrated YOLOv4-CNN Approach for Rapid and Accurate Point-of-Care Colorimetric Antioxidant Testing in Saliva

Youssef Amin, Paola Cecere, Tania Pomili, and Pier P. Pompa*

Istituto Italiano di Tecnologia (IIT), Nanobiointeractions & Nanodiagnostics, Via Morego 30, Genova 16163, Italy

ABSTRACT: This study introduces a machine learning (ML)-based method for point-of-care (POC) colorimetric testing of total antioxidant concentration (TAC) in saliva, an important biomarker for health monitoring. The approach leverages ML to accurately classify color intensity in the POC test. Saliva samples were collected and imaged at specific intervals during the colorimetric reaction, generating a dataset representative of various antioxidant levels. Four classifiers (Convolutional Neural Network, Support Vector Machine, K-Nearest Neighbors, and Single-layer Feed-forward Neural Network) were evaluated on distinct datasets, with Convolutional Neural Network (CNN) consistently achieving superior performance. To enhance classification accuracy, stacking-based ensemble learning was applied, combining CNN predictions with a Support Vector Machine (SVM) meta-classifier, achieving up to 92% accuracy. Additionally, YOLOv4-tiny was utilized for object detection to isolate regions of interest in the images, creating a refined dataset that a CNN model is then classified with ca. 98% accuracy. This YOLOv4-CNN approach not only improved accuracy but also simplified the model architecture. The integrated object detection and CNN models were deployed on an Android application, enabling real-time TAC analysis on a smartphone with 98% accuracy and a fast readout time of 2 minutes. This method offers a robust, efficient, and accessible solution for POC antioxidant testing.

1. INTRODUCTION

Point-of-care (POC) diagnostics has emerged as a transformative approach in modern healthcare, enabling rapid, accurate, and on-site detection [1]. Unlike traditional laboratory techniques, POC diagnostics can bring the testing process closer to the patient, improving accessibility and facilitating timely clinical decisions. POC testing often integrates colorimetric assays by combining chemical reactions with visual/optical detection systems, enabling fast biomarker measurement [2, 3]. These tests have been widely used across various fields and applications, including blood analysis [4], food allergen testing [5], urine albumin analysis [6], and pH quantification [7].

Among various biomarkers, salivary total antioxidant concentration (TAC) is recognized as a promising noninvasive tool to monitor oxidative stress levels [8, 9] and body's overall antioxidant defense system, providing critical insights into health and wellness [10]. TAC is also associated with the pathophysiology of cardiovascular diseases [11, 12], aging process [13, 14], and cancer [15, 16]. Hence, monitoring TAC levels provides valuable insights into health and wellness [17]. Besides its potential in health monitoring [1, 18, 19], TAC is also used for quality control in the food industry [20–22]. However, traditional colorimetric tests often depend on visual interpretation by users, which introduces variability due to differences in human color perception [23, 24], further complicating the use of naked-eye or color chart-based methods [3]. Additionally, ambient lighting conditions and camera optics significantly impact the accuracy of colorimetric analyses, which require advanced algorithms, such as machine learning (ML), to address all these issues [7, 25–33].

ML algorithms offer automated decision-making and self-learning capabilities, enabling more objective and consistent results in colorimetric testing. Recent advances in smartphone technology have further promoted the integration of ML models into mobile platforms [34–37] and other embedded devices (e.g., STM32 [38] and ARM microcontrollers) [39–41]. This led to enhanced usability, portable solutions, automated detection, and personalized diagnostics. Moreover, deep learning, which is a subset of ML, has shown remarkable success in image-based applications, including object detection and classification tasks [42]. Its ability to learn hierarchical representations from data makes it ideal for analyzing complex colorimetric reactions [43]. However, despite its potential, several challenges arise, such as computational demands, data requirements, and deployment limitations, especially when using deep learning algorithms for POC devices [44–47]. Therefore, very often establishing a balance between the computational cost and the performance of the model is important, especially in resource-constrained-devices [48].

Several studies have highlighted the use of ML/deep, where they encountered some challenges and limitations. Mutlu et al. (2017) reported the difficult classification of closely spaced levels of pH using Least Squares SVM classifier (LS-SVM), due to overlapping color intensities [7]. Furthermore, Kim et

* Corresponding author: Pier Paolo Pompa (pierpaolo.pompa@iit.it).

doi:10.2528/PIER24120505

al. (2016) developed a smartphone-based system for classifying saliva alcohol concentrations. Three classifiers (i.e., Linear Discriminant Analysis, Artificial Neural Network, and SVM) were evaluated on four color spaces. Artificial neural network (ANN) models achieved the best accuracy but struggled to classify intermediate concentration levels, achieving 80% average accuracy. The authors also faced challenges scaling the mass deployment for the stand-alone approach [49]. Recent advancements in smartphone-integrated colorimetric assays have demonstrated their potential for enhancing accessibility and accuracy in quantitative and semiquantitative analysis. Nevertheless, environmental factors such as ambient lighting and camera optics introduce variability in image quality, potentially affecting colorimetric accuracy. For example, Sajed et al. (2020) developed a smartphone-based system for estimating lead ion concentrations, employing ML to improve the precision of colorimetric measurements extracted from smartphone-captured images. However, it required additional preprocessing steps and controlled experimental setups to ensure reliable performance [36]. Additionally, Bhatt et al. (2023) developed a smartphone-based colorimetric system for quantifying urinary albumin using ML algorithms. Chemically impregnated dipsticks were analyzed through RGB, HSV, and Lab color spaces, with K-Nearest Neighbor (kNN) achieving the highest accuracy of 96% under constant illumination from the smartphone [35]. However, challenges included dependence on controlled lighting (Flash ON) and limited dataset diversity. In [26], the authors incorporated 15 reference color patches in a custom-designed color chart placed alongside the assay vial to help improve the model classification of thiocyanate concentrations in saliva. These patches allowed the system to estimate and compensate for variations in lighting and camera processing effects using a CNN architecture via input-aware Neural Architecture Search (NAS), achieving an increase of accuracy by about 14%. Similarly in [50], the authors demonstrated that combining the sample's color with a reference color improved prediction accuracy when analyzing colorimetric assays on paper-based devices (PADs), achieving a maximum accuracy of 96.6% for food dye using ANN model and 90.8% for pesticide assays using SVM model. Yet, they highlighted the computational challenges due to high-dimensional features on some ML algorithms such as SVM. Furthermore, Feng et al. (2023) utilized a CNN-based approach with powerful parallel processing for urine glucose detection, achieving an 87.6% correlation with standard blood glucose tests [51]. Instead, [27] focused on glucose detection in saliva utilizing different classifiers such as Linear Discriminant Analysis (LDA), Gradient Boosting Classifier (GBC), and Random Forest (RF), achieving 98.24% classification accuracy with Tetramethylbenzidine (TMB) under various lighting and smartphone conditions. Moreover, Duan et al. (2023) employed four deep learning algorithms, for enzyme-linked immunosorbent assays (ELISA), achieving over 97% accuracy with a GoogLeNet-trained model. However, their dependence on cloud-based processing created challenges for real-time, resource-limited applications [27, 52]. Additionally, Random Weight Neural Networks (RBNs) were proposed to quantify antioxidant levels in saliva using colorimetric reactions in a vial relying on a centralized server for inference [1]. Besides that, collecting and labeling the dataset can be expensive and time-consuming causing limited adaptability to more complex biological matrices. In such cases, Deep Neural Networks (DNN) may face overfitting when small datasets are involved. In [53], each sample of Cis-Pt required 20-minute reaction time to be collected. On the other hand, in [21], antioxidant concentration level was detected using a lateral flow device after 10 minutes of reaction time, whereas in [36] the system required 30 min reaction time. These challenges and limitations highlight the need for fast, more robust, and adaptable methodologies that can deliver consistent performance.

Addressing some of the main challenges, herein we designed an in-house colorimetric assay to assess the TAC in human saliva. The test was fine-tuned to develop a POC-compliant tool, featuring a rapid time-to-result within a few minutes. Additionally, a data augmentation technique was applied to improve the performance of the color classification models in the absence of object detection methods. We evaluated two approaches for TAC classification: (1) ensemble learning based on Stacking, and (2) an alternative approach integrating object detection (i.e., YOLOv4-tiny). By capturing an image just 2 minutes after the test starts, our system uses rapid reaction dynamics for efficient classification, ensuring both speed and precision. This streamlined architecture reduced computational complexity while achieving high classification accuracy. We designed preprocessing and ML algorithms that cope with the deployment on a resource-constrained device. We deployed the algorithms (i.e., for the YOLOv4-CNN approach) on a smartphone achieving real-time, on-device processing, and inference, overcoming the dependency on cloud infrastructure found in other applications. The YOLOv4-CNN approach for TAC classification showed consistent performance when being deployed on a smartphone, matching the results observed during offline testing. These contributions not only enhance classification accuracy but also demonstrate the feasibility of practical, efficient ML-driven diagnostic solutions for other POC applications.

The structure of this paper is as follows. Section 2 details the methodology. Section 3 presents the experimental setup. Section 4 provides the results and discussion, and Section 5 concludes the paper.

2. METHODOLOGY

2.1. Imaging Setup for Consistent Image Quality

In imaging-based applications, consistent lighting conditions are critical for ensuring reliable image quality, particularly for machine learning tasks. Variations in ambient light can significantly affect image clarity, contrast, and color representation, leading to inconsistencies in downstream processing. To address this challenge, we designed a 3D-printed black-closed box with an aperture for smartphone placement. This design enables users to capture images under controlled lighting conditions using the smartphone's flashlight as the sole illumination source. The box minimizes the effects of external light variations and provides a neutral, standardized background, ensuring consistent image clarity and reproducibility.

2.2. Assessing TAC in Saliva

To assess the total antioxidant capacity (TAC) in human saliva, we utilized an in-house designed colorimetric assay based on the well-known competitive reaction of antioxidants with the TMB/H_2O_2/HRP redox system [18, 54]. In this system, the enzyme horseradish peroxidase (HRP) first catalyzes the generation of hydroxyl radicals ($OH^\cdot$) via cleavage of hydrogen peroxide (H_2O_2). These hydroxyl radicals then oxidize the chromogenic substrate $3,3'$,5, $5'$-tetramethylbenzidine, converting its colorless reduced form into the blue-colored oxidized product [18, 54–56]. The resulting shift of the solution color from transparent to dark blue depends on the concentration of the oxidized TMB and occurs in a few minutes.

Antioxidants present in human saliva can neutralize the HRP-generated hydroxyl radicals, effectively competing with this catalytic oxidation of TMB and preventing the solution's blue color development [18]. In our semiquantitative assay, the colorimetric detection harnessed this inverse correlation between antioxidant levels in saliva samples and the blue intensity of the test solution, ranging from transparent to a darker blue color as the antioxidant concentration decreases. The assay was fine-tuned to develop a POC-compliant tool, featuring a rapid time-to-result within a few minutes and a dynamic range of approximately 0.1–1.0 absorbance units, thus allowing visual colorimetric readout. The method was then validated with a standard ABTS-based commercial kit, assessing the correlation between the antioxidant concentrations in human saliva samples and the absorbance kinetics of our colorimetric reactions.

In our POC test, five antioxidant levels were classified. The middle level represented the physiological range of antioxidants in saliva, while the lower and higher levels indicated non-physiological ranges.

Accordingly, we performed a digital image analysis to study the kinetic of the solution color development correlated to each antioxidant level. Five samples, corresponding to the five different antioxidant classes, were tested. During such analysis, each sample was placed on a white card inside a black box, and images were captured with a smartphone every 30 seconds for 5 minutes. Later, a small region of interest (ROI) located within the tube was selected manually and cropped from each image, to measure the average color intensity. Fig. 1 shows the average color intensity over time for each antioxidant class. The kinetic curves confirmed that the common trend observed, with a decrease in signal over time, depended on the antioxidant level: lower antioxidant concentrations (class 0 and 1) exhibited a rapid decrease in the average color intensity over time, due to the shift of the color solution from transparent to blue, unlike higher concentrations (class 3 and 4), which showed a slower reaction kinetic. Moreover, all the kinetic curves tended to achieve a similar average color intensity after 4–5 minutes. Therefore, the timeframe between 1 and 2.5 minutes appeared as the best interval to distinguish among different antioxidant levels. This time-resolved analysis was particularly important for understanding the kinetic of the reactions and finding the perfect window to maximize the discrimination efficiency among the 5 antioxidant levels, since they reached similar end-point color intensities while following different kinetic paths.

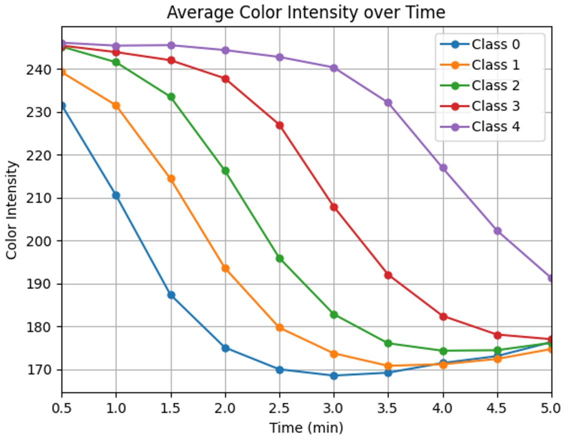

FIGURE 1. Average color intensity over time showing the kinetic behaviors for different classes of TAC.

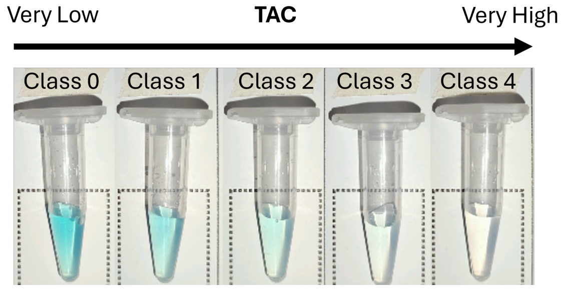

FIGURE 2. Representative images of vial samples corresponding to the 5 classes, illustrating the distinct color intensities observed at time = 2 min.

The naked-eye readout of the 5 test tubes at 2 minutes confirmed the kinetic study results, with a solution color gradient from blue to transparent as the antioxidant level increases, as shown in Fig. 2.

2.3. Data Collection and Preprocessing

Saliva samples were collected multiple times from several healthy donors over a few months. During the data collection, the colorimetric test was applied, and each reaction was divided into two parts: one part was employed for image collection using a smartphone (i.e., Xiaomi 11 lite), whereas the other half was used to measure the absorbance value at 2 minutes with a UV-vis spectrophotometer. For each test reaction, four images were captured using the smartphone at the following timestamps: $t_{\text{test}} = 60$ sec, $t_{\text{test}} = 90$ sec, $t_{\text{test}} = 120$ sec, and $t_{\text{test}} = 150$ sec.

As a result, we constructed 4 datasets, namely 1_00, 1_30, 2_00, and 2_30 (corresponding to the specific timestamp at 60, 90, 120, and 150 sec, respectively). The absorbance value at 120 seconds was used as the sample label. Table 1 lists the dataset for the 5 assay classes, associated with a specific absorbance range and antioxidant level. A total of 285 tests were conducted, providing a balanced dataset.

TABLE 1. Dataset representation in terms of absorbance range, antioxidant levels across different classes, and number of samples in each class.

	Absorbance range (a.u.)	**Antioxidant level**	**Nb. of samples**
Class 0	> 0.90	Very low	(57)
Class 1	0.70–0.90	Low	(57)
Class 2	0.50–0.70	Average	(57)
Class 3	0.30–0.50	High	(57)
Class 4	< 0.30	Very high	(57)

2.4. Data Augmentation

Creating a dataset is a major challenge, particularly due to the process of preparing, collecting, and labeling each test sample which can be time-consuming and expensive. In our test, the reaction vial should be placed within a designated white area marked by dotted lines (Fig. 3). However, human errors, time constraints, and lack of precision often cause the vials to fall outside the specified location. In our application, we needed to extract a small image of the vial to reduce the computational load. However, cropping an image at fixed coordinates increases the risk of capturing only a small part of the reaction solution within the vial, potentially affecting the model decision later. Additionally, the size of the dataset was insufficient for training an ML model to solve a 5-class classification problem with a high accuracy. Therefore, we tried to address these challenges by applying data augmentation as shown in Fig. 3. First, a cropped image of size 150 × 150 Pixels is divided into four equal square crops, each of size 75 × 75. Then, a fifth crop is applied by extracting an image of the same size from the central region. Thus, each image sample is augmented into five distinct images, effectively expanding the dataset to improve model training. This augmentation approach should allow the models to accurately classify images even when only a small portion of the vial is presented within the fixed crop. We applied the same approach for each of the 4 datasets, increasing the number of samples from 285 to 1425 images. The augmented datasets were named 1_00_Aug, 1_30_Aug, 2_00_Aug, and 2_30_Aug as reported in Table 2.

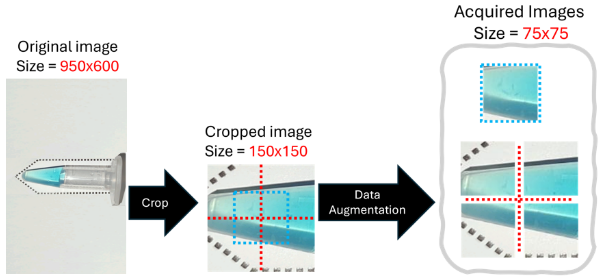

FIGURE 3. Data augmentation process for a single input image.

TABLE 2. Augmented dataset.

Dataset Name	Nb. of classes	Nb. of samples for each class	Total Nb. of samples
1_00_Aug 1_30_Aug 2_00_Aug 2_30_Aug	(5)	(285)	(1425)

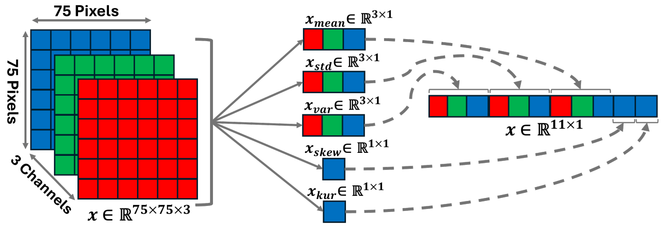
75 Pixels
75 Pixels
3 Channels
$x \in \mathbb{R}^{75\times75\times3}$
$x_{mean} \in \mathbb{R}^{3\times1}$
$x_{std} \in \mathbb{R}^{3\times1}$
$x_{var} \in \mathbb{R}^{3\times1}$
$x_{skew} \in \mathbb{R}^{1\times1}$
$x_{kur} \in \mathbb{R}^{1\times1}$
$x \in \mathbb{R}^{11\times1}$

FIGURE 4. Feature extraction.

2.5. Feature Extraction

To create an effective representation of the image data and reduce its dimensionality, we extracted specific statistical features from each sample image. Before that, all images were converted to the HSV color space as part of the preprocessing of the dataset. These features included the mean (x_{mean}), standard deviation (x_{std}), and variance (x_{var}) of the color channels, and the skewness (x_{skew}) and kurtosis (x_{kur}) of the blue channel only. Consequently, each image X was transformed into a one-dimensional array $x \in \mathbb{R}^{11\times1}$, representing one sample in the dataset as sketched in Fig. 4. Therefore, the new datasets, created from extracting the features of the datasets presented in Table 2, will be named as 1_00_Feat, 1_30_Feat, 2_00_Feat, and 2_30_Feat, respectively.

2.6. Classifiers

To classify the solution's color intensity, we employed four machine learning algorithms: a Support Vector Machine (SVM), a Single-layer Feed-Forward Neural Network (SLFNN), a K-Nearest Neighbors (KNN), and a Convolution Neural Network (CNN).

The SVM, SLFNN, and KNN were trained using the datasets after the feature extraction. These classifiers are the state-of-the-art tools used for classification problems. They can provide high performance when being trained on handcrafted features (i.e., statistical parameters) [41]. On the other hand, the CNN was trained directly on the data-augmented datasets (Table 2). These algorithms were employed to solve a 5-class color intensity classification problem for TAC. In the following, the algorithms are briefly described.

2.6.1. Support Vector Machine

The Support Vector Machine (SVM) classifier is a supervised ML algorithm that computes the hyperplane that maximizes the margin to the nearest samples of the two classes (i.e., the support vectors). SVM is capable of capturing complex relationships between data points. We have implemented only the linear kernel due to its efficiency in terms of memory and computation [57]. The inference of one input datum $\mathbf{z}$ follows the One-vs-One (OvO) strategy that splits the multi-class classification into one binary classification problem per each pair of antioxidant levels, solving (1):

$$y = \text{sign}(\mathbf{w} \cdot \mathbf{z} + b) \tag{1}$$

where $\mathbf{w}$ and b are the support vector and bias, respectively. The eventual label of $\mathbf{z}$ is assigned according to the majority of votes among the predicted classes.

2.6.2. Single-Layer Feed-Forward Neural Network

Single-layer Feed-Forward Neural Networks (SLFNNs) are fully connected networks trained through backpropagation. These models mimic the structure of the human brain's neural network such that all neurons in one layer are connected to the neurons in the next layer. The prediction function of an SLFNN is:

$$f(\mathbf{z}) = \text{softmax}\left(\sum_{i=1}^{Neu} \beta_{i,j}\phi(\mathbf{z}\cdot\mathbf{w}_i + b_i) + B_{i,j} \text{ with } j = 1,\dots,5\right) \tag{2}$$

where **z** is the tested datum; Neu is the number of hidden neurons; $\beta_{i,j}$ and $B_{i,j}$ are the weights and biases between the hidden and output layer, respectively; and ϕ is the ReLU activation function. b_i and **w** are the biases and weights between the input and hidden layers, respectively, and $softmax$ is the Softmax function to predict the label. The output layer contains 5 neurons, equal to the number of antioxidant concentration classes, indexed by j in the equation.

2.6.3. K-Nearest Neighbors

The K-Nearest Neighbors (KNN) is a simple supervised ML algorithm that assigns a class to a given input sample **z**, based on the majority class among its $K > 1$ nearest neighbors in the previously labeled points. The neighbors are determined using a distance metric between the input sample and the points in the training dataset. Its efficiency and performance depend on the number of neighbors (k), distance metric, and training data size. The training phase produces a very simple model (K parameter), but the inference phase requires exploring the whole training set.

The relation between the predicted class $\hat{y}$**z** for an input sample **z** is given by:

$$\hat{y}(\mathbf{z}) = \arg\max_{y\in\mathcal{Y}} \sum_{i=1}^{K} I(y_i = y), \tag{3}$$

2.6.4. Convolutional Neural Network

A Convolutional Neural Network (CNN) is one of the most popular deep neural networks used in a multitude of applications. CNN is composed of different building blocks such as convolutional, pooling, and fully connected layers. Unlike the SVM, SLFNN, and KNN which use hand-crafted features, CNNs can combine feature extraction and classification into a single learning body. Moreover, CNNs proved their effectiveness when being applied to image data decoding [58]. In this work, the input of the CNN is the image collected in the datasets presented in Table 2. The CNN consists of several functional blocks composed of one 2-D convolutional layer with ReLU activation, a dropout layer, and an average pooling layer. The blocks are stacked sequentially based on the number of convolutions chosen by the designer. Here, the number of blocks was set through the $filter$ parameter listed below. At the bottom of the functional blocks, a dense layer with 10 neurons and ReLU activation is stacked, followed by a dense layer with a softmax activation function to perform classification. The 10-neuron dense layer receives the features extracted by the functional blocks and is flattened through a $Flatten$ layer.

2.7. Ensemble Learning and Stacking

Ensemble learning is a machine learning paradigm that combines predictions from multiple models to achieve superior performance compared to individual models alone. The primary objective of ensemble methods is to utilize the unique strengths and complementary weaknesses of various models to increase robustness, reduce generalization error, and improve predictive accuracy. Common ensemble techniques include bagging, boosting, and stacking, each of which enhances model performance through distinct strategies. In this work, we used Stacking also called stacked generalization, which is an ensemble learning technique where predictions from multiple base models are used as input features for a higher-level model, often referred to as a meta-learner. In this approach, base models are trained independently on different datasets derived from the same experimental trial, and their outputs — typically probability predictions for each class — serve as new features for the meta-learner. The meta-learner then learns to interpret these outputs to make a final prediction. This method effectively integrates diverse predictive perspectives, leading to improved accuracy as the meta-learner can exploit patterns in the predictions of the base models that might not be apparent from any single model alone.

In this study, we employed stacking by first training four CNNs on different datasets for a 5-class classification task. The probability predictions from these CNN models were then used as input features to train a Support Vector Machine (SVM) meta-learner. This ensemble approach was designed to utilize the discriminative power of CNNs in feature extraction and classification. At the same time, the SVM used the combined feature space to refine and enhance predictive accuracy. Stacking, therefore, was selected as an optimal technique for integrating predictions from complementary models, thus addressing the limitations of individual CNNs and yielding a more robust and precise final classification.

2.8. Object Detection

Furthermore, we implemented the YOLOv4-tiny (You Only Look Once, Version 4) object detection model to precisely locate and detect regions of interest within the collected dataset. YOLOv4 is a state-of-the-art deep learning model known for its high-speed, real-time object detection capabilities and accuracy in identifying multiple objects within an image. Unlike traditional approaches that rely on fixed cropping to isolate regions for classification, YOLOv4 processes the entire image in a single pass, allowing for the precise detection of specific regions — such as in our case the reaction solution within a vial or the absence of the vial itself — in a single forward operation. By training the smaller variant YOLOv4, namely YOLOv4-tiny, on our dataset, we aimed to extract regions of interest dynamically and accurately, thereby analyzing if we can enhance the accuracy of a single model without the need for ensemble methods. The adoption of YOLOv4-tiny was intended to optimize detection precision and provide a dynamic focus on regions critical for classification.

3. EXPERIMENTAL SETUPS

3.1. Algorithms Hyper-Parameters

The training procedure involved model selection, i.e., tuning the classifier architecture hyper-parameters. Such a procedure explored a grid of candidates for each algorithm. The best candidates were selected by evaluating the accuracy of the validation set during the training phase.

The hyper-parameter for the SVM was the regularizer λ that has been chosen as the norm L2 in the range $\lambda = [10^i, with\ i = -4, -3, \dots, 4]$ during the training procedure.

The hyper-parameters for the KNN were:

- number of neighbors, $k = \{3, 5, 7, 9\}$;
- weighting strategy, weights $= \{\text{uniform, distance}\}$;
- distance metric, metric $= \{\text{euclidean, manhattan}\}$.

The hyper-parameters, including the value of k weighting strategy and the choice of distance metric were optimized through cross-validation to maximize classification accuracy. The hyper-parameters for the SLFNN were:

- hidden neurons $Neu = [10 * i, with\ i = 1, 2, \ldots, 10]$;
- L2 regularizer $\lambda = [10^i, with\ i = -4, -3, \ldots, 4]$.

The hyper-parameters for the CNN were:

- number of convolutional layers from 2 to 4 (the filter candidates were: $(5,5)$, $(8,8)$, $(12,12)$, $(12,24)$, $(5,8,16)$, $(5,5,5)$, $(8,8,8)$);
- kernel size $Ks = \{6, 8, 12, 16\}$;
- dropout percentage 20%.

The filters represented the number of kernels applied to the input features in each functional block. For example, setting $f = (5,8,16)$ implied the use of three functional blocks, the application of 4 kernels to the input tensor, dropout, average pooling, and doubling in the number of kernels at each next functional block. For each of the training algorithms, a combination of these parameters was evaluated during training, and the model with the highest validation accuracy was selected for testing.

The hyper-parameters for the YOLOv4-tiny model were configured to optimize object detection performance. These parameters included:

- **Batch size** and **subdivisions**: Set to 64 and 16, respectively, to balance training efficiency and GPU memory usage.
- **Input resolution**: The width and height were both set to 416 pixels, a common setting that balances detection accuracy and computational efficiency.
- **Learning rate**: Initialized at 0.001 with a burn-in period of 1000 iterations, allowing the model to stabilize before adjusting weights.
- **Maximum iterations**: Configured to 6000, providing sufficient training epochs given the dataset size and model convergence rate.
- **Classes and labels**: Defined for two object classes, “Reaction solution” and “No vial”.

The model was trained using Darknet on Google Colab, utilizing GPU acceleration. Validation metrics, including mean Average Precision (mAP), were monitored throughout the training to assess and refine model performance.

3.2. Training Strategy

As part of the preprocessing pipeline, all images were converted from RGB color space to HSV color space. This transformation was applied due to HSV’s improved capability to capture color intensity features, which was observed to enhance model performance compared to using RGB color space directly. Before training, the datasets were also normalized to ensure that all input features were on a comparable scale, which is essential for improving model convergence and performance. A MinMaxScaler was applied to the reshaped training data, scaling the values between 0 and 1. This normalization step helps in the stabilization of the training process and makes it easier for the models to learn efficiently by reducing the impact of varying feature magnitudes.

3.2.1. Single Learning

To evaluate the performance of the algorithms on color intensity classification for our colorimetric antioxidant test, the SVM, KNN, and SLFNN were trained on each of the datasets containing the handcrafted features (i.e., 1_00_Feat, 1_30_Feat, 2_00_Feat, and 2_30_Feat), which were extracted after data augmentation as described in Section 2.5. CNN models were trained directly on each of the Augmented datasets shown in Table 2. Each dataset was randomly split into training, validation, and testing sets to ensure an unbiased evaluation. Specifically, each dataset contained 1425 samples, divided as follows: 70% of the data was allocated for training (200 image samples per class) and the remaining 30% for testing (85 image samples per class). This split allowed us to train robust models while retaining a substantial test set for performance assessment.

3.2.2. Ensemble Learning

To evaluate the performance of our ensemble learning approach (mentioned in Section 2.7), we trained and validated the CNN models on augmented data, splitting each dataset into three distinct groups to facilitate the training and testing of the meta-learner (SVM). The datasets were randomly divided as follows: 53.68% for training the CNN models, 28.77% for testing the CNN models (with their prediction probabilities subsequently used to train the SVM meta-classifier), and the remaining 17.55% reserved for testing the SVM. In this final group, images were first passed through the CNN models, and the resulting prediction probabilities were then utilized as input features to evaluate the SVM model’s performance.

We investigated two configurations for the ensemble learning process. In the first configuration, we selected only two CNNs trained on the 1_30_Aug and 2_00_Aug datasets. The prediction probabilities from these two models were used as input features for the SVM meta-classifier, resulting in a 10-dimensional input array (5 classes × 2 models). This approach allowed us to examine the effect of a streamlined ensemble that focused on models trained on datasets corresponding to different critical time points (90 and 120 seconds), which showed strong performance in preliminary analysis. Fig. 5 illustrates this configuration.

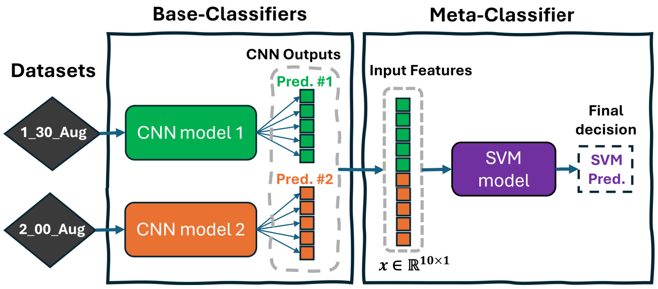

FIGURE 5. Ensemble learning with two CNN base models feeding prediction probabilities into an SVM meta-classifier.

In the second configuration, we increased the number of base models by using the prediction probabilities from all four CNN models trained on the augmented datasets (1_00_Aug, 1_30_Aug, 2_00_Aug, and 2_30_Aug) as input features for the SVM meta-classifier. Given that each CNN model outputs a probability distribution over five classes, the combined input feature for the SVM consisted of a 20-dimensional array (5 classes × 4 models), capturing the predictive perspectives of each CNN model across the four datasets. By using these configurations, we aimed to determine whether combining predictions from all four models can provide a substantial advantage over a more selective two-model ensemble.

TABLE 3. Summary of training strategies for single learning, ensemble learning, and learning with object detection approaches, including the learning algorithms, datasets, number of samples, and number of trained models.

Learning approaches		Learning Algorithms	Train samples (per class)	Test samples (per class)	Used datasets	N. of trained models
Single Learning		SVM	200	85	All feature extracted datasets	4
		SLFN	200	85	All feature extracted datasets	4
		KNN	200	85	All feature extracted dataset	4
		CNN	200	85	Data augmented datasets: Table 2	4
Ensemble Learning	1st config.	CNN	153	82	1_30_Aug 2_00_Aug	2
		SVM	82	50	The pred. prob. of the 2 CNN models	1
	2nd config.	CNN	153	82	1_00_Aug 1_30_Aug 2_00_Aug 2_30_Aug	4
		SVM	82	50	The pred. prob. of the 4 CNN models	1
Learning with Object detection		CNN	39	16	2_00_yolo	1

3.2.3. Learning With Object Detection

To train the object detection model, we utilized a dataset of 500 images containing vials with solutions of varying color intensities and images without vials. For annotation, we exploited the open-source Python tool LabelImg [59], which facilitates the creation of bounding boxes to label objects in images. Using this tool, the regions containing the reaction solution in images with vials were labeled as "Reaction solution", while images without vials were labeled as "No vial". The tool generated annotation files containing the object labels and coordinates, which were later used along with the original images to train the YOLOv4-tiny model.

After training, a 16-bit quantization was applied to the YOLOv4-tiny model to reduce its size, optimizing it for more efficient deployment. Following this, the original dataset named 2_00, as described in Section 2.3, was processed using the YOLOv4-tiny model. Post-processing steps were then applied to extract and refine the detected objects (i.e., the reaction solution), by resizing the extracted image to a size of 110x70 pixels. The resulting processed dataset, containing cropped images of the detected solutions, was named 2_00_yolo.

To evaluate the impact of object detection on classification performance, a single CNN model was trained on the 2_00_yolo dataset. This model was trained without applying data augmentation. The training strategy was similar to the one described in Section 3.2.1 for the CNN model. The dataset contained a total of 285 samples, randomly divided as follows: 68.4% for training (39 image samples per class) and the remaining 31.6% for testing (18 image samples per class). This setup should enable us to assess whether the YOLOv4-CNN approach can improve the classification accuracy of the TAC level, compared to the ensemble methods, by focusing the model's attention on the most important part of the image.

The training strategies for single learning, ensemble learning, and learning with object detection are summarized in Table 3, highlighting the learning algorithms used, datasets, sample splitting, and the number of trained models for each approach.

4. RESULTS AND DISCUSSION

The results of this study provide insights into the performance and suitability of different machine learning models for classifying color intensity levels in the colorimetric antioxidant test. Various models, including SVM, SLFNN, KNN, and CNN, along with three different learning approaches, were evaluated on multiple datasets representing distinct time intervals and total antioxidant concentration levels. This section presents and analyzes the classification accuracy of these models, examining trends across datasets and exploring the impact of ensemble and object detection methods to enhance classification precision.

Table 4 presents the classification accuracies of CNN, SVM, KNN, and SLFNN across four datasets. For dataset 1_00_Aug, CNN

TABLE 4. Models accuracy.

Datasets	CNN	SVM	KNN	SLFNN
1_00_Aug	73.26	60.05	47.27	54.54
1_30_Aug	75.28	67.36	59.02	65.45
2_00_Aug	77.76	76.54	70.09	76.36
2_30_Aug	64.74	58.18	58.18	61.81

achieved the highest accuracy at 73.26%, outperforming SVM, KNN, and SLFNN, which showed comparatively lower accuracies. In dataset 1_30_Aug, CNN maintained strong performance at 75.28%, while SVM and SLFNN reached 67.36% and 65.45%, respectively, making them possible alternatives. The dataset 2_00_Aug yielded the highest overall accuracies, with CNN reaching 77.76% and SVM and SLFNN close behind, indicating it as the most favorable dataset for classification. In dataset 2_30_Aug, all models showed reduced performance, suggesting unique challenges with this dataset. Overall, CNN consistently outperformed other models, highlighting its robust classification ability across the different datasets.

Upon examining the confusion matrices of the CNN models trained on the four augmented datasets, as shown in Fig. 6, distinct classification trends were observed across different time intervals. For the dataset collected at 60 seconds (1_00_Aug), the CNN model showed higher accuracy in classifying the lower antioxidant concentration classes (0,1). This suggests that, at this early time point, the visual differences among lower concentration samples were more distinguishable, aiding classification accuracy for those classes. At 90-seconds (1_30_Aug), the CNN model displayed improved performance in identifying the middle concentration classes (2,3). This shift implies that, as time progresses, the reaction stabilizes in a way that makes middle concentrations more distinct, thus enhancing classification accuracy for these classes. For the 120-second (2_00_Aug) dataset, the CNN achieved the highest accuracy among all datasets, reaching 77.76% accuracy. This dataset appears to provide the most balanced information across all classes, allowing for consistent classification performance across lower, middle, and higher antioxidant concentrations. In contrast, for the 150-second (2_30_Aug) dataset, the CNN model showed better accuracy in classifying higher antioxidant concentration classes (3,4). This could be due to a more prominent reaction effect at this time point, making higher concentrations distinct while reducing contrast in lower concentration classes. However, overall accuracy was lower for this dataset (64.74%), suggesting that in the late stage of the reaction, it is challenging to classify accurately.

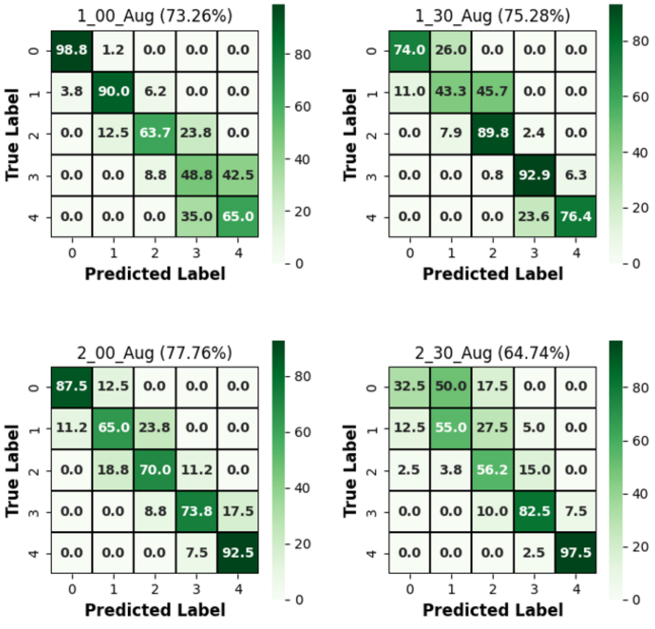

FIGURE 6. Confusion matrices for CNN models across four datasets, illustrating classification performance variations across different time intervals and antioxidant concentration levels.

Given these observations from Fig. 6, it was evident that the individual CNN models were each better suited to specific concentration ranges, depending on the dataset. To address this limitation and improve overall predictive accuracy, we implemented an ensemble learning strategy testing two configurations (see Section 2.7). In the first configuration, the SVM meta-classifier achieved a classification accuracy of 90%, providing a significant improvement of approximately 12% over the best individual CNN model (77.76%). The second configuration (all 4 models combined together) further increased the accuracy to 92%; however, the additional improvement was insufficient considering the increased complexity of stacking four CNN models. This suggests that the selective use of fewer but well-performing base models (as in the first configuration) can achieve comparable results with reduced complexity. The confusion matrices for the SVM meta-classifiers of these two configurations are shown in Fig. 7, demonstrating that the accuracy across all classes was consistently high, without significant disparities between class performances. Moreover, other configurations were also tested but showed either lower or almost similar accuracies but were not reported here.

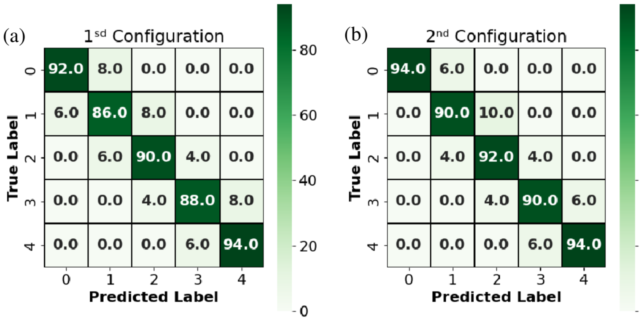

FIGURE 7. Confusion matrices for the two stacking ensemble configurations: (a) the first configuration achieving 90% accuracy; and (b) the second configuration combining all four CNN models, achieving 92% accuracy.

The improved accuracy in both ensemble configurations demonstrated the effectiveness of stacking in utilizing the complementary strengths of models trained on different datasets. By combining models specialized in detecting specific concentration ranges, we were able to create a more robust classifier capable of handling the full range of total antioxidant concentrations. These findings suggest that stacking, particularly with a diverse selection of time-specific models, can significantly enhance classification performance. It also shows the trade-off between complexity and performance, as the selective use of fewer models still provided a significant accuracy compared to the complex stacking of 4 CNN models. The results also emphasize the potential of ensemble learning for complex classification tasks, where individual models struggle to capture all details in the data.

We then explored the potential of YOLOv4-tiny to further improve our detection accuracy. After training the YOLOv4-tiny model, we

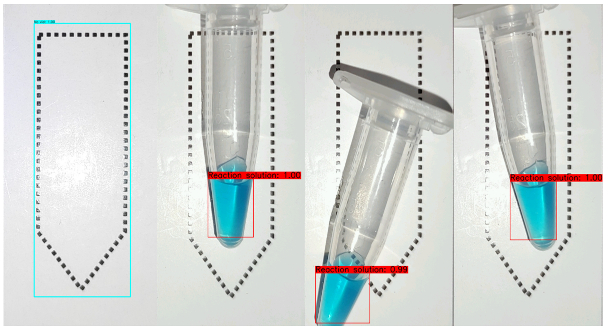

FIGURE 8. Detection results from the YOLOv4–tiny model for both objects: 'No vial' and 'Reaction solution', having different vial positions.

evaluated its performance using different test images. The model consistently detected the objects in all the images with high confidence, accurately identifying and localizing the regions of interest (ROI) (i.e., the 'Reaction solution' and 'No vial'). The output from the model was used to draw a bounding box around the detected objects, as shown in Fig. 8. Recognizing that our initial stacking model with four CNNs followed by an SVM meta-classifier (five models in total) achieved promising results, we hypothesized that using only an object detection model followed by a CNN model could significantly improve accuracy and efficiency. The results were very interesting, and the CNN model achieved an impressive classification accuracy of 97.78% solving the five-class classification task. This approach, consisting of only two models, a YOLOv4-tiny model followed by a CNN, outperformed the original ensemble of four CNNs plus an SVM meta-classifier, which required a more complex stacking configuration. Hence, by reducing the number of models to just two blocks, we observed not only an increase in the accuracy but also a simplification of the overall model structure. The confusion matrix for this CNN model is shown in Fig. 9, highlighting its excellent classification performance over the whole TAC concentration range.

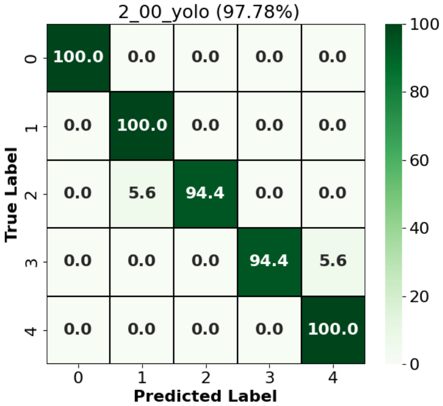

FIGURE 9. Confusion matrix for the CNN model trained on the 2_00_YOLO dataset, achieving 97.78% accuracy in the five-class classification task.

This significant improvement underscores the efficacy of integrating object detection with classification for tasks requiring precise localization. The ability of YOLOv4-tiny to detect precisely the relevant region allowed the CNN to focus exclusively on the critical area, thereby improving classification accuracy and eliminating the need for additional model layers. These results demonstrate that, in applications where regions of interest can be accurately localized, employing object detection in conjunction with classification can yield a streamlined, high-performance model. This approach also reduces the computational load, as it bypasses the need for stacking multiple models and instead uses a targeted pre-processing step to enhance classification accuracy. Moreover, the quantization step enabled the model to maintain its detection capabilities while significantly decreasing memory requirements, decreasing the size of the model from approximately 22.47 MB to 11.28 MB, making it suitable for real-time applications on resource-constrained devices.

Ultimately, the YOLOv4-tiny and CNN models were converted to TensorFlow Lite format (.tflite) for efficient real-time deployment of the YOLOv4-CNN approach on Android devices, utilizing Kotlin and TensorFlow Lite's native support. The phone application was programmed to capture the reaction image at $t = 2$ minutes, apply post-processing through the YOLOv4-tiny model to localize the ROI, and extract the detected region. The extracted ROI was then resized to a standardized size (e.g., 110 × 70 pixels) and normalized using the MinMax scaler. The scaler parameters were precomputed and saved in Python, allowing for consistent normalization during deployment. Finally, the processed ROI was passed to the CNN for classification. As a result, real-time processing was achieved on the phone, ensuring low latency and high accuracy. This integration not only validated the feasibility of deploying our architecture for mobile environments but also paved the way for its use in other practical point-of-care applications requiring real-time processing and decision-making.

5. CONCLUSION

In this work, we have developed a novel approach for smartphone-based testing antioxidant concentration in saliva, utilizing smartphone-based machine learning and object detection techniques to achieve accurate and efficient classification. Given the importance of TAC levels as a biomarker for health, this method addresses a critical need for accessible, real-time health monitoring. By capturing images of the reaction vials at different intervals and building a comprehensive dataset, we trained and evaluated multiple classifiers — CNN, SVM, KNN, and SLFNN — on color intensity classification, with CNN models consistently achieving the highest performance.

To further enhance classification accuracy, we implemented an ensemble learning approach through stacking, combining the outputs of four CNN models with an SVM meta-classifier, resulting in a classification accuracy of 92%. However, the integration of object detection with YOLOv4-tiny enabled the extraction of precise regions of interest within the images, allowing a CNN model trained on this refined dataset to achieve an impressive 97.78% accuracy using one image that is extracted 2 minutes after the antioxidant test starts. This YOLOv4-CNN approach simplified the model architecture, provided a very fast test readout (i.e., 2 min), reduced computational requirements, and maintained high accuracy, demonstrating the viability of object detection as an alternative to ensemble methods for region-based classification tasks.

The successful deployment of the YOLOv4-tiny and CNN models on an Android application further highlights the practical applicability of this system, enabling real-time analysis on a smartphone platform. This advancement provides a robust and efficient solution for POC antioxidant testing, with the potential for broader applications in other biomarker-based health assessments. Future work may focus on generalizing this approach to accommodate diverse clinical assays, such as those for glucose, proteins, or pH levels, by utilizing the framework's modularity and retraining models on assay-specific datasets. Additionally, we aim to enhance the robustness of the system by addressing variability introduced by differences in device optics, ensuring consistent image quality and reliable performance across a range of hardware setups and extending the system to cover different biological fluids (e.g., sweat, urine) or food matrices. Finally, continued efforts will aim to further optimize models' performance on mobile devices to enhance accessibility and diagnostic reliability, making it a scalable solution for real-world healthcare applications.

REFERENCES

[1] Ragusa, E., V. Mastronardi, D. Pedone, M. Moglianetti, P. P. Pompa, R. Zunino, and P. Gastaldo, "Random weights neural network for low-cost readout of colorimetric reactions: Accurate detection of antioxidant levels," in *International Conference on System-Integrated Intelligence*, 95–104, 2022.

[2] Yager, P., T. Edwards, E. Fu, K. Helton, K. Nelson, M. R. Tam, and B. H. Weigl, "Microfluidic diagnostic technologies for global public health," *Nature*, Vol. 442, No. 7101, 412–418, 2006.

[3] Tania, M. H., K. T. Lwin, A. M. Shabut, M. Najlah, J. Chin, and M. A. Hossain, "Intelligent image-based colourimetric tests using machine learning framework for lateral flow assays," *Expert Systems with Applications*, Vol. 139, 112843, 2020.

[4] Zhu, H., I. Sencan, J. Wong, S. Dimitrov, D. Tseng, K. Nagashima, and A. Ozcan, "Cost-effective and rapid blood analysis on a cell-phone," *Lab on a Chip*, Vol. 13, No. 7, 1282–1288, 2013.

[5] Coskun, A. F., J. Wong, D. Khodadadi, R. Nagi, A. Tey, and A. Ozcan, "A personalized food allergen testing platform on a cellphone," *Lab on a Chip*, Vol. 13, No. 4, 636–640, 2013.

[6] Coskun, A. F., R. Nagi, K. Sadeghi, S. Phillips, and A. Ozcan, "Albumin testing in urine using a smart-phone," *Lab on a Chip*, Vol. 13, No. 21, 4231–4238, 2013.

[7] Mutlu, A. Y., V. Kılıç, G. K. Özdemir, A. Bayram, N. Horzum, and M. E. Solmaz, "Smartphone-based colorimetric detection via machine learning," *Analyst*, Vol. 142, No. 13, 2434–2441, 2017.

[8] Kohen, R. and A. Nyska, "Invited review: Oxidation of biological systems: Oxidative stress phenomena, antioxidants, redox reactions, and methods for their quantification," *Toxicologic Pathology*, Vol. 30, No. 6, 620–650, 2002.

[9] Mastalerz-Migas, A., A. Steciwko, M. .Pokorski, I. Pirogowicz, J. Drobnik, A. Bunio, A. Muszyńska, and A. Jasińska, "What influences the level of oxidative stress as measured by 8-hydroxy-2'-deoxyguanosine in patients on hemodialysis?" *Journal of Physiology and Pharmacology: An Official Journal of the Polish Physiological Society*, Vol. 57, 199–205, 2006.

[10] Buczko, P., A. Zalewska, and I. Szarmach, "Saliva and oxidative stress in oral cavity and in some systemic disorders," *Journal of Physiology and Pharmacology*, Vol. 66, 1–7, 2015.

[11] Popolo, A., G. Autore, A. Pinto, and S. Marzocco, "Oxidative stress in patients with cardiovascular disease and chronic renal failure," *Free Radical Research*, Vol. 47, No. 5, 346–356, 2013.

[12] Mangge, H., K. Becker, D. Fuchs, and J. M. Gostner, "Antioxidants, inflammation and cardiovascular disease," *World Journal of Cardiology*, Vol. 6, No. 6, 462–477, 2014.

[13] Fusco, D., G. Colloca, M. R. L. Monaco, and M. Cesari, "Effects of antioxidant supplementation on the aging process," *Clinical Interventions in Aging*, Vol. 2, No. 3, 377–387, 2007.

[14] Bjørklund, G., M. Shanaida, R. Lysiuk, H. Antonyak, I. Klishch, V. Shanaida, and M. Peana, "Selenium: An antioxidant with a critical role in anti-aging," *Molecules*, Vol. 27, No. 19, 6613, 2022.

[15] Reuter, S., S. C. Gupta, M. M. Chaturvedi, and B. B. Aggarwal, "Oxidative stress, inflammation, and cancer: How are they linked?" *Free Radical Biology and Medicine*, Vol. 49, No. 11, 1603–1616, 2010.

[16] Bahar, G., R. Feinmesser, T. Shpitzer, A. Popovtzer, and R. M. Nagler, "Salivary analysis in oral cancer patients: DNA and protein oxidation, reactive nitrogen species, and antioxidant profile," *Cancer*, Vol. 109, No. 1, 54–59, 2007.

[17] Tulunoglu, O., S. Demirtas, and I. Tulunoglu, "Total antioxidant levels of saliva in children related to caries, age, and gender," *International Journal of Paediatric Dentistry*, Vol. 16, No. 3, 186–191, 2006.

[18] Pedone, D., M. Moglianetti, M. Lettieri, G. Marrazza, and P. P. Pompa, "Platinum nanozyme-enabled colorimetric determination of total antioxidant level in saliva," *Analytical Chemistry*, Vol. 92, No. 13, 8660–8664, 2020.

[19] Sateanchok, S., S. Wangkarn, C. Saenjum, and K. Grudpan, "A cost-effective assay for antioxidant using simple cotton thread combining paper based device with mobile phone detection," *Talanta*, Vol. 177, 171–175, 2018.

[20] Frankel, E. N., "Antioxidants in lipid foods and their impact on food quality," *Food Chemistry*, Vol. 57, No. 1, 51–55, 1996.

[21] Scarsi, A., D. Pedone, and P. P. Pompa, "A multi-line platinum nanozyme-based lateral flow device for the colorimetric evaluation of total antioxidant capacity in different matrices," *Nanoscale Advances*, Vol. 5, No. 8, 2167–2174, 2023.

[22] Puangbanlang, C., K. Sirivibulkovit, D. Nacapricha, and Y. Sameenoi, "A paper-based device for simultaneous determination of antioxidant activity and total phenolic content in food samples," *Talanta*, Vol. 198, 542–549, 2019.

[23] Choices, N., "Colour vision deficiency (colour blindness)," 2016, [Online Accessed: Sep. 1, 2024]. [Online]. Available: nhs.uk.

[24] Awareness, C. B., "Colour blindness," 2018, [Online Accessed: Nov. 1, 2024].

[25] Liu, W., S. Liu, K. Fan, Z. Li, Z. Guo, D. Cheng, and G. Liu, "Machine-learning-based colorimetric sensor on smartphone for salivary uric acid detection," *IEEE Sensors Journal*, Vol. 24, No. 20, 32 991–33 000, 2024.

[26] Taccioli, T., E. Ragusa, T. Pomili, P. Gastaldo, and P. P. Pompa, "Semi-quantitative determination of thiocyanate in saliva through colorimetric assays: Design of CNN architecture via input-aware NAS," *IEEE Sensors Journal*, Vol. 23, No. 23, 29 869–29 876, 2023.

[27] Mercan, O. B., V. Kılıç, and M. Şen, "Machine learning-based colorimetric determination of glucose in artificial saliva with different reagents using a smartphone coupled µpad," *Sensors and Actuators B: Chemical*, Vol. 329, 129037, 2021.

[28] Fan, K., W. Liu, Y. Miao, Z. Li, and G. Liu, "Engineering strategies for advancing optical signal outputs in smartphone-enabled point-of-care diagnostics," *Advanced Intelligent Systems*, Vol. 5, No. 6, 2200285, 2023.

[29] Jia, M.-Y., Q.-S. Wu, H. Li, Y. Zhang, Y.-F. Guan, and L. Feng, "The calibration of cellphone camera-based colorimetric sensor array and its application in the determination of glucose in urine," *Biosensors and Bioelectronics*, Vol. 74, 1029–1037, 2015.

[30] Lopez-Ruiz, N., V. F. Curto, M. M. Erenas, F. Benito-Lopez, D. Diamond, A. J. Palma, and L. F. Capitan-Vallvey, "Smartphone-based simultaneous ph and nitrite colorimetric determination for paper microfluidic devices," *Analytical Chemistry*, Vol. 86, No. 19, 9554–9562, 2014.

[31] Jung, Y., J. Kim, O. Awofeso, H. Kim, F. Regnier, and E. Bae, "Smartphone-based colorimetric analysis for detection of saliva alcohol concentration," *Applied Optics*, Vol. 54, No. 31, 9183–9189, 2015.

[32] Shen, L., J. A. Hagen, and I. Papautsky, "Point-of-care colorimetric detection with a smartphone," *Lab on a Chip*, Vol. 12, No. 21, 4240–4243, 2012.

[33] Fisher, R., K. Anderson, and J. B. Christen, "Using machine learning to objectively determine colorimetric assay results from cell phone photos taken under ambient lighting," in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, 467–470, Lansing, MI, USA, 2021.

[34] Geng, Z., Y. Miao, G. Zhang, and X. Liang, "Colorimetric biosensor based on smartphone: State-of-art," *Sensors and Actuators A: Physical*, Vol. 349, 114056, 2023.

[35] Bhatt, S., S. Kumar, M. K. Gupta, S. K. Datta, and S. K. Dubey, "Colorimetry-based and smartphone-assisted machine-learning model for quantification of urinary albumin," *Measurement Science and Technology*, Vol. 35, No. 1, 015030, 2023.

[36] Sajed, S., M. Kolahdouz, M. A. Sadeghi, and S. F. Razavi, "High-performance estimation of lead ion concentration using smartphone-based colorimetric analysis and a machine learning approach," *ACS Omega*, Vol. 5, No. 42, 27 675–27 684, 2020.

[37] Ashraf, S. and T. Ahmed, "Sagacious intrusion detection strategy in sensor network," in *2020 International Conference on UK-China Emerging Technologies (UCET)*, 1–4, Glasgow, UK, 2020.

[38] STMicroelectronics, "STM32-bit Arm Cortex MCUs," 2022, [Online Accessed: Nov. 20, 2024]. [Online]. Available: https://\protect\penalty-\@Mwww.st.com/en/microcontrollers-microprocessors/stm32-32-bit-arm-cortex-mcus.html.

[39] Cui, F., Y. Yue, Y. Zhang, Z. Zhang, and H. S. Zhou, "Advancing biosensors with machine learning," *ACS sensors*, Vol. 5, No. 11, 3346–3364, 2020.

[40] Kadian, S., P. Kumari, S. Shukla, and R. Narayan, "Recent advancements in machine learning enabled portable and wearable biosensors," *Talanta Open*, Vol. 8, 100267, 2023.

[41] Amin, Y., C. Gianoglio, and M. Valle, "Embedded real-time objects' hardness classification for robotic grippers," *Future Generation Computer Systems*, Vol. 148, 211–224, 2023.

[42] Athira, M. V. and D. M. Khan, "Recent trends on object detection and image classification: A review," in *2020 International Conference on Computational Performance Evaluation (ComPE)*, 427–435, Shillong, India, 2020.

[43] Tania, M. H., K. T. Lwin, A. M. Shabut, K. J. Abu-Hassan, M. S. Kaiser, and M. A. Hossain, "Assay type detection using advanced machine learning algorithms," in *2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, 1–8, Island of Ulkulhas, Maldives, 2019.

[44] Chen, X.-W. and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE Access*, Vol. 2, 514–525, 2014.

[45] Ballard, Z. S., H.-A. Joung, A. Goncharov, J. Liang, K. Nugroho, D. D. Carlo, O. B. Garner, and A. Ozcan, "Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors," *NPJ Digital Medicine*, Vol. 3, No. 1, 66, 2020.

[46] Luo, Y., H.-A. Joung, S. Esparza, J. Rao, O. Garner, and A. Ozcan, "Quantitative particle agglutination assay for point-of-care testing using mobile holographic imaging and deep learning," *Lab on a Chip*, Vol. 21, No. 18, 3550–3558, 2021.

[47] Joung, H.-A., Z. S. Ballard, J. Wu, D. K. Tseng, H. Teshome, L. Zhang, E. J. Horn, P. M. Arnaboldi, R. J. Dattwyler, O. B. Garner, D. D. Carlo, and A. Ozcan, "Point-of-care serodiagnostic test for early-stage Lyme disease using a multiplexed paper-based immunoassay and machine learning," *ACS Nano*, Vol. 14, No. 1, 229–240, 2019.

[48] Amin, Y., C. Gianoglio, and M. Valle, "Towards a trade-off between accuracy and computational cost for embedded systems: A tactile sensing system for object classification," in *International Conference on System-Integrated Intelligence*, 148–159, 2022.

[49] Kim, H., O. Awofeso, S. Choi, Y. Jung, and E. Bae, "Colorimetric analysis of saliva-alcohol test strips by smartphone-based instruments using machine-learning algorithms," *Applied Optics*, Vol. 56, No. 1, 84–92, 2016.

[50] Khanal, B., P. Pokhrel, B. Khanal, and B. Giri, "Machine-learning-assisted analysis of colorimetric assays on paper analytical devices," *ACS Omega*, Vol. 6, No. 49, 33 837–33 845, 2021.

[51] Feng, F., Z. Ou, F. Zhang, J. Chen, J. Huang, J. Wang, H. Zuo, and J. Zeng, "Artificial intelligence-assisted colorimetry for urine glucose detection towards enhanced sensitivity, accuracy, resolution, and anti-illuminating capability," *Nano Research*, Vol. 16, No. 10, 12 084–12 091, 2023.

[52] Duan, S., T. Cai, J. Zhu, X. Yang, E. G. Lim, K. Huang, K. Hoettges, Q. Zhang, H. Fu, Q. Guo, *et al.*, "Deep learning-assisted ultra-accurate smartphone testing of paper-based colorimetric ELISA assays," *Analytica Chimica Acta*, Vol. 1248, 340868, 2023.

[53] Ragusa, E., R. Zunino, V. Mastronardi, M. Moglianetti, P. P. Pompa, and P. Gastaldo, "Design of a quantitative readout in a point-of-care device for cisplatin detection," *IEEE Sensors Letters*, Vol. 6, No. 11, 1–4, 2022.

[54] Liu, X., Q. Wang, Y. Zhang, L. Zhang, Y. Su, and Y. Lv, "Colorimetric detection of glutathione in human blood serum based on the reduction of oxidized TMB," *New Journal of Chemistry*, Vol. 37, No. 7, 2174–2178, 2013.

[55] Josephy, P. D., T. Eling, and R. P. Mason, "The horseradish peroxidase-catalyzed oxidation of 3, 5, 3', 5'-tetramethylbenzidine. Free radical and charge-transfer complex intermediates," *Journal of Biological Chemistry*, Vol. 257, No. 7, 3669–3675, 1982.

[56] Lin, S., D. Zheng, A. Li, and Y. Chi, "Black oxidized 3, 3′, 5, 5′-tetramethylbenzidine nanowires (oxTMB NWs) for enhancing Pt nanoparticle-based strip immunosensing," *Analytical and Bioanalytical Chemistry*, Vol. 411, 4063–4071, 2019.

[57] Sakr, F., F. Bellotti, R. Berta, and A. D. Gloria, "Machine learning on mainstream microcontrollers," *Sensors*, Vol. 20, No. 9, 2638, 2020.

[58] Elngar, A. A., M. Arafa, A. Fathy, B. Moustafa, O. Mahmoud, M. Shaban, and N. Fawzy, "Image classification based on CNN: A survey," *Journal of Cybersecurity and Information Management*, Vol. 6, No. 1, 18–50, 2021.

[59] Tzutalin, "Labelimg: A graphical image annotation tool and label object bounding boxes in images," 2015. [Online]. Available: https://github.com/HumanSignal/labelImg