



Cite this: DOI: 10.1039/d6ay00780e

Smartphone-based colorimetric analysis of pH strips using machine learning

Ece Yıldız,^a Mustafa Şen ^a and Mehmet Akif Özdemir ^{*ab}

This study introduces a machine learning (ML)-enhanced smartphone application designed for the precise colorimetric quantification of pH strips. To ensure the robustness of the system against environmental variations, a comprehensive dataset was constructed by capturing images of pH strips under diverse illumination conditions and camera angles. Following region of interest extraction, an initial set of 33 colorimetric features was employed to train and evaluate 15 different regression models. To ensure model interpretability and computational efficiency, a SHapley Additive exPlanations (SHAP)-based analysis was implemented, successfully identifying six critical descriptors (including color channel skewness, entropy, and intensity metrics) that primarily govern the pH prediction. The best-performing model ($R^2 = 0.99$) was subsequently integrated into a user-friendly Android application, *pHScoper*. This application enables image capture, interactive cropping, and offline, on-device quantitative analysis without cloud reliance. Overall, the developed platform demonstrates strong potential for reliable, low-cost pH measurements in resource-limited settings.

Received 24th April 2026

Accepted 7th June 2026

DOI: 10.1039/d6ay00780e

rsc.li/methods

Introduction

pH measurement is fundamental in healthcare, given that many biological processes are inherently pH-dependent, and deviations from normal pH ranges are often associated with various disease states.^{1–4} Conventional pH meters provide high accuracy; however, their cost, maintenance requirements, and susceptibility to electrode degradation limit their applicability in field and point-of-care settings.^{5,6} As a result, pH indicator strips and paper-based microfluidic devices (μ PADs) have emerged as a simple and cost-effective alternative for rapid pH assessment.⁷ These devices can be incorporated into compact, on-chip platforms, enabling low-volume and on-site analysis. These assays indicate the acidity or alkalinity of a sample through a visible color change in response to hydrogen-ion (H^+) concentration.⁸ Sample pH is typically assessed qualitatively or semi-quantitatively using reference color charts; however, such visual interpretation is inherently subjective and affected by both user perception and environmental conditions. To overcome this limitation, the colorimetric signals of these low-volume assays can be digitized and analyzed computationally. This integration of digital image processing seamlessly transforms simple pH strips into compact, quantitative point-of-care analytical platforms.⁵ By eliminating human bias,

a computational approach may ensure reproducibility and analytical reliability.

To enable this digital quantification, smartphones have emerged as an ideal platform, owing to their advanced computational capabilities and embedded high-resolution cameras as optical sensors.⁹ This technological convergence has facilitated the integration of colorimetric detection into portable analytical systems across diverse biological, chemical, and healthcare applications.^{10–14} In practice, smartphone-based systems process captured images to extract specific quantitative descriptors, primarily utilizing color spaces such as RGB (Red, Green, Blue) and HSV (Hue, Saturation, Value).¹⁵ By correlating these image-derived features with analyte concentrations, robust calibration models can be established, effectively translating complex chromatic responses into precise quantitative measurements.^{16,17}

Unlike conventional rule-based or simple analytical methods, which struggle to model the highly non-linear and non-monotonic colorimetric responses of multi-pad strips,¹⁸ artificial intelligence (AI) architectures are inherently equipped to map these complex feature interactions, enabling accurate, continuous pH interpolation between discrete integer values. In scenarios requiring fast, portable, and efficient analysis, these data-driven approaches play a pivotal role in enhancing colorimetric detection. Consequently, smartphone-based colorimetric methods employing such algorithms have gained popularity due to their affordability, adaptability, and portability.^{19–21} AI-driven colorimetric detection has been successfully applied across environmental, healthcare, and food safety fields. For instance, Mutlu *et al.*¹⁹ used smartphone images of pH strips as a training

^aDepartment of Biomedical Engineering, Izmir Katip Celebi University, Izmir, Turkiye. E-mail: makif.ozdemir@ikcu.edu.tr

^bBiomedical Research in AI & Neuroscience Laboratory, Izmir Katip Celebi University, Izmir, Turkiye



set for a Least Squares Support Vector Machine (SVM) classifier to evaluate illumination effects. Mercan *et al.*²² introduced *GlucoseSensing*, a machine learning (ML)-based portable μ PAD platform for glucose quantification from a smartphone. Similarly, Feng *et al.*²³ developed a nanosensor using convolutional neural networks for glucose detection, and Liu *et al.*²⁴ applied ML to microfluidic paper strips to detect salivary uric acid. These AI-enhanced colorimetric strategies enable low-volume, on-chip analysis, transforming subjective, semi-quantitative assays into reliable quantitative methods. Nevertheless, most AI models operate as “black boxes”, where only the final predictions are observable, making it challenging to quantify the information embedded in the inputs.^{25–27} To address this limitation in scientific and medical contexts, explainable AI (XAI) methods such as SHapley Additive exPlanations (SHAP)²⁸ are increasingly applied to interpret model outputs and identify the most important features. In the context of colorimetric analysis, feature-attribution techniques are particularly relevant, as they quantify and visualize the contribution of image-derived features to model predictions.²⁹ By ranking and aggregating Shapley values, SHAP-based interpretation may facilitate evaluation of whether predictive performance is driven by physically meaningful colorimetric descriptors rather than by illumination variations or background interference.

The strategic integration of explainable AI into smartphone-based colorimetric systems is still largely unexplored, despite the acknowledged necessity for interpretability. Alongside transparency, feature importance assessment is a crucial tool for model optimization in this field. SHAP enables the identification and elimination of redundant or noise-sensitive variables by measuring the precise contribution of particular colorimetric descriptors. This targeted feature selection is highly advantageous for smartphone platforms, as it facilitates the development of lightweight ML models for efficient, on-device computation without the need for high computational power. Furthermore, isolating physically meaningful color features ensures robust and consistent analytical performance across varying environmental and illumination conditions. Addressing these gaps, this study introduces *pHScoper*, a portable and user-friendly Android application that combines

SHAP-based feature optimization with ML-driven colorimetric analysis for quantitative pH determination, providing a reliable point-of-care tool for resource-limited settings.

Materials and methods

The schematic workflow of smartphone-based colorimetric pH detection is illustrated in Fig. 1. The process begins with the immersion of pH strips into 15 different solutions corresponding to discrete pH values (*e.g.*, pH 0, 1, 2, ..., 13, 14). Images of the strips were captured using a smartphone under varying illumination and camera angle conditions to construct a representative dataset. The acquired images were subsequently pre-processed, and colorimetric features were extracted for ML model training. To ensure robust performance across diverse conditions, the initial models underwent validation for generalizability. Afterwards, XAI techniques were employed to interpret feature importance, enabling the identification of a significant subset. The models were then retrained on this reduced feature set, and the optimal model was deployed as a custom smartphone application.

Materials

The materials used included aqueous solutions adjusted to pH values ranging from 0 to 14.0, prepared by controlled titration of sodium hydroxide (NaOH) and hydrochloric acid (HCl). Commercial pH indicator strips (MQuant® 109535.0001, Merck KGaA, Darmstadt, Germany) were used as the colorimetric sensing medium. All solutions were homogenized using a vortex mixer (MX-S, Scilogex LLC, Rocky Hill, USA). Prior to any test strip immersion and image acquisition, the reference pH values of all prepared solutions were experimentally validated using a calibrated benchtop pH meter (AZ 86505 Affordable Benchtop pH & Conductivity Meter, AZ Instrument Corp., Taichung, Taiwan) to ensure the reliability of the reference measurements used for ML model training and evaluation.

Dataset collection

pH strip images were captured using a 12 MP smartphone camera (Apple iPhone 11) with a native resolution of 3024 ×



Fig. 1 The workflow for ML-based pH detection on a smartphone application.



4032 px, a 26 mm focal length lens, and an $f/1.8$ aperture. All experiments and image acquisitions were conducted under stable laboratory conditions at approximately 25 °C. Since measurements were performed within a narrow and controlled temperature range, no temperature correction was applied. The reference pH values were verified using a calibrated benchtop pH meter under the same conditions, consistent with previous studies reporting room-temperature pH measurements for paper-based sensing systems.³⁰ The acquisition of images was performed under approximately constant laboratory ambient conditions while varying illumination and camera angles to reflect real-world variability. Each pH-adjusted solution produced a distinct color change on the pH strip, and each strip corresponding to a given pH value was imaged at least 15 times to expand the dataset and capture experimental variability, resulting in a dataset comprising 1035 images.

To better reflect real-world operational environments, strict hardware-based color calibration and lux measurements were intentionally omitted during image acquisition. Instead,

experimental variability was systematically introduced to evaluate the computational robustness. As shown in Fig. 2, images were captured under diverse, uncalibrated indoor and outdoor conditions. For indoor setups, laboratory fluorescent bulbs (Philips 12 W) provided a neutral-to-cool illumination at approximately 4000 K. Outdoor images were acquired under partially cloudy daylight (~5000–6500 K) to represent diffuse natural lighting. To simulate practical scenarios involving multiple light sources, dual-illumination configurations were also evaluated by combining these ambient sources with the smartphone's integrated LED flashlight (~5000–6500 K).²⁰ Alongside these varying illumination conditions, five camera angles (center, left, right, leftmost, and rightmost) were employed across 15 discrete pH values, resulting in 300 distinct image capture configurations. While the distance between the smartphone camera and the pH strip was kept constant to ensure scale consistency, collecting a diverse, uncalibrated dataset ensures that the proposed framework can computationally compensate for complex illumination variations

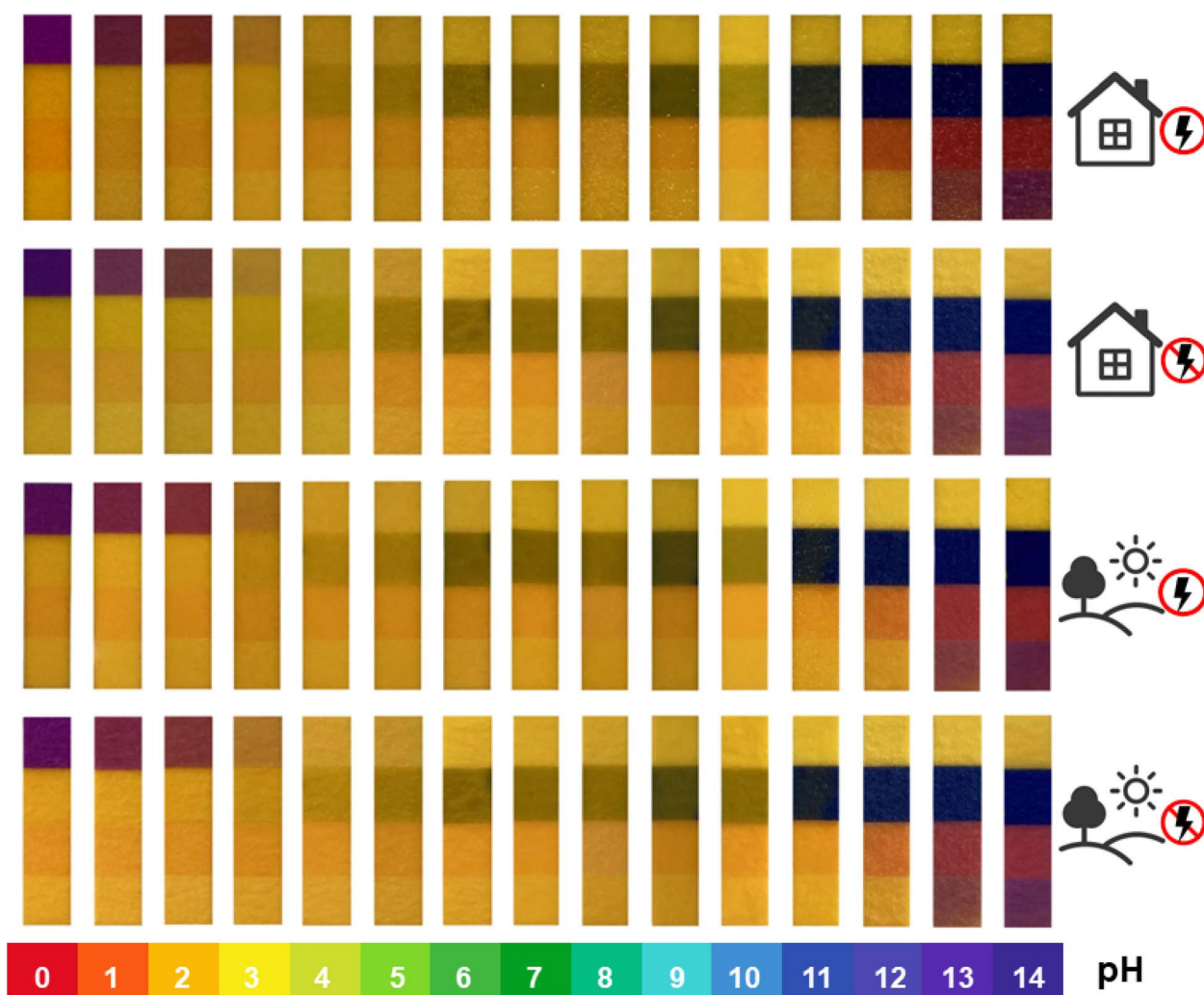


Fig. 2 Representative images of pH strips captured under four illumination conditions. Each row (top to bottom) corresponds to indoor with flashlight, indoor without flashlight, outdoor with flashlight, and outdoor without flashlight. Each column represents one of the 15 discrete pH values.



without requiring external calibration targets from the end-user. Additionally, to assess device-dependent variability, an additional independent image set corresponding to a full pH range (0–14; examples are presented in SI Fig. S1) was acquired using a second smartphone (Huawei Mate 10 Lite) under varying illumination conditions and camera angles.

Image processing and feature extraction

Each input image was pre-processed in MATLAB R2023b (MathWorks, Natick, USA) by cropping a region of interest (ROI) corresponding to the colored area of the pH strip, consisting of four indicator pads. The ROI was defined by manually selecting the corner coordinates using the `drawpolygon` function, enabling consistent isolation of the sensing region across all samples. The extracted coordinates were stored as text files (*.txt*) and converted to *.mat* format to generate binary masks for ROI extraction.

Fig. 2 illustrates the images of pH strips acquired under four illumination conditions across different pH levels. To quantitatively analyze the variations of color associated with pH, RGB values were converted into HSV and CIELAB color spaces. To ensure a comprehensive analysis, features were extracted from these multiple color spaces. Statistical parameters, including mean, skewness, and kurtosis, were included to characterize the distribution properties and asymmetry of color intensity values within the multi-pad sensing region. Additionally, texture- and intensity-based features (such as contrast, correlation, energy, entropy, homogeneity, and mean intensity) were incorporated to capture spatial heterogeneity and inter-pad transitions. A total of 33 complementary descriptors were chosen due to their proven physical relevance in image-based chemical sensing applications.⁴⁸ This initial comprehensive feature set served as a base for the subsequent feature importance analysis, which was utilized to identify the most critical descriptors and eliminate redundant variables.

Training and testing ML models

Various regression models, including linear regressors, decision trees, ensemble methods, support vector machines (SVM), kernel-based models, and neural networks, were trained and evaluated in Python 3.10 (Python Software Foundation, Delaware, USA) using the PyCharm IDE (JetBrains N.V., Amsterdam,

Netherlands). To improve the robustness and generalizability of model evaluation while reducing variance, a *k*-fold cross-validation strategy was applied during model development. Because preliminary screening indicated slow or redundant performance for some models, and certain models (e.g., Gaussian Process Regression (GPR) and its variants) have integration requirements that are not well aligned with lightweight mobile deployment, other available regression models were not included. In order to improve the performance of generalization and to reduce the likelihood of overfitting, model selection was guided by both predictive performance and fold-to-fold consistency. Table 1 lists the regression models evaluated in this study.

The evaluated regression models exhibit distinct characteristics in modeling colorimetric pH behavior. Linear models assume proportional relationships between extracted colorimetric features and pH values, providing interpretability and computational efficiency but limited expressiveness for modeling nonlinear color transitions across the full pH range. Decision tree-based models partition the feature space through rule-based splits, enabling nonlinear mapping; however, coarse variants may underfit complex calibration behavior. Ensemble methods, including bagged and boosted trees, improve predictive stability through model aggregation, contributing to variance reduction and enhanced robustness to illumination-induced variability. Kernel-based SVM models address nonlinear feature-target relationships by projecting features into higher-dimensional spaces, with performance influenced by regularization parameters and kernel selection. Neural networks are capable of capturing higher-order feature interactions inherent in colorimetric responses, offering flexible nonlinear approximations well-suited for representing pH-dependent color transitions. Model selection was guided not only by predictive performance but also by computational efficiency and suitability for lightweight mobile deployment.

The dataset was partitioned into a training set (80%) and a test set (20%). A 10-fold cross-validation procedure was applied to the training subset during model development and selection, while the independent test subset was used for final performance evaluation. The labeled data used for supervised regression consisted of 33 colorimetric features as input variables and their corresponding pH values as continuous target outputs.^{31,32} Feature importance analysis was subsequently

Table 1 Regression models with learned parameter counts and approximate storage

Category	Model	Number of parameters	Size (KB)
Linear models	Linear regression/efficient linear regression	34	~0.27
Trees	Coarse tree/medium tree/fine tree	~175/375/495	~1/~2/~2.7
SVM	Linear SVM	17 647	~138
	Quadratic SVM	9827	~76.8
	Cubic SVM	9011	~70.4
	Gaussian SVM	9759	~76.2
	Kernel (exp.) SVM	2048	~16
	Kernel least-squares	2048	~16
Ensemble methods	Bagged trees	13 380	~105
	Boosted trees	3150	~25
Neural networks	Narrow neural network (10)	351	~2.7
	Wide neural network (100)	3501	~27



conducted on the selected regression model using the SHAP method. The resulting SHAP values were used to estimate the contribution of individual features to the model output and to rank colorimetric features according to their relative importance. This ranking was further used for feature reduction by selecting the most influential features and retraining models with reduced feature sets to determine whether similar predictive performance could be maintained with fewer input features.

Android app design

An Android application, *pHScoper*, was developed to extract colorimetric information and predict pH values from commercial pH strip images using an ML regression model. The final selected regression model was developed in Python and subsequently converted into the TensorFlow Lite (*.tflite*) format using the *TensorFlow Lite Converter API*. To ensure minimal memory footprint (as detailed in SI Table S1), standard model optimization flags were applied during this conversion to enable lightweight, efficient on-device inference. The mobile application, *pHScoper*, was implemented in Kotlin using Android Studio (Fig. 3a). By embedding the TensorFlow Lite interpreter directly into the application architecture, inference is performed locally on the smartphone. This approach eliminates the requirement for cloud connectivity, thereby ensuring operational portability, computational efficiency, and user data privacy. The practical application workflow consists of four sequential stages: (i) image acquisition *via* the device's camera or gallery import (Fig. 3b); (ii) ROI extraction using an adjustable cropping interface to select the indicator pads (Fig. 3c); (iii) on-device feature extraction from the cropped region (implemented utilizing native Kotlin Bitmap processing to compute the colorimetric and statistical descriptors); and (iv) local ML-

based prediction. Upon tapping the Analyze button, the continuous pH prediction is immediately computed and displayed on the screen (Fig. 3d).

Evaluation metrics

The performance of the regression models was evaluated using the coefficient of determination (R^2) and error metrics such as mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE):

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (1)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3)$$

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (4)$$

where i represents the test samples ($i = 1, \dots, N$), N is the number of samples in the test set, \hat{y} represents the predicted value, and y denotes the actual value.

Results

A qualitative assessment of the acquired images revealed that the illumination source significantly impacts camera-based color perception. As illustrated in Fig. 2, distinct color variations were observed across identical pH strips depending on the

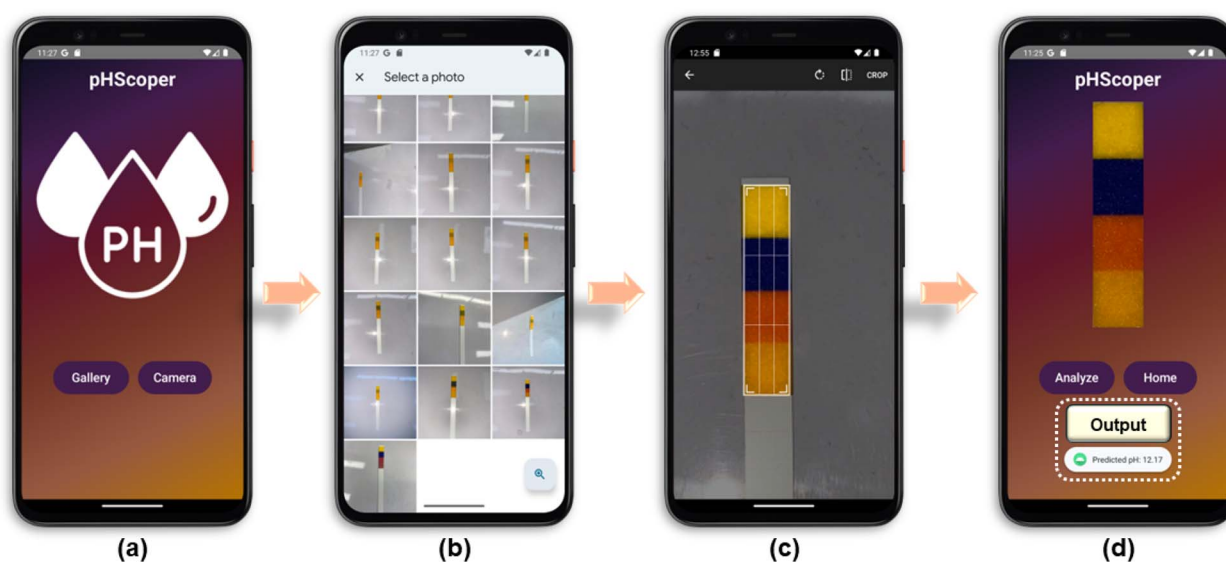


Fig. 3 Screenshots of the *pHScoper* mobile application. (a) Startup screen displaying the application logo and home interface, with options to capture an image by selecting one from the gallery or capture a picture from the camera. (b) Image gallery to upload an image of a pH strip. (c) A cropping screen is displayed where the user isolates the ROI from the background. (d) Result screen showing the assessed pH value after analysis, with an option to return to the home screen.



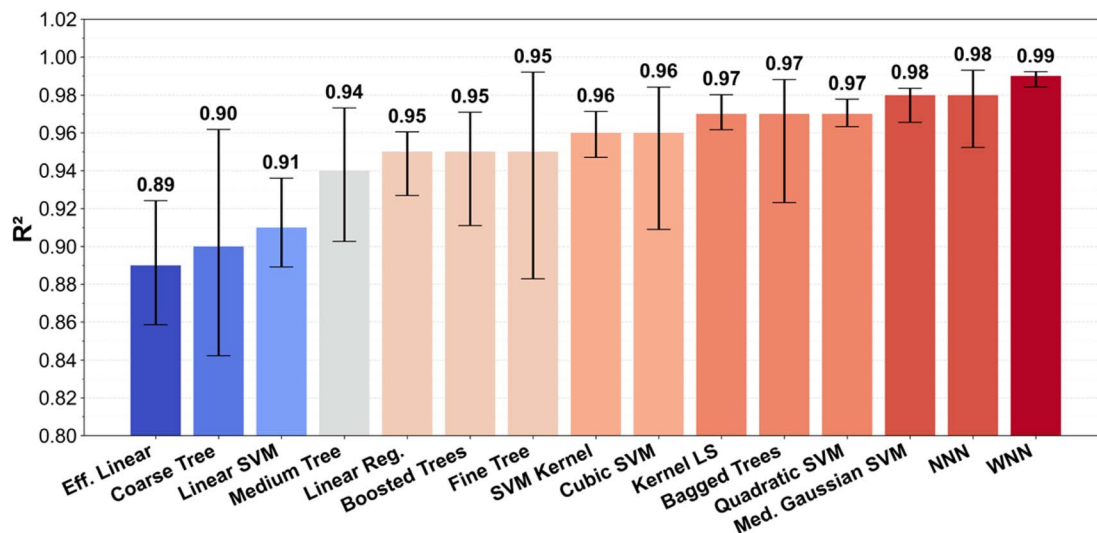


Fig. 4 Comparison of 15 regression models based on R^2 values with error bars.

lighting setup. Specifically, images captured under indoor lighting without a flashlight exhibited cooler tones compared to other setups. Conversely, under outdoor conditions with the flashlight enabled, the yellow pads (pH 0) displayed reduced luminosity compared to their counterparts captured without the flashlight. To mitigate these illumination-induced variations, we generated a comprehensive dataset featuring identical pH strips captured under diverse luminosity levels. The ROI-derived features were compiled into a structured dataset for ML model training in Python, comprising 33 colorimetric, texture-, and intensity-based input variables used for regression

analysis. Based on their R^2 values, a total of 15 regression models were evaluated.

The efficient linear regression and coarse tree models yielded the lowest R^2 values and were therefore excluded from this task. SVMs and ensemble-based methods demonstrated better performance. The wide neural network (WNN) achieved the highest predictive performance ($R^2 = 0.99$) and was consequently selected for further analysis. The variability of the R^2 values across cross-validations was represented by error bars, as shown in Fig. 4. The central bar corresponds to the mean R^2 value for each regression model, while the upper and lower

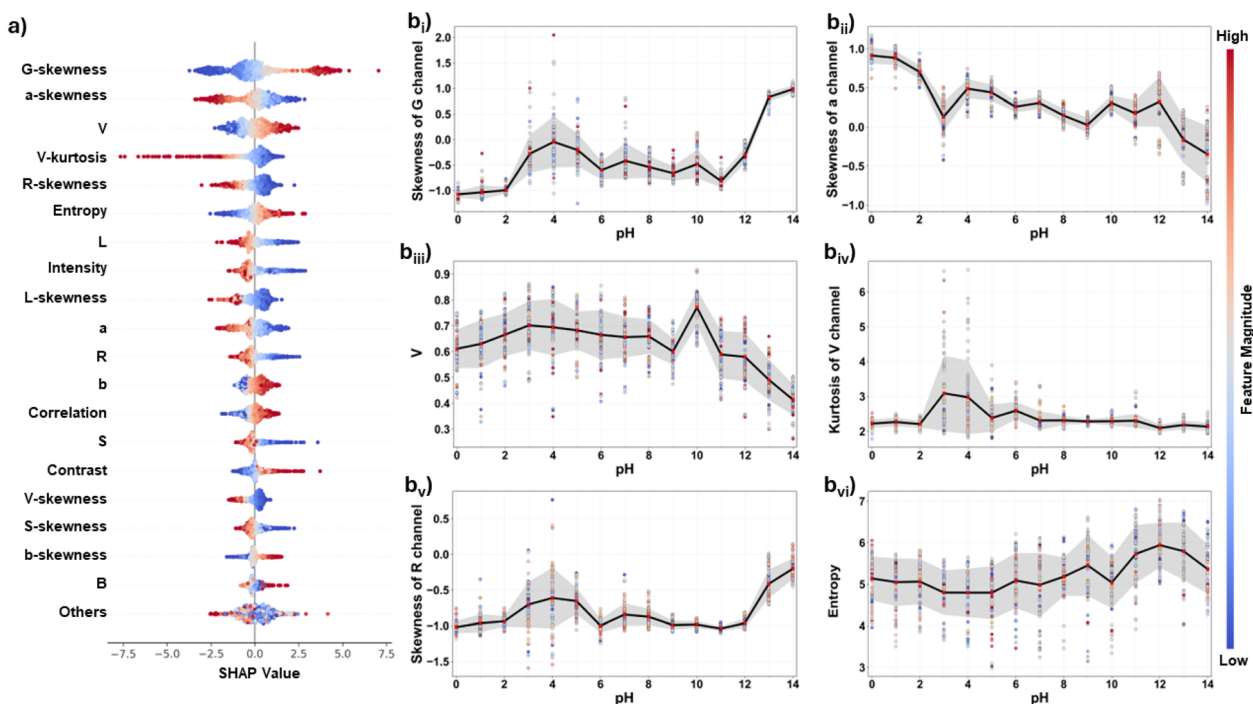


Fig. 5 (a) SHAP-based feature importance analysis of the WNN model and (b) pH-dependent variations of the top 6 features: (b_i) G-skewness, (b_{ii}) a-skewness, (b_{iii}) V, (b_{iv}) V-kurtosis, (b_v) R-skewness, and (b_{vi}) entropy.



limits of the error bars represent the maximum and minimum R^2 values obtained during 10-fold cross-validation. Based on the error bars, the highest consistency was achieved by the WNN model, whereas the largest variation was reflected by the coarse tree model. The relatively narrow error bars observed for the WNN model indicate stable predictive behavior across validation folds and support its robustness against overfitting.

A feature importance analysis was carried out to quantify the relative impact of each feature on the regression model's output. Moreover, this analysis can improve model interpretability and predictive performance.³³ Therefore, the best-performing model was interpreted using SHAP values in Python. Among the 33 extracted colorimetric features, the model identified G-skewness, a-skewness, V, V-kurtosis, R-skewness, and entropy as the most influential features for the colorimetric determination of pH, as shown in Fig. 5a. Based on the plots in Fig. 5(b_i-b_{vi}), the skewness of the green channel exhibited a pronounced peak under highly basic conditions, reflecting enhanced asymmetry in pixel intensity distributions at high pH values. A secondary peak was also observed around pH 4, suggesting increased asymmetry in the acidic range. The skewness of the red channel showed a similar pattern, indicating that variations in skewness within these channels contribute to the model's capacity to differentiate between pH levels. In contrast, the skewness of the a channel gradually decreased with increasing pH, indicating that lower values of this feature are characteristic of more basic conditions and

therefore contribute to higher predicted pH values, following an approximately linear trend. The V channel showed a local maximum around pH 3 and a pronounced global maximum near pH 10, after which it decreased toward very high pH values. This variation reflects changes in brightness in the colorimetric response and provides a discriminative feature for identifying alkaline conditions. The kurtosis of the V channel exhibited a distinct peak around pH 3 while remaining relatively stable across the rest of the pH range. Finally, entropy remained relatively stable across most pH levels, with a local minimum around pH 10 and a global maximum near pH 12, reflecting localized changes in pixel intensity distributions at strongly basic pH levels.

Relative importance (RI) metrics based on SHAP values were calculated to quantify the contribution of each feature to the model predictions and to identify the most informative descriptors governing the colorimetric response. This analysis identified six features as the most influential contributors to the model predictions. To evaluate the impact of feature reduction, feature subset experiments were conducted in which different combinations and numbers of input features were tested against the full 33-feature model (Fig. 6a), which was defined as the baseline with an RI of 100%. In contrast, models trained on reduced feature subsets exhibited noticeably lower R^2 values. For instance, the reduced feature subsets of the individual V-channel descriptors (Fig. 6b) and texture-only features (Fig. 6c) resulted in substantial performance degradation,

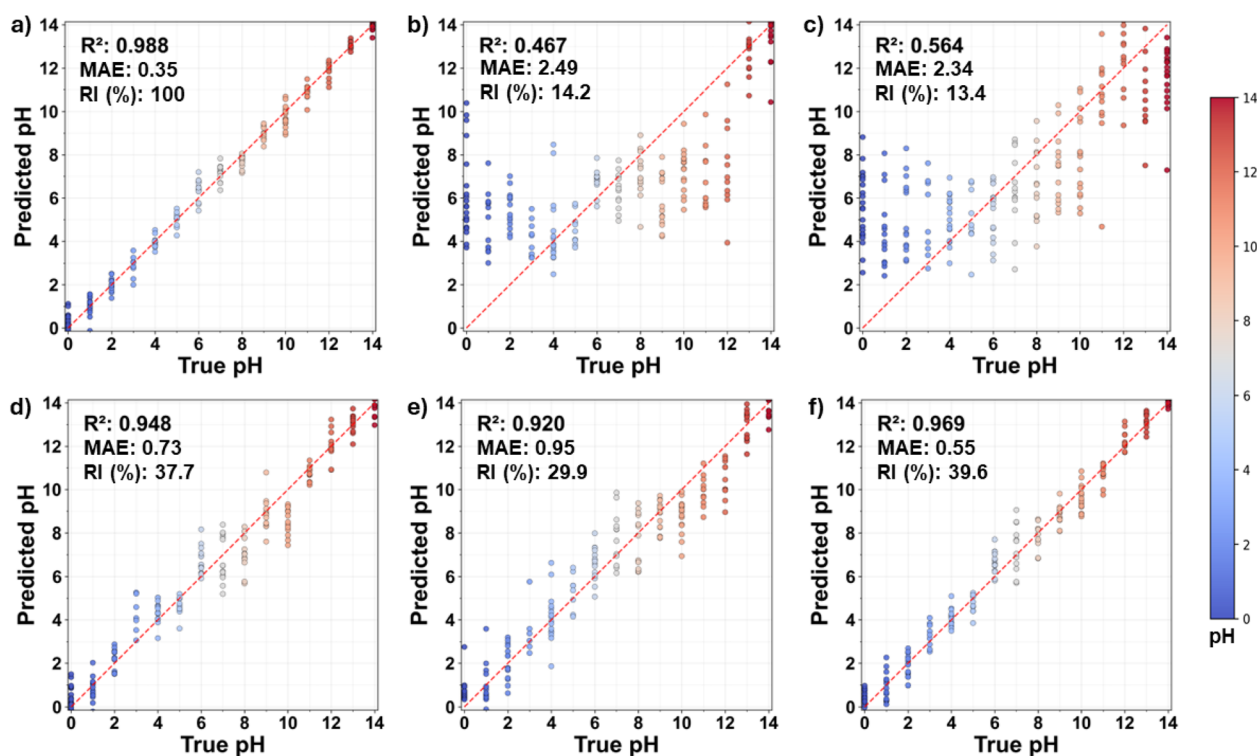


Fig. 6 Predicted pH versus true pH for the WNN model using different feature subsets. (a) The model using all 33 extracted features provides the highest predictive performance ($R^2 = 0.988$, MAE = 0.35, RI = 100%). (b) V-channel descriptors only and (c) texture-only features, showing substantial performance degradation ($R^2 < 0.60$, MAE > 2.3). (d) Model using only the skewness values of the color channels, demonstrating improved predictive accuracy ($R^2 = 0.948$, MAE = 0.73, RI = 37.7%). (e and f) Optimized subsets of the four and six most influential features with reduced feature sets ($R^2 = 0.920$, MAE = 0.95, RI = 29.9% and $R^2 = 0.969$, MAE = 0.55, RI = 39.6%, respectively).



corresponding to RI of only 14.2% and 13.4%, respectively. The decline in R^2 was most pronounced when limited channel feature subsets were used, as shown in Fig. 6b and c. Based on their RI values, texture-only features provided a higher R^2 and a lower MAE compared to the individual V-channel descriptors, despite having a lower RI. Models using optimized subsets of four (Fig. 6e) and six (Fig. 6f) influential features still achieved strong R^2 values of 0.920 and 0.969, respectively, corresponding to the RI of 29.9% and 39.6%, although slightly lower than that of the full feature configuration. Similarly, the model trained exclusively on skewness-based features (Fig. 6d) achieved an improved R^2 value of 0.948, with a RI of 37.7%. These results demonstrate that although RI values decreased when fewer features were used, the skewness-only, top four-feature, and top six-feature subsets retained most of the predictive information contained in the full 33-feature model. These results demonstrate that a reduced set of physically meaningful colorimetric features can maintain predictive performance while enhancing computational efficiency, supporting the implementation of lightweight, robust smartphone-based sensing.

As a final validation step, the cross-device generalization of the developed application was evaluated using the independent dataset acquired from the second smartphone. Because these images were strictly excluded from the model development phase, they served as an external test for cross-device generalization. The deployed model maintained highly accurate predictions ($R^2 = 0.97$, MAE = 0.48, RMSE = 0.59; the regression plot is presented in SI Fig. S2), confirming that *pHScoper* effectively tolerates variations in different smartphone camera sensors and internal image processing pipelines.

To evaluate the computational efficiency of the proposed lightweight deployment strategy, the original 33-feature model and the SHAP-optimized 6-feature model were quantitatively compared within the smartphone application. To ensure a fair comparison, both models were converted to TensorFlow Lite format and benchmarked under identical experimental conditions, including the same image resolution and preprocessing workflow. As summarized in SI Table S1, the SHAP-optimized model reduced the input features from 33 to 6 (an 81.8% reduction in dimensionality) and decreased the TensorFlow Lite model size from 16 KB to 6 KB (a 62.5% reduction). Despite the smaller model size, both exhibited comparable runtime memory usage (~0.18 MB), indicating that the fixed TensorFlow Lite runtime overhead and image preprocessing operations dominate the overall memory footprint. Notably, evaluated over 1000 repeated runs on the Android platform, the optimized 6-feature model achieved an average inference time of 0.06 ± 0.02 ms, firmly supporting its applicability for real-time mobile deployment.

Discussion

The camera's color perception is influenced by its color matching functions, illumination conditions, and the reflective properties of the pH strip, which may necessitate color calibration to enhance the accuracy of smartphone-based pH analysis. Although the use of the flashlight under indoor

conditions produced slightly more vibrant color responses, this effect was not consistently observed under outdoor conditions, highlighting the importance of incorporating illumination variability in the design of robust colorimetric pH sensing systems. *pHScoper* does not require custom hardware or cloud computation, ensuring accessibility in low-cost and resource-limited settings. Moreover, the application minimizes the need for repeated strip usage by enabling an efficient analysis in a single measurement, thereby reducing material consumption. Fig. 3 illustrates the workflow of the application, in which the selected image is cropped before further processing. Similar deployment strategies based on TensorFlow have been reported in recent smartphone-based colorimetric analysis studies, demonstrating their effectiveness for on-device regression scenarios.^{34,35}

The pH-dependent variations of the top six features (Fig. 5b) highlight a nonlinear relationship between the extracted image features and pH, resulting in feature-specific response patterns. This nonlinear calibration behavior underscores the limitations of conventional pH estimation methods, which may not be well-suited to capture the complex mappings between image-derived descriptors and chemical responses on the pH strip, thereby motivating the use of ML-based regression approaches.¹⁸ At very low and very high pH values, the slightly increased prediction variability is closely related to the nonlinear response behavior of the indicator dyes near the boundaries of their effective transition ranges. Because colorimetric pH indicator dyes operate through protonation–deprotonation equilibria, highly acidic or basic conditions cause the dyes to reach chemical saturation. These dominant color states exhibit reduced sensitivity to incremental pH changes, thereby limiting color discrimination near the pH extremes.⁴¹ This chemical limitation directly contributes to the comparatively higher variance observed near the pH boundaries in Fig. 6, despite the model's overall strong predictive performance. To enable continuous pH prediction, regression modeling combined with feature importance analysis was employed. Feature importance analysis further enhances interpretability by quantifying the contribution of each feature to the model output. In particular, SHAP-based analysis enables post-hoc interpretation of trained models, providing insight into the relative importance and directional influence of colorimetric features.⁴² The comparable predictive performance of the skewness-only, top four-feature, and top six-feature subsets relative to the full 33-feature model indicates that the skewness of the RGB, HSV, and CIE-LAB color channels, together with the V channel, its kurtosis, and entropy, play a dominant role in capturing pH-dependent color transitions on the pH strip (Fig. 6d–f). These subsets achieved RI values of 37.7%, 29.9%, and 39.6%, respectively, demonstrating that reasonable colorimetric pH prediction performance can be maintained with reduced feature dimensionality. These findings highlight the importance of feature selection and demonstrate the potential for further model simplification while preserving predictive performance.

Although individual indicator pads appear visually uniform, the analyzed ROI in this study encompassed four distinct pads, each formulated to exhibit different color responses across the



pH scale. Consequently, texture-based descriptors (e.g., contrast, correlation, entropy) are not merely capturing image noise or random local variations. Rather, they physically quantify the spatial color gradients and inter-pad transitions that dynamically change as each pad responds differently to a given pH level. While the SHAP analysis indicates that color distribution features are the dominant predictors (Fig. 6c), these texture metrics capture complementary spatial context regarding the multi-pad chemical reaction. Furthermore, capturing images under varying positions and illumination conditions introduced the necessary dataset variability to rigorously evaluate model robustness. This variability directly impacted algorithm performance across different data splits; for instance, the larger variation in R^2 observed for the coarse tree model during cross-validation (Fig. 4) reflects its inherent sensitivity to fold-specific feature distributions. Because coarse decision trees rely on a limited number of rigid, threshold-based splits, minor variations in image-derived features across data folds can significantly alter the learned decision boundaries, leading to inconsistent predictive performance.⁴³ In contrast, the WNN model exhibited greater stability and lower variance. This improved tolerance to heterogeneous acquisition conditions stems from the network's capacity to effectively model complex, nonlinear interactions among colorimetric features, ensuring robust performance even under varying illumination.⁴⁴

The findings of this study are also consistent with recent research on AI-based colorimetric pH sensing using smartphone cameras or other portable sensing systems. More recently, paper-based diagnostic platforms have increasingly incorporated computational image analysis and advanced sensing strategies to improve assay sensitivity, robustness, and quantitative interpretation. For instance, Du *et al.* introduced a rapid deep learning-based quantitative lateral flow assay integrating residual neural networks and temporal modeling to enable reliable quantification within the early stages of assay development, substantially reducing interpretation time while maintaining high quantitative accuracy.⁴⁵ Similarly, Park *et al.* developed a smartphone-assisted deep learning framework combined with a bioengineered enrichment strategy to improve the sensitivity of lateral flow assays and enhance the interpretation of weak colorimetric responses in noninvasive HIV screening.⁴⁶ Lee *et al.* reported a deep learning-assisted point-of-care diagnostic framework that combines image-based region-of-interest detection with sequential analysis for rapid assay interpretation, demonstrating the potential of computational support for improving analytical performance in paper-based tests.⁴⁷ Similarly, Han *et al.* introduced a deep learning-enhanced vertical flow paper-based assay incorporating nanoparticle amplification and computational analysis to achieve highly sensitive quantitative detection.⁴⁸ Although these studies target different analytes and sensing formats, they demonstrate the growing transition of paper-based assays from subjective interpretation toward more robust quantitative systems. In this context, the proposed framework contributes to these ongoing efforts by integrating quantitative image analysis with ML-assisted pH prediction. While many of these existing studies

rely on classification approaches or directly utilize raw RGB measurements, the present study specifically adopts a regression-based approach to enable more precise, continuous pH estimation. Unlike classification methods that merely assign discrete pH labels, our regression framework successfully captures finer intermediate variations in the colorimetric response. In addition, this study incorporates engineered colorimetric descriptors from multiple color spaces and evaluates their relative importance using SHAP-based feature analysis. Table 2 summarizes selected representative studies.

This study presents several advantages beyond predictive performance. First, it covers the complete pH range from 0 to 14, whereas prior studies have reported improved precision when estimation is restricted to narrower ranges. For instance, Kadian *et al.*³⁸ achieved high precision within a pH range of 2–12, while Alhaqi *et al.*⁴⁰ reported balanced $F1$ -scores associated with relatively moderate precision values within the pH range of 4–8. Extending such approaches to broader pH ranges may require additional methodological adjustments. A second advantage lies in the composition and diversity of the dataset. While Elsenety *et al.*³⁷ reported high prediction accuracy using a larger dataset acquired under controlled illumination, the present work intentionally incorporates challenging acquisition settings, such as camera angle and natural illumination, to better reflect real-world scenarios. Third, nonlinear colorimetric behavior is explicitly evaluated through comprehensive feature importance analysis. While ML has been widely applied in related studies,^{19,37,38,40} explicit feature importance analyses have not been systematically reported, with model behavior primarily addressed in an end-to-end manner. To build upon earlier work, our study integrates feature importance analysis to interpret nonlinear and feature-specific responses of individual colorimetric features across the continuous pH range, with model inference performed locally on the mobile device. This study reduces reliance on specialized hardware by using a standard smartphone camera and commercially available pH strips, thereby lowering system cost and complexity. In contrast, methods based on fixed imaging systems³⁶ or custom hardware^{38,39} may improve measurement consistency but often rely on additional hardware configurations that limit portability and scalability. The present approach demonstrates reliable pH estimation under practical conditions, supporting field-deployable and portable applications.

Despite the high predictive performance and successful on-device integration, several limitations should be acknowledged. The colorimetric features were derived exclusively from MQuant® pH strips. Because indicator compositions and resulting colorimetric responses vary across commercial manufacturers, the generalizability of the proposed model to other strip brands remains an open question. To address this, future iterations will implement transfer learning strategies. Under this approach, the current architecture can serve as a pre-trained base model. By fine-tuning this model using smaller, brand-specific supplementary datasets acquired under similar conditions, the system could efficiently adapt to different indicators without the computational and practical burden of complete model redevelopment. Although the proposed



Table 2 Comparison of colorimetric pH sensing studies^a

Reference	Sensor	Method	Dataset	Features	Performance
19	Smartphone + pH strips	LS-SVM	N/A	Raw RGB values	Classification accuracy = 100%
36	Smartphone + pH strips	None (color adaptation)	N/A	CIE 1976 u'v' color space	Error \approx 0.25 pH units
37	Smartphone + pH strips	KNN	2689 experimental samples	RGB color features	$R^2 \approx$ 0.99
38	Microneedle patch + web	CNN (VGG16)	1466 images	RGB image data	ACC = 0.98
39	AI-WMCS	1D-CNN-GRU	Real and artificial tear samples	Multi-channel color patches	$R^2 \approx$ 0.99
40	Smartphone camera	Decision tree	1025 data points (pH 4–8)	RGB, hue, saturation, luminance, grayscale, coloration	ACC = 0.67
This study	Smartphone + pH strips	WNN	1035 images	RGB, HSV, CIELAB features	$R^2 \approx$ 0.99

^a LS-SVM = Least-squares SVM; KNN = K-nearest neighbors; CNN = convolutional neural network; VGG = visual geometry group; AI-WMCS = AI-assisted wearable microfluidic colorimetric sensor; GRU = gated recurrent unit; DT = decision tree; WNN = wide neural network.

framework demonstrates robustness to moderate camera-angle variations included in the training data, explicit perspective-normalization algorithms were not incorporated into the current preprocessing pipeline. Because the ROI extraction relies on controlled orientation, severe angular distortions or completely unrestricted strip placements could still negatively impact predictive performance, representing a limitation of the current study. To further enhance operational reliability in unconstrained real-world environments, future iterations of the mobile application will integrate algorithmic geometric correction and automated ROI detection—defining the reaction zone dynamically rather than relying on predefined geometric shapes. While the current study focuses on colorimetric pH estimation, the implemented framework is adaptable to other biochemical assays, such as ammonia, lactate, or uric acid. By adapting the ROI templates and retraining the model with analyte-specific datasets, the current pH system serves as a base model for cost-effective and extensible quantitative analysis. To achieve this, future datasets for new colorimetric sensors must similarly incorporate variability in light conditions and camera angles. Furthermore, since the reported results were validated on two smartphone platforms, expanding validation across a wider range of devices, camera systems, and alternative light sources (e.g., laser illumination) remains an important future objective. Finally, further optimization of the TensorFlow Lite integration will ensure this generalized framework remains efficient and robust for diverse paper-based analytical deployments.

Conclusion

In this study, a smartphone-based quantitative colorimetric analysis of pH strips was successfully developed through the integration of ML. By training regression models on images captured under diverse, uncalibrated illumination conditions and multiple camera angles, the framework demonstrated high robustness against real-world imaging variability. A key finding of this work is the implementation of SHAP-based feature importance analysis, which successfully reduced the feature

space dimensionality by 81.8% (from 33 to 6 features) while maintaining strong predictive performance ($R^2 = 0.99$). The optimized model was efficiently embedded into an Android application, *pHScoper*, using TensorFlow Lite. This user-friendly architecture allows for offline, localized ROI feature extraction and continuous pH prediction without cloud dependency. While future work must address current limitations regarding automated geometric correction and cross-brand strip generalizability, the current *pHScoper* framework demonstrates strong potential for straightforward adaptation to point-of-care diagnostics. Its scalable design, operational simplicity, and efficient local data processing make it highly suitable for reliable chemical analysis in remote and resource-limited environments.

Author contributions

Ece Yıldız: methodology, data curation, formal analysis, investigation, software, visualization, writing – original draft. Mustafa Şen: conceptualization, methodology, visualization, supervision, validation, project administration, writing – review & editing. Mehmet Akif Özdemir: conceptualization, methodology, formal analysis, visualization, resources, supervision, validation, writing – review & editing. All authors provided input into drafts and approved the final draft of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d6ay00780e>.



Acknowledgements

Open access funding for this article was provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). The authors gratefully acknowledge Ece Minel Bursalı for her valuable contributions during the revision process of this manuscript.

References

- 1 S.-H. Kuo, C.-J. Shen, C.-F. Shen and C.-M. Cheng, *Diagnostics*, 2020, **10**, 107.
- 2 J. Kim, J. Park, P. Kim, C. Lee, K. Choi and K. Choi, *Ecotoxicology*, 2009, **19**, 662–669.
- 3 S. Yoshida, T. Miyake, S. Yamamoto, S. Furukawa, T. Niiya, H. Senba, S. Kanzaki, O. Yoshida, T. Ishihara, M. Koizumi, M. Hirooka, T. Kumagi, M. Abe, K. Kitai, B. Matsuura and Y. Hiasa, *J. Diabetes Investig.*, 2017, **9**, 769–775.
- 4 L. Zhang, Q. Zhao, Z. Jiang, J. Shen, W. Wu, X. Liu, Q. Fan and W. Huang, *Biosensors*, 2021, **11**, 282.
- 5 M. González-González, R. F. Toba, M. Ortiz-Martínez and M. Rito-Palomares, *J. Chem. Technol. Biotechnol.*, 2025.
- 6 M. S. Fathimal and S. Jothiraj, *Lecture Notes in Networks and Systems*, Springer, 2022, pp. 101–107.
- 7 S. Wang, W. Chen, A. Du, Y. Kou, X. Xu, D. Hu and Z. Lu, *Adv. Compos. Hybrid Mater.*, 2025, **8**, 260.
- 8 J. Hu, S. Wang, L. Wang, F. Li, B. Pingguan-Murphy, T. J. Lu and F. Xu, *Biosens. Bioelectron.*, 2013, **54**, 585–597.
- 9 Y. Yoo and W. S. Yoo, *Sensors*, 2020, **20**, 6418.
- 10 G. M. Fernandes, W. R. Silva, D. N. Barreto, R. S. Lamarca, P. C. F. L. Gomes, J. F. Da S Petrucı and A. D. Batista, *Anal. Chim. Acta*, 2020, **1135**, 187–203.
- 11 W. Liu, J. Tian, C. Mao, Z. Wang, J. Liu, R. A. Dahlgren, L. Zhang and X. Wang, *Anal. Chim. Acta*, 2020, **1127**, 246–255.
- 12 M. Zhang, Y. Zhang, C. Yang, C. Ma and J. Tang, *Talanta*, 2020, **224**, 121840.
- 13 C. Choi, S. M. Shaban, B. Moon, D. Pyun and D. Kim, *Anal. Chim. Acta*, 2021, **1170**, 338630.
- 14 G. D. Ozdemir, M. A. Ozdemir, M. Sen and U. K. Ercan, *Adv. Intell. Syst.*, 2024, **6**, 2400029.
- 15 M. Maruthupandi and N. Y. Lee, *Sensors*, 2026, **26**, 439.
- 16 N. Monisha, K. Shrivasa, T. Kant, S. Patel, R. Devi, N. S. Dahariya, S. Pervez, M. K. Deb, M. K. Rai and J. Rai, *J. Hazard. Mater.*, 2021, **414**, 125440.
- 17 K. Su, Q. Zou, J. Zhou, L. Zou, H. Li, T. Wang, N. Hu and P. Wang, *Sens. Actuators, B*, 2015, **216**, 134–140.
- 18 E. M. Bursalı, M. A. Özdemir and M. Şen, *Microchim. Acta*, 2025, **192**, 850.
- 19 A. Y. Mutlu, V. Kılıç, G. K. Özdemir, A. Bayram, N. Horzum and M. E. Solmaz, *Analyst*, 2017, **142**, 2434–2441.
- 20 M. E. Solmaz, A. Y. Mutlu, G. Alankus, V. Kılıç, A. Bayram and N. Horzum, *Sens. Actuators, B*, 2017, **255**, 1967–1973.
- 21 Y. Xiao, Y. Huang, J. Qiu, H. Cai and H. Ni, *Chem. Pap.*, 2024, **78**, 8849–8862.
- 22 Ö. B. Mercan, V. Kılıç and M. Şen, *Sens. Actuators, B*, 2020, **329**, 129037.
- 23 F. Feng, Z. Ou, F. Zhang, J. Chen, J. Huang, J. Wang, H. Zuo and J. Zeng, *Nano Res.*, 2023, **16**, 12084–12091.
- 24 W. Liu, S. Liu, K. Fan, Z. Li, Z. Guo, D. Cheng and G. Liu, *IEEE Sens. J.*, 2024, **24**, 32991–33000.
- 25 H. Tan, *2022 International Joint Conference on Neural Networks (IJCNN)*, 2023, pp. 1–8.
- 26 M. A. Özdemir, G. D. Özdemir, M. Gül, O. Güren and U. K. Ercan, *Mach. Learn.: Sci. Technol.*, 2023, **4**, 015030.
- 27 U. K. Ercan, G. D. Özdemir, M. A. Özdemir and O. Güren, *Plasma Processes Polym.*, 2023, **20**, e2300066.
- 28 S. Gupte and J. Paparrizos, *Proc. ACM Manag. Data.*, 2025, **3**, 1–31.
- 29 A. Parakh, A. Awate, S. M. Barman, R. K. Kadu, D. P. Tulaskar, M. B. Kulkarni and M. Bhaiyya, *Trends Environ. Anal. Chem.*, 2025, **48**, e00280.
- 30 Y. Li, Y. Wang, S. Chen, Z. Wang and L. Feng, *Anal. Chim. Acta*, 2021, **1154**, 338275.
- 31 M. A. Özdemir, G. D. Özdemir, M. Gül, O. Güren and U. K. Ercan, *Mach. Learn.: Sci. Technol.*, 2023, **4**, 015030.
- 32 M. W. Berry, A. Mohamed and B. W. Yap, *Unsupervised and Semi-supervised Learning*, Springer, 2019.
- 33 M. Mirzaei, I. Furxhi, F. Murphy and M. Mullins, *Nanomaterials*, 2021, **11**, 1774.
- 34 Z. Meng, M. Tayyab, Z. Lin, H. Raji and M. Javanmard, *Analyst*, 2024, **149**, 1719–1726.
- 35 M. Baştürk, E. Yüzer, M. Şen and V. Kılıç, *Adv. Intell. Syst.*, 2024, **6**, 2400202.
- 36 S. D. Kim, Y. Koo and Y. Yun, *Sensors*, 2017, **17**, 1604.
- 37 M. M. Elsenety, M. B. I. Mohamed, M. E. Sultan and B. A. Elsayed, *Sci. Rep.*, 2022, **12**, 22584.
- 38 S. Kadian, P. Kumari, S. S. Sahoo, S. Shukla and R. J. Narayan, *Microchem. J.*, 2024, **200**, 110350.
- 39 Z. Wang, Y. Dong, X. Sui, X. Shao, K. Li, H. Zhang, Z. Xu and D. Zhang, *npj Flexible Electron.*, 2024, **8**, 35.
- 40 M. A. D. Alhaqi, *J. Inf. Syst. Eng. Manag.*, 2025, **10**, 324–333.
- 41 A. Steinegger, O. Wolfbeis and S. Borisov, *Chem. Rev.*, 2020, **120**, 12357–12489.
- 42 S. M. Lundberg and S.-I. Lee, *Adv. Neural Inf. Process. Syst.*, 2017, **30**, 4766.
- 43 L. Breiman, *Mach. Learn.*, 1996, **24**, 123–140.
- 44 D. Hendrycks, S. Basart, N. Mu, S. Kadavath, F. Wang, E. Dorundo, R. Desai, T. L. Zhu, S. Parajuli, M. Guo, D. Song, J. Steinhardt and J. Gilmer, *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2020, pp. 8320–8329.
- 45 J. Du, C. Cao, Z. Xue, W. Wang, X. Lu, Y. Wei, J. Huang, L. Zhao, L. Wang, F. Xu, C. Yao, T. Wen and M. You, *Anal. Chim. Acta*, 2025, **97**, 24196–24208.
- 46 J. S. Park, S. Lee, H. Woo, J. H. Hong, D. S. Yoon, S. Chung and J. H. Lee, *ACS Nano*, 2026, **20**, 12639–12650.
- 47 S. Lee, J. S. Park, H. Woo, Y. K. Yoo, D. Lee, S. Chung, D. S. Yoon, K.-B. Lee and J. H. Lee, *Nat. Commun.*, 2024, **15**, 1695.
- 48 G.-R. Han, A. Goncharov, M. Eryilmaz, H.-A. Joung, R. Ghosh, G. Yim, N. Chang, M. Kim, K. Ngo, M. Veszpremi, K. Liao, O. B. Garner, D. Di Carlo and A. Ozcan, *ACS Nano*, 2024, **18**, 27933–27948.

