A Machine Learning Approach to pH Monitoring: Mango Leaf Colorimetry in Aquaculture

Hajar Rastegari[1], Romi Fadilah Rahmat[2], Farhad Nadi[3]
Higher Institution Center of Excellence (HICoE)-Institute of Tropical Aquaculture and Fisheries,
Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia[1]
Department of Information Technology-Faculty of Computer Science and Information Technology,
University of Sumatera Utara Medan, North Sumatra, Indonesia[2]
School of Information Technology, UNITAR International University, 47301 Kelana Jaya, Selangor, Malaysia[3]
Center for Innovation and Technology Adoption (CITA), UNITAR International University,
47301 Kelana Jaya, Selangor, Malaysia[3]

***Abstract*—Maintaining optimal water quality is crucial for successful aquaculture. This necessitates careful management of various water quality parameters, including pH levels within their ideal range. There is growing interest in creating affordable optical pH sensors that provide accurate readings across a wide range of pH values. Development of sensors that are both accurate and cost-effective remains a challenge. To this end, this study demonstrates the use of machine learning with mango leaf extract as a colorimetric indicator to achieve accurate and cost-effective pH estimation for aquaculture practices. Mango leaf was utilized as the pH indicator, covering a range from 1 to 13. RGB color extraction and Exif data were used for image analysis to extract relevant features. The XGBoost algorithm, optimized through stepwise hyperparameter tuning with early stopping, was used to train three different models on this dataset to predict pH values. Three classification models, namely Y3, Y5, and Y13, were trained with 3, 5, and 13 output classes, respectively. The overall precision achieved by each model was 0.94, 0.85, and 0.72, respectively. This demonstrates the potential of this approach for developing a user-friendly yet cost-effective sensor for pH detection applicable in aquaculture practices. The proposed method could help aquaculture farmers an affordable and intelligent smartphone-based pH detection tool, enhancing water quality management while reducing the need for expensive instruments and eliminating the need for additional costly and time-consuming experimental work, thereby contributing to the sustainability of aquaculture practices.**

***Keywords*—Aquaculture; machine learning; XGboost; water quality; sustainable aquaculture practices; water quality monitoring; mango leaf extract**

I. Introduction

Water quality is a pivotal factor in aquaculture, exerting significant influence on fish growth, survival, and reproduction [1], [2], [3]. Precise and consistent monitoring of water quality parameters is essential for effective aquaculture management, aiding in disease prevention, treatment, and overall productivity [2]. To maintain optimal water conditions, various physical, chemical, and biological treatments are applied to aquaculture pond [3]. Key water quality parameters, including temperature, dissolved oxygen, salinity, and pH, provide valuable insights into aquaculture system health and performance [4], [5]. The primary challenge with traditional water quality assessment methods is their lack of cost-effectiveness and the significant labour they require. This highlights the need for more innovative monitoring solutions. Although technological advancements like the Internet of Things (IoT) and artificial intelligence have enabled real-time monitoring and analysis of water quality [6], the high cost of sensors remains a major barrier to the widespread adoption of these modern technologies [7].

Several studies have highlighted that pH is a crucial parameter for assessing water quality and plays a fundamental role in aquatic ecosystems. It has been reported that abnormal pH levels can negatively impact the health of aquatic organisms and the overall quality of water [8], [9]. Therefore, maintaining optimal pH levels is essential for successful aquaculture practices. However, research consistently shows that the ideal pH range varies among fish species, as detailed in Table I, which presents the acceptable pH ranges for different aquatic species.

Colorimetric methods are simpler, more cost-effective, and capable of providing realtime results, making them a promising solution for pH measurement in various fields, such as environmental monitoring [10], [11], [12], [13]. In aquaculture, colorimetric pH sensors demonstrated continuous, in-situ monitor-ing, which allows for early detection of pH fluctuations and enables prompt corrective actions [14]. This study aims to advance the development of more efficient and effective aquaculture practices by addressing the limitations of traditional water quality monitoring techniques and leveraging the benefits of colorimetric pH sensors.

Supervised learning is a subfield of machine learning algorithms that trains models, f, on la-belled data. This data comprises feature vectors, x, and their corresponding labels, y. The primary objective is to minimize a loss function, denoted by $L(f)$, which quantifies the discrepancy between the predicted labels, $f(x)$, and the true labels, y [23].

Ensemble learning is a subcategory of supervised learning algorithms that aims to construct a robust learner, F, by combining multiple weaker learners, fi. A prominent example of ensemble learning is XGBoost, which utilizes decision trees as base learners. These decision trees recursively partition the feature space based on specific decision rules, ultimately predicting a class or a continuous value [24]. XGBoost operates in

TABLE I. OPTIMAL RANGE OF PH VALUES AS REPORTED IN LITERATURE

Description	Lower Range	Upper Range	Reference
Optimal pH range for nile tilapia is between 5 and 8.	5	8	[15]
Efficient nitrification activity in aquaculture biofilters 7.0 9.0			
Reproduction and infectivity of c. Irritans.	6	9	[16]
C. Irritans can survive in pH 5-10 in aquaculture.	5	10	[16]
PH ranged from 7.4 to 9.6 in eutrophic aquaculture ponds.	7.4	9.6	[8]
pH range 6-8 commonly acceptable in aquaculture.	6	8	[17]
pH ranges vary based on species, but generally 6.5-9.0.	6.5	9.0	[18]
Optimal pH for most aquaculture species is 7.0-8.5.	7.0	8.5	[18]
pH levels between 7.51 and 8.00 maintained in aquaculture.	7.51	8.00	[19]
Atlantic salmon embryos have lower lethal limits around pH 3.0-4.0.	3.0	4.0	[20]
Silver catfish juveniles survive in pH range of 4.0-9.0.	4.0	9.0	[21]
pH range in aquaculture: 6.5-9.0 suggested, optimal range varies by species.	6.5	9.0	[22]
Results suggest that 8.0–8.5 is the best pH range for survival and growth of the larvae of the silver catfish (rhamdia quelen) larvae	8.0	8.5	[22]

an iterative manner, gradually building an ensemble of decision trees. Each subsequent tree, f_i, focuses on rectifying the errors made by its predecessors using a technique called gradient boosting. Mathematically, this translates to minimizing $L(f)$ by strategically adding trees that target the residuals of the existing ensemble, $F_{i-1}(x)$. This approach enables XGBoost to effectively handle intricate feature relationships and achieve significant minimization of the loss function, $L(f)$.

II. BACKGROUND

Traditionally, pH measurement has relied on physical sensors, both analog and digital [25]. Despite their precision, these sensors necessitate frequent calibration to maintain accuracy, which can be expensive and time-consuming. Coupled with manual operation, this has prompted a search for more affordable and user-friendly alternatives. Recent advancements have facilitated the integration of digital pH sensors into IoT systems for continuous water quality monitoring [26], [25], [27]. While these solutions are suitable for larger aquaculture operations, their high implementation costs often preclude their adoption by smaller farms seeking economical options. An alternative approach to pH measurement involves pH test strips, a traditional and cost-effective method for assessing water quality in aquaculture [28], [29]. While less precise than sensor-based methods, their simplicity, affordability, and portability make them valuable tools in various applications. However, relying on visual color comparisons with a reference chart can introduce subjectivity and potential inaccuracies. Furthermore, the presence of other water constituents may interfere with test strip accuracy [30], [31]. Despite these limitations, the ease of use and absence of calibration requirements make pH test strips an economical option, particularly for small-scale farms with limited resources.

To mitigate the subjectivity inherent in interpreting pH test strip colors, researchers have explored the use of machine learning models to enhance the accuracy of pH measurements derived from paper strips [32], [33], [34]. Concurrently, the development of novel reagents for pH determination remains an active research area. These innovative reagents, when integrated with smartphone or machine learning-based methodologies, hold the potential to significantly improve pH measurement accuracy. A novel method for directly measuring the pH of airborne particles or droplets has been developed, combining pH indicator paper with RGB-based colorimetric analysis. This approach established a linear correlation between RGB values and pH, surpassing the accuracy and applicability of previous models. Hydrion® Brilliant pH dip stciks (lot no. 3110, Sigma-Aldrich), with their wide pH detection range and resistance to interference, were deemed optimal for analyzing ambient aerosols. Initial findings suggest that aerosol pH can be estimated with an uncertainty of 0.5 units or less, casting doubt on the reliability of traditional pH color charts and emphasizing the need for in situ calibration of pH papers using standardized pH buffers [11].

A smartphone-based colorimetric analysis technique employing a pH-sensitive photonic gel was reported. The gel exhibited color variations corresponding to different pH levels, which were captured and analyzed using a smartphone camera and image processing algorithms. This method demonstrated accurate real-time pH measurement, suggesting its potential applications in environmental monitoring and medical diagnostics [35].

A smartphone-based colorimetric method was developed for detecting enzyme-substrate reactions using pH-responsive gold nanoparticle assemblies. The pH-induced color changes in the gold nanoparticles, resulting from enzyme-substrate interactions, were captured and analyzed using a smartphone. This method offers a simple, cost-effective, and portable platform for monitoring enzyme-substrate reactions in various fields such as biochemistry and environmental science [36].

Traditionally, pH measurement in aquaculture has relied on physical sensors and pH test strips, which provide accurate results but often suffer from portability and usability constraints. In contrast, smartphone-based approaches, utilizing colorimetric changes and sophisticated image processing, offer high precision, portability, and user-friendliness, making them suitable for remote and real-time monitoring. However, these methods may necessitate specific reagents and con-trolled conditions for optimal performance. This research underscores the potential for developing cost-effective and user-friendly smartphone-based pH detection systems with broader applications.

III. MATERIALS AND METHODS

A. Chemicals

Ortho-Phosphoric acid (85-88%) and sodium hydroxide (99%) were purchased from R & M Chemicals. All chemicals were of analytical grade and used as received without further purification. Distilled water was used in all experiments. Green

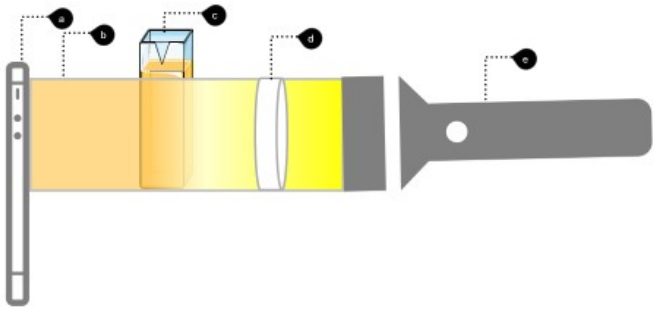


Fig. 1. Schematic illustration of the smartphone colorimetric sensors data gathering. (a) smart phone, (b) PVC pipe of, (c) Cuvette containing sample solution, (d) light diffuser that is a one-centimeter-thick hot glue, (e) torch light as the source of light.

mango leaves were collected from apple mango trees in Kuala Nerus, Terengganu, in November and December 2023.

B. Sample Preparation for Colorimetric Studies

Mango leaves powder was prepared by thoroughly washing the leaves with tap water, followed by three additional rinses with distilled water. The washed leaves were then dried in an air oven at 65°C for 48 h. After drying, the leaves were ground and sieved to obtain a powder with a particle size of 120 µm. To prepare standard solutions covering a pH range from 1 to 13, a sequential dilution process was conducted. This process involved diluting stock solutions of 1M H3PO4 and 1M NaOH with distilled water. The dilution ratios were exactly adjusted to achieve the desired pH levels for each standard solution. For colorimetric studies, 40 mL of each standard solution was mixed with 0.1 g of mango leaves powder. The mixture was manually shaken for 1 min at room temperature and then centrifuged at 8000 rpm for 15 min. The supernatants were removed and filtered using a Nylon syringe filter with a pore size of 0.22 µm. The filtered samples were then transferred to cuvettes for colorimetric analysis.

C. Dataset

Seven different smartphones were used to capture three sets of photographs for each sample with pH values ranging from 1 to 13. A simple photography setup, as illustrated in Fig. 1, was employed for image acquisition. The apparatus consisted of a commonly available 1-inch PVC pipe (Fig. 1(b)) designed to produce consistent images. A $22cm$ long polyethylene pipe with a $4.5cm$ inner diameter was used for the setup. A white LED flashlight (Eveready LC1L2A) was placed at one end of the pipe to illuminate the samples (Fig. 1(e)). To position the samples, a hole was created $7cm$ from the light source. A $0.5cm$ thick layer of translucent hot melt glue was applied inside the pipe to evenly distribute the light (Fig. 1(d)). The smartphone camera was positioned at the opposite end of the pipe to capture images of the samples (Fig. 1(a)). All photographs were taken of samples placed in standard spectrophotometer cuvettes with dimensions of $4.5cm$ height, $3.5mL$ capacity, and an optical range of $190-2500nm$ (Fig. 1(c)).

Image files were renamed to accurately reflect sample type details and subsequently transferred to a computer system.

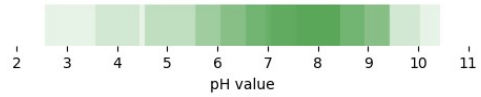


Fig. 2. Heat map chart of suitable pH ranges in aquaculture, as reported in the literature.

iPhone images were converted to JPG format post-transfer. Image boundaries were then delineated using the VIA annotation tool [37].

A dataset of 819 images was initially collected, reduced to 817 images after excluding two due to incorrect labeling. A Python script was developed to extract dominant colors within annotated regions of interest from each image. This process involved applying K-Means clustering with K=1 to determine the predominant color within the annotated regions of the image.

In addition to the dominant color's RGB value, Exif metadata was extracted from each image and integrated into the dataset. Feature engineering was applied to create additional features, such as average image intensity (mean of RGB values) and grayscale values calculated using,

$$0.299R + 0.587G + 0.114B,$$

The red-to-green ratio was calculated and added as an additional feature to the dataset. Subsequently, RGB values were converted to XYZ, HSL, and LAB color spaces using the Python colormath library.

D. pH Class Construction

The original dataset contained a discrete target variable, pH, with 13 distinct classes, forming a 13-class classification problem (Y13). To create datasets with reduced class numbers, a label binarization process was applied. The pH values were grouped into three and five classes for the Y3 and Y5 datasets, respectively. These new datasets represent multi-class classification problems with reduced cardinality. Table II outlines the binning criteria used to transform the original 13 classes into the desired number of classes.

TABLE II. CLASSIFICATION SCENARIOS

Scenario	Label	pH low	pH high
Y3	Acidic	1	3
	Neutral	4	6
	Basic	7	13
Y5	Lethal Acidic	1	3
	Acidic	4	5
	Neutral	6	9
	Basic	9	10
	Lethal Basic	11	13

Fig. 2 presents a heat-map generated from Table I data that was used as the reference points for determining class boundaries. Colour intensity within the heat-map indicates the frequency with which a pH range was reported as optimal in the literature. These ranges were used as a reference to define pH class boundaries for the Y3 and Y5 datasets.

E. XGBoost Classifier

The XGBoost was selected as it offers a compelling alternative to complex deep learning mod-els, particularly for tabular data problems, due to its practicality, robustness, and effectiveness (Harrison, 2023). This algorithm belongs to a family of Ensemble learning, a meta-algorithmic framework that amalgamates multiple base models to enhance predictive accuracy and robustness, has witnessed substantial growth in recent years. Within this domain, boosting has gained recognition for its effectiveness as a sequential approach where models are iteratively developed to correct the errors of their predecessors. Among boosting algorithms, XGBoost stands out as a leading option, known for its computational efficiency, scalability, and superior predictive accuracy. Unlike traditional gradient boosting methods, XGBoost introduces key enhancements such as regularization, optimized tree construction, and robust handling of missing data. The primary objective of XGBoost is to minimize a loss function that includes both a differentiable error component and a regularization term, which helps prevent over-fitting. This approach is mathematically represented as:

$$L(\theta) = \Sigma_{i=1}^{n} l(y_i, \hat{y}_i) + \Omega(\theta)$$

The algorithm iteratively builds decision trees, with each tree refining the predictions made by the previous ensemble. XGBoost accelerates the training process by utilizing techniques such as approximation, weighted quantile sketches, and columnar storage. Furthermore, it incorporates a regularized objective function that includes both L1 and L2 regularization, which helps prevent overfitting and enhances generalization. XGBoost's ability to efficiently handle missing values, along with its parallel computing capabilities, has solidified its dominance in various machine learning competitions and real-world applications.

F. Hyperparameter Tuning

Hyperparameter tuning was performed utilizing a stepwise approach with Hyperopt. This method iteratively explores the space of possible hyperparameter values. In each iteration, a configuration is drawn from the search space, an XGBoost model is trained with those parameters, and the model's performance is evaluated using a chosen metric. This information is then used by Hyperopt to update its internal model of the search space, prioritizing regions that are more likely to contain high-performing configurations. This process continues until a stopping criterion, such as a maximum number of iterations, is met. The main advantage of this approach is its faster execution time compared to other methods. Table III shows the value of the hyper parameters for each classification scenario.

TABLE III. THE HYPER PARAMETER VALUES AS THE RESULT OF HYPERPARAMETER TUNING FOR EACH SCENARIO

Parameter	Y3	Y5	Y13
max_depth	4	8	5
min_child_weight	0.69	0.44	0.22
subsample	0.87	0.92	0.67
colsample_bytree	0.53	0.72	0.96
reg_alpha	0.59	0.89	0.47
reg_lambda	4.57	1.75	8.64
gamma	0.00	0.01	0.00
learning rate	0.52	0.27	0.48

IV. METHODOLOGY

As previously described, three distinct classification scenarios, labeled Y3, Y5, and Y13, were established. To optimize model performance, hyper-parameter tuning with early stopping was implemented for each scenario. The identified optimal parameters were subsequently employed to train respective models. A comprehensive evaluation of these models was undertaken, en-compassing precision, recall, F1-score, and support metrics. To provide a visual representation of model performance, a variety of diagnostic plots were generated. These included confusion matrix heat-maps to visualize classification accuracy, prediction error charts to identify patterns in mis-classifications, classification charts to assess overall performance, and AUC-ROC curves to evaluate the model's ability to discriminate between positive and negative classes.

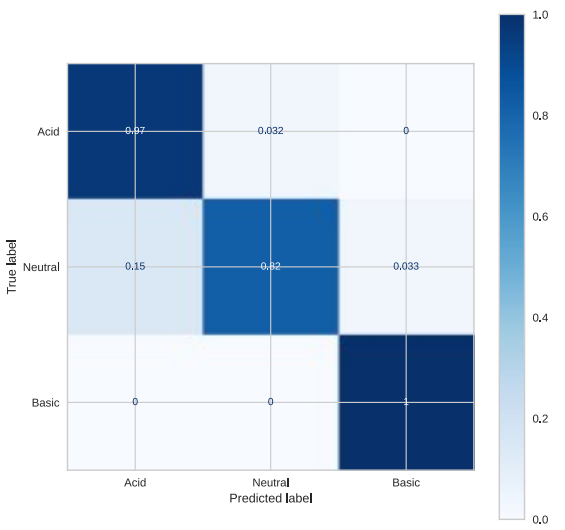
V. RESULTS

Fig. 3 presents the confusion matrices for the three scenarios. The results indicate that model performance generally improved as the number of classes decreased. In the Y3 scenario, the model achieved perfect classification of all Basic samples. Similarly, all samples in the "lethal acidic" category were correctly classified in the Y5 scenario. A common trend emerged, demonstrating superior model performance for samples at the extreme ends of the class spectrum. Conversely, classification accuracy tended to diminish for samples in the intermediate classes, a pattern particularly evident in the Y13 scenario.

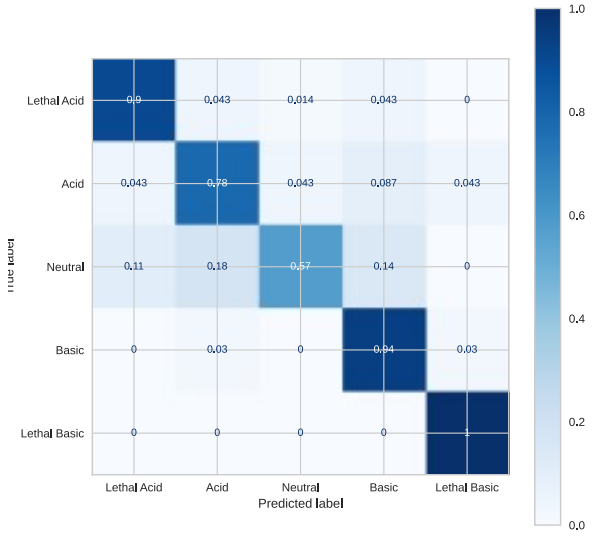
Fig. 4 illustrates the class prediction error profiles for the three classification models using stacked bar charts. Each bar represents a true class, segmented into stacked bars indicating the predicted classes. This visualization provides insights into the types of errors made by the models.

In the Y3 and Y5 scenarios shown in Fig. 4a and Fig. 4b respectively, the models demonstrated relatively strong performance, with most observations correctly classified within their respective true classes. However, the Y13 (show in Fig. 4c) model exhibited a more pronounced error pattern. The stacked bars for pH 9 and pH 5 classes were notably taller than others, indicating higher rates of mis-classification for these categories. This suggests potential class overlap, where instances of pH 9 might share similar feature characteristics with other classes, leading to confusion during the classification process. Additionally, the model might have been underrepresented in training data for these classes, contributing to lower classification accuracy. These findings highlight the challenges inherent in multi-class classification tasks, where class boundaries can be less distinct and model performance is influenced by factors such as data quality and feature engineering.

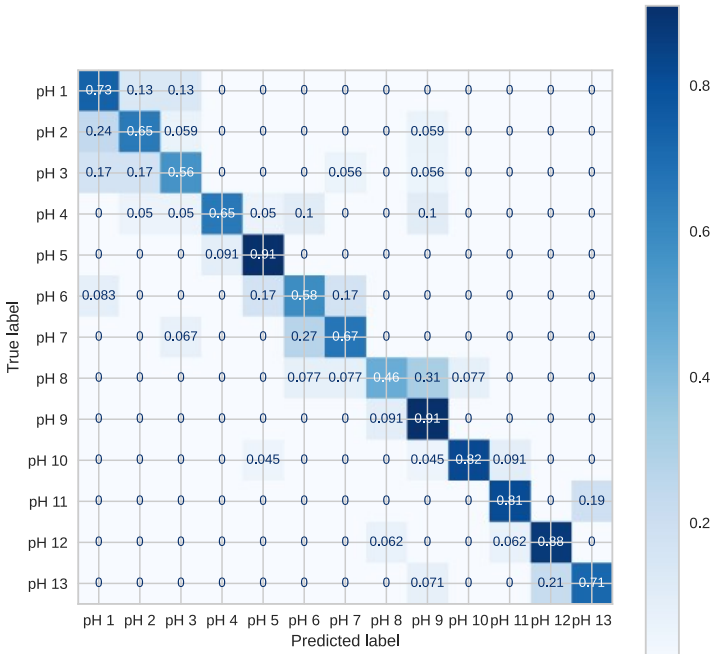
Fig. 5 presents a heatmap visualization of key classification metrics computed on a per-class basis. The heatmap encapsulates four critical performance indicators: precision, recall, F1-score, and support. Support represents the number of instances within each class, providing a measure of class distribution. Precision quantifies the accuracy of positive predictions, essentially the proportion of correctly predicted positive instances among all predicted positives. Recall, conversely, assesses a model's ability to identify all relevant instances, calculated as the ratio of correctly predicted positive instances to the total



(a) Y3



(b) Y5



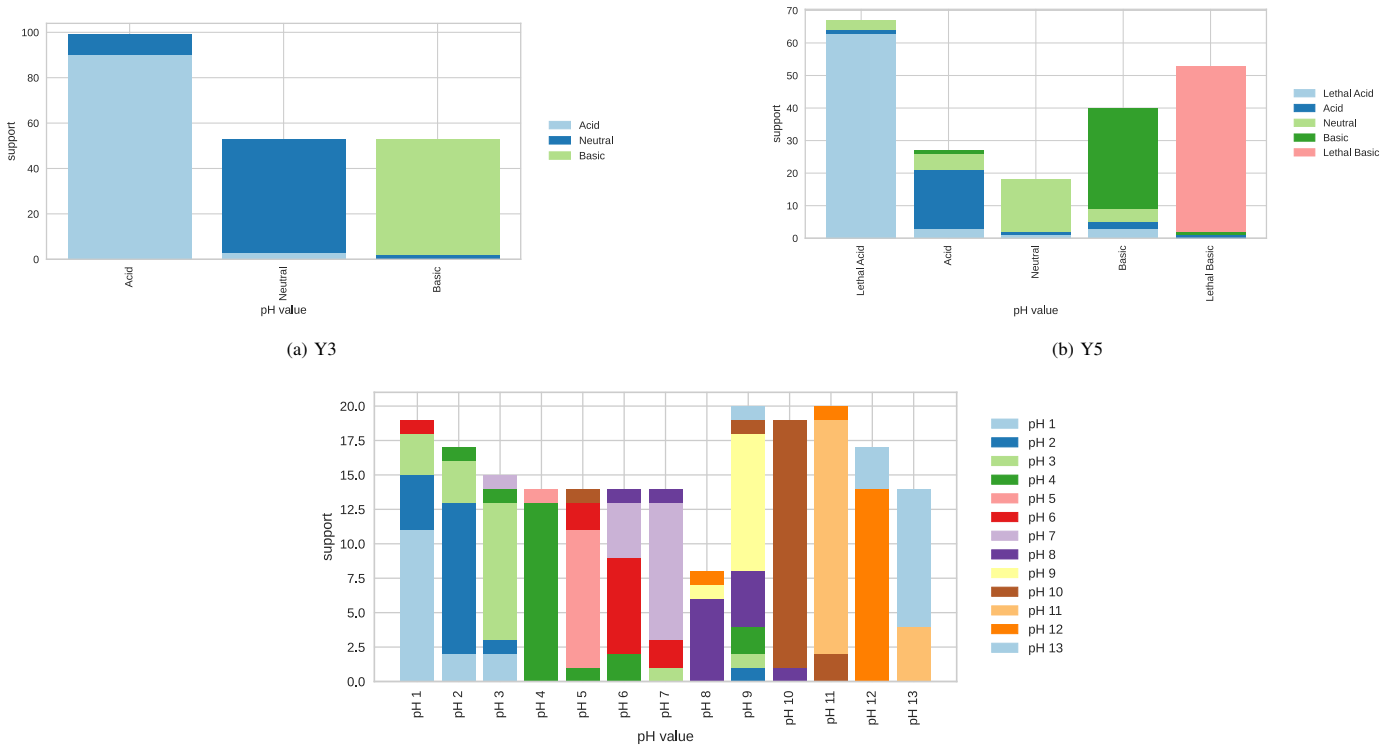
(c) Y13

Fig. 3. Confusion matrices.

number of actual positive instances. The F1-score, a harmonic mean of precision and recall, offers a balanced evaluation of a model's performance, providing a single metric that considers both precision and recall. This metric ranges from 0 to 1, with higher values indicating superior performance. A weighted average of the F1-scores across all classes is commonly employed to compare the overall effectiveness of different classification models.

Referring to Fig. 5c we can observe that while several classes demonstrated robust classification capabilities, as indicated by F1-scores exceeding 0.7, a subset of classes exhibited suboptimal performance. These classes with lower F1-scores suggest potential challenges in accurately identifying and classifying instances within these categories, warranting further investigation into factors such as class imbalance, data quality, or model complexity.

The original model, i.e. Y13 model, exhibited variable performance across classes, with several notably lower F1-scores. While the distribution of instances across classes appeared balanced, as indicated by the support column in the classification report, the intrinsic challenge of the problem became apparent. The data, primarily composed of images with subtle variations in shades of yellow, presented a complex classification task. The model's difficulties in distinguishing



(a) Y3

(b) Y5

(c) Y13

Fig. 4. Class prediction errors.

be-tween these closely related visual features, coupled with the limited discriminative power of the available features, likely contributed to the sub-optimal performance. These findings suggest that enhancing feature engineering or exploring more sophisticated image processing techniques might be necessary to improve classification accuracy for Y13 scenario.

The Y3 and Y3 scenario exhibit notably higher average F1-scores compared to the original thirteen-class model. This suggests that consolidating the original classes into fewer categories has led to improved classification performance. The Y3 model, in particular, demonstrates consistently strong performance across all classes, with F1-scores above 0.88. The Y5 model also shows promising results, with three classes achieving F1-scores above 0.7. These findings imply that the complexity introduced by the thirteen-class model might have hindered its ability to accurately discriminate between closely related classes. By reducing the number of classes, the models were able to focus on more distinct categorical boundaries, resulting in enhanced classification accuracy.

The ROC AUC curves, as shown in Fig. 6, provide valuable insights into the discriminative power of the models across different class scenarios. The Y3 model (see Fig. 6a) exhibits exceptional performance, with all classes achieving AUC values close to 1. This indicates an outstanding ability to differentiate between the three classes. While the Neutral class in Y3 shows a slightly lower AUC of 0.94 compared to the other two classes, the overall performance remains exceptionally high. The same is true for the Y5 (see Fig. 6b) model as it has shown comparable performance with the Y3 model.

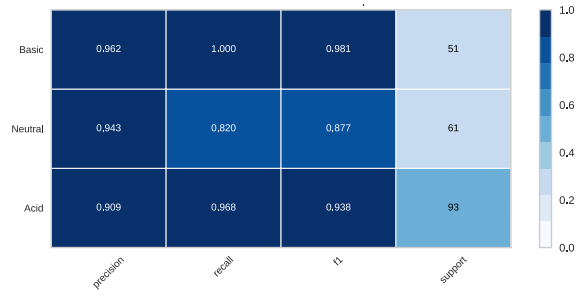
In contrast, the Y13 model (see Fig. 6c) presents a more complex picture. While the macro and micro average AUC values of 0.96 are still commendable, the individual class performance varies. Classes “pH 6” and “pH 3” show notably lower AUC values, suggesting challenges in distinguishing these classes from others. This aligns with the previously discussed difficulties in classifying closely related shades of yellow.

Overall, the ROC AUC analysis reinforces the findings from the F1-score evaluation. The Y3 model demonstrates superior discriminative power, likely due to the increased class separability. The Y13 model, while showing good overall performance, struggles with certain classes, emphasizing the impact of feature similarity and the potential limitations of the current feature set.

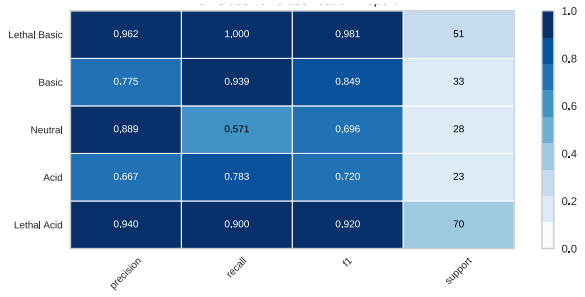
These results further support the conclusion that refining feature engineering or exploring more advanced image processing techniques could be crucial for improving the performance of the thirteen-class model.

This research evaluated the feasibility of developing a cost-effective smartphone-based pH detection system for aquaculture. While commercial pH sensors provide precise measurements, their high cost limits their adoption by small-scale farmers.

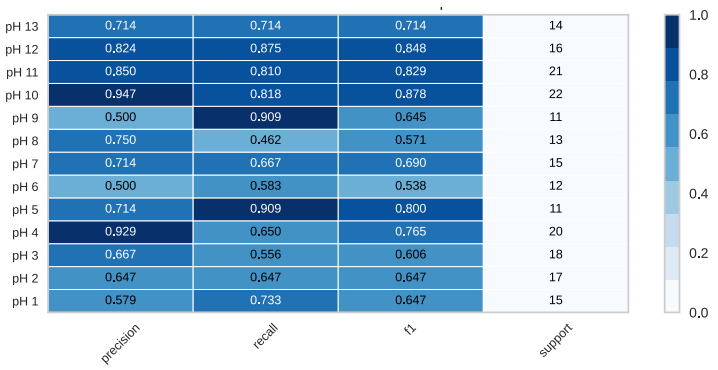
To address this gap, we developed a method using readily available materials like mango leaves, smartphones, and simple



(a) Y3



(b) Y5



(c) Y13

Fig. 5. Classification report.

photography equipment. Our results demonstrate that while models like Y13 may not achieve the precision of commercial sensors, models Y3 and Y5 can effectively estimate pH ranges. This level of accuracy aligns with the practical needs of most aquaculture farmers, as precise pH values are often less critical than maintaining pH within specific ranges (see Table I).

Our proposed method offers a low-cost, accessible alternative to traditional pH monitoring. By reducing reliance on expensive equipment, it has the potential to improve water quality management in aquaculture. However, to ensure reliable results, farmers should conduct multiple tests and consider factors such as water change frequency and potential water treatment measures.

VI. Conclusion

This study successfully demonstrated the potential of a smartphone-based sensor for pH monitoring in aquaculture settings. By integrating the colorimetric properties of mango leaf extract with advanced image processing and machine learning techniques, we developed a predictive model for pH levels. This approach offers a sustainable and economically viable alternative to traditional pH sensors.

The use of a natural indicator aligns with eco-friendly aquaculture practices while the smartphone platform enhances accessibility. Farmers can potentially utilize this device for regular water quality assessments, enabling proactive pond management.

While the initial results are promising, further research is imperative to improve the model's precision and reliability under diverse environmental conditions. Extensive field trials are necessary to validate the sensor's effectiveness in real-world aquaculture scenarios. Overcoming these challenges is crucial for transforming the technology into a practical and indispensable tool for aquaculture practitioners.

Ultimately, the development of cost-effective and user-friendly water quality monitoring solutions is essential for the sustainable growth of the aquaculture industry. This research represents a significant step towards achieving this goal. By providing farmers with data-driven insights into water quality, we can enhance the overall health and productivity of aquaculture systems while minimizing environmental impact.

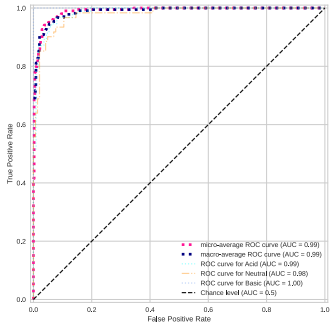
Acknowledgment

The authors would like to thank the Research Management Office (RMO) at Universiti Malaysia Terengganu, for the financial support of this research through Talent and Publication Enhancement-Research grant 2023 (UMT/TAPE-RG/55497)

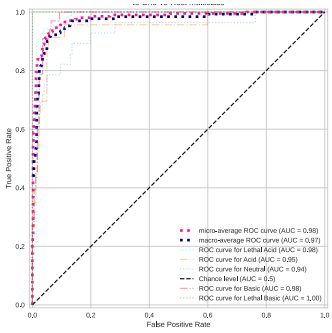
References

[1] N. Abdullah, A. S. Chowdhury, M. M. Hossain, O. Dutta, and J. Uddin, "Analysis of aquaculture for cultivating different types of fish," in *Applied Informatics for Industry 4.0*. Chapman and Hall/CRC, 2023, pp. 83–96.

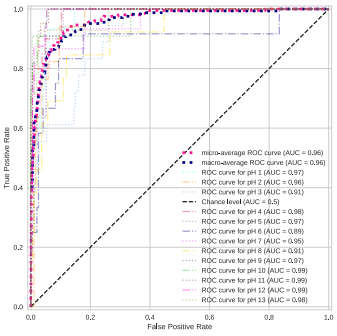
[2] P. Lindholm-Lehto, "Water quality monitoring in recirculating aquaculture systems," *Aquaculture, Fish and Fisheries*, vol. 3, no. 2, pp. 113–131, Apr. 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/aff2.102



(a) Y3



(b) Y5



(c) Y13

Fig. 6. Receiver Operating Characteristic (ROC) to One-vs Rest runtime.

[3] T. Y. Wei, E. S. Tindik, C. F. Fui, H. Haviluddin, and M. H. A. Hijazi, "Automated water quality monitoring and regression-based forecasting system for aquaculture," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, pp. 570–579, Feb. 2023. [Online]. Available: https://beei.org/index.php/EEI/article/view/4464

[4] D. Kültz, "213Abiotic parameters," in *A Primer of Ecological Aquaculture*. Oxford University Press, 09 2022. [Online]. Available: https://doi.org/10.1093/oso/9780198850229.003.0016

[5] P. C. Lindholm-Lehto, "Water quality monitoring in recirculating aquaculture systems," *Aquaculture, fish and fisheries*, vol. 3, no. 2, pp. 113–131, 2023.

[6] . Flura, M. Moniruzzaman, M. A. Alam, M. H. Rashid, M. H. Rahman, M. A. Rahman, and Y. Mahmud, "An Assessment of the Water Quality Factors: A Case of Hilsa Fishery River Areas," *Asian Journal of Fisheries and Aquatic Research*, pp. 30–47, Oct. 2022. [Online]. Available: https://journalajfar.com/index.php/AJFAR/article/view/486

[7] H. Rastegari, F. Nadi, S. S. Lam, M. Ikhwanuddin, N. A. Kasan, R. F. Rahmat, and W. A. W. Mahari, "Internet of Things in aquaculture: A review of the challenges and potential solutions based on current and future trends," *Smart Agricultural Technology*, vol. 4, p. 100187, Aug. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772375523000175

[8] J. A. Hargreaves, L. D. Sheely, and F. S. To, "A Control System to Simulate Diel pH Fluctuation in Eutrophic Aquaculture Ponds," *Journal of the World Aquaculture Society*, vol. 31, no. 3, pp. 390–402, Sep. 2000. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1749-7345.2000.tb00889.x

[9] N. Mohamed Ramli, C. Giatsis, F. Md Yusoff, J. Verreth, and M. Verdegem, "Resistance and resilience of small-scale recirculating aquaculture systems (RAS) with or without algae to pH perturbation," *PLOS ONE*, vol. 13, no. 4, p. e0195862, Apr. 2018. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0195862

[10] L. Di Costanzo and B. Panunzi, "Visual pH Sensors: From a Chemical Perspective to New Bioengineered Materials," *Molecules*, vol. 26, no. 10, p. 2952, may 2021. [Online]. Available: https://www.mdpi.com/1420-3049/26/10/2952

[11] G. Li, H. Su, N. Ma, G. Zheng, U. Kuhn, M. Li, T. Klimach, U. Pöschl, and Y. Cheng, "Multifactor colorimetric analysis on pH-indicator papers: an optimized approach for direct determination of ambient aerosol pH," *Atmospheric Measurement Techniques*, vol. 13, no. 11, pp. 6053–6065, nov 2020. [Online]. Available: https://amt.copernicus.org/articles/13/6053/2020/

[12] K. Vizarova, I. Vajova, N. Krivonakova, R. Tino, Z. Takac, S. Vodny, and S. Katuscak, "Regression Analysis of Orthogonal, Cylindrical and Multivariable Color Parameters for Colorimetric Surface pH Measurement of Materials," *Molecules*, vol. 26, no. 12, p. 3682, jun 2021. [Online]. Available: https://www.mdpi.com/1420-3049/26/12/3682

[13] W. X. C. S. W. X. W. L. K. Q. Y. M. L. X. X. J. Z. C. X. Y., "pH-Sensitive Dye-Based Nanobioplatform for Colorimetric Detection of Heterogeneous Circulating Tumor Cells," *ACS Sensors*, vol. 6, no. 5, pp. 1925–1932, may 2021. [Online]. Available: https://pubs.acs.org/doi/10.1021/acssensors.1c00314

[14] H. Chen, F. Ding, Z. Zhou, X. He, and J. Shen, "FRET-based sensor for visualizing pH variation with colorimetric/ratiometric strategy and application for bioimaging in living cells, bacteria and zebrafish," *The Analyst*, vol. 145, no. 12, pp. 4283–4294, 2020. [Online]. Available: https://xlink.rsc.org/?DOI=D0AN00841A

[15] Y.-J. Wang, T. Yang, and H.-J. Kim, "pH Dynamics in Aquaponic Systems: Implications for Plant and Fish Crop Productivity and Yield," *Sustainability*, vol. 15, no. 9, p. 7137, apr 2023. [Online]. Available: https://www.mdpi.com/2071-1050/15/9/7137

[16] L. Zhou, J. Huang, Y. Jiang, J. Kong, X. Xie, and F. Yin, "pH Regulates the Formation and Hatching of *Cryptocaryon irritans* Tomonts, Which Affects Cryptocaryoniasis Occurrence in *Larimichthys crocea* Aquaculture," *Applied and Environmental Microbiology*, vol. 88, no. 7, pp. e00 058–22, apr 2022. [Online]. Available: https://journals.asm.org/doi/10.1128/aem.00058-22

[17] E. D. Chang, R. M. Town, S. F. Owen, C. Hogstrand, and N. R. Bury, "Effect of Water pH on the Uptake of Acidic (Ibuprofen) and Basic (Propranolol) Drugs in a Fish Gill Cell Culture Model," *Environmental Science & Technology*, vol. 55, no. 10, pp. 6848–6856, may 2021. [Online]. Available: https://pubs.acs.org/doi/10.1021/acs.est.0c06803

[18] J. R. Bowman and J. E. Lannan, "Evaluation of Soil pH-Percent Base Saturation Relationships for Use in Estimating the Lime Requirements of Earthen Aquaculture Ponds," *Journal of the World Aquaculture Society*, vol. 26, no. 2, pp. 172–182, jun 1995. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1749-7345.1995.tb00241.x

[19] C. M. McGraw, C. E. Cornwall, M. R. Reid, K. I. Currie, C. D. Hepburn, P. Boyd, C. L. Hurd, and K. A. Hunter, "An automated pH-controlled culture system for laboratory-based ocean acidification experiments," *Limnology and Oceanography: Methods*, vol. 8, no. 12, pp. 686–694, dec 2010. [Online]. Available: https://aslopubs.onlinelibrary.wiley.com/doi/10.4319/lom.2010.8.0686

[20] P. G. Daye and E. T. Garside, "Lower lethal levels of pH for embryos and alevins of Atlantic salmon, *Salmo salar* L." *Canadian Journal of Zoology*, vol. 55, no. 9, pp. 1504–1508, sep 1977. [Online]. Available: http://www.nrcresearchpress.com/doi/10.1139/z77-194

[21] L. S. D. Andrade, R. L. B. D. Andrade, A. G. Becker, L. V. Rossato, J. F. D. Rocha, and B. Baldisserotto, "Interaction of Water Alkalinity and Stocking Density on Survival and Growth of Silver Catfish, *Rhamdia quelen* , Juveniles," *Journal of the World Aquaculture*

Society, vol. 38, no. 3, pp. 454–458, sep 2007. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/j.1749-7345.2007.00118.x

[22] J. Lopes, L. Silva, and B. Baldisserotto, "Survival and growth of silver catfish larvae exposed to different water pH," *Aquaculture International*, vol. 9, no. 1, pp. 73–80, 2001. [Online]. Available: http://link.springer.com/10.1023/A:1012512211898

[23] L. Takahashi, Keisuke; Takahashi, "Supervised Machine Learning," in *Materials Informatics and Catalysts Informatics*. Singapore: Springer Nature Singapore, 2024, pp. 191–226. [Online]. Available: $https://link.springer.com/10.1007/978-981-97-0217-6_8$

[24] L. Vanneschi and S. Silva, "Ensemble Methods," in *Lectures on Intelligent Systems*. Cham: Springer International Publishing, 2023, pp. 283–288, series Title: Natural Computing Series. [Online]. Available: $https://link.springer.com/10.1007/978-3-031-17922-8_11$

[25] G. Aryotejo, P. W. Adi, and E. A. Sarwoko, "Water quality monitoring with an early warning system for enhancing the shrimp aquaculture production," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, p. 1042, may 2024. [Online]. Available: https://ijeecs.iaescore.com/index.php/IJEECS/article/view/35583

[26] F. Akhter, H. R. Siddiquei, M. E. E. Alahi, K. P. Jayasundera, and S. C. Mukhopadhyay, "An IoT-Enabled Portable Water Quality Monitoring System With MWCNT/PDMS Multifunctional Sensor for Agricultural Applications," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 14 307–14 316, aug 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9390294/

[27] M. Singh, K. S. Sahoo, and A. Nayyar, "Sustainable IoT Solution for Freshwater Aquaculture Management," *IEEE Sensors Journal*, vol. 22, no. 16, pp. 16 563–16 572, aug 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9827934/

[28] K. Devarayan, D. Anandakumar Muthurani, G. Kannusamy, A. Theivasigamani, Y. Palanisamy, G. Mohan, M. Sukumaran, E. U. Siluvai John, R. Marimuthu, and H. Anjappan, "On-site colorimetric determination of pH in brackish water aquaculture," *Pigment & Resin Technology*, may 2024. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/PRT-11-2023-0099/full/html

[29] A. Jain, S. Wadhawan, V. Kumar, and S. K. Mehta, "pH-Sensing Strips Based on Biologically Synthesized Ly-MgO Nanoparticles," *ACS Omega*, vol. 4, no. 26, pp. 21 647–21 657, dec 2019. [Online]. Available: https://pubs.acs.org/doi/10.1021/acsomega.9b01306

[30] S. Borsci, P. Buckle, J. Huddy, Z. Alaestante, Z. Ni, and G. B. Hanna, "Usability study of pH strips for nasogastric tube placement," *PLOS ONE*, vol. 12, no. 11, p. e0189013, nov 2017. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0189013

[31] N. A. Metheny, E. M. Gunn, C. S. Rubbelke, T. F. Quillen, U. R. Ezekiel, and K. L. Meert, "Effect of pH Test-Strip Characteristics on Accuracy of Readings," *Critical Care Nurse*, vol. 37, no. 3, pp. 50–58, jun 2017. [Online]. Available: https://aacnjournals.org/ccnonline/article/37/3/50/3578/Effect-of-pH-TestStrip-Characteristics-on-Accuracy

[32] S. D. Kim, Y. Koo, and Y. Yun, "A Smartphone-Based Automatic Measurement Method for Colorimetric pH Detection Using a Color Adaptation Algorithm," *Sensors*, vol. 17, no. 7, p. 1604, jul 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/7/1604

[33] A. Mutlu, A. Y. Mutlu, V. Kılıç, V. Kilic, G. K. Özdemir, G. K. Özdemir, A. Bayram, A. Bayram, N. Horzum, N. Horzum, M. Solmaz, and M. E. Solmaz, "Smartphone-based colorimetric detection via machine learning," *Analyst*, 2017.

[34] E. V. Woodburn, K. D. Long, and B. T. Cunningham, "Analysis of Paper-Based Colorimetric Assays With a Smartphone Spectrometer," *IEEE Sensors Journal*, vol. 19, no. 2, pp. 508–514, jan 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8494768/

[35] H. Park, Y. G. Koh, and W. Lee, "Smartphone-based colorimetric analysis of structural colors from pH-responsive photonic gel," *Sensors and Actuators B: Chemical*, vol. 345, p. 130359, oct 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925400521009278

[36] L. Zou, C. Mai, M. Li, and Y. Lai, "Smartphone-assisted colorimetric sensing of enzyme-substrate system using pH-responsive gold nanoparticle assembly," *Analytica Chimica Acta*, vol. 1178, p. 338804, sep 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0003267021006309

[37] A. Dutta and A. Zisserman, "The VIA Annotation Software for Images, Audio and Video," in *Proceedings of the 27th ACM International Conference on Multimedia*, ser. MM '19. New York, NY, USA: ACM, 2019, event-place: Nice, France. [Online]. Available: https://doi.org/10.1145/3343031.3350535