International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 3
ISSN: 2088-5334

Ensemble Learning Regression Method for Glucose Concentration Prediction System using Colorimetric Paper-based and Smartphones

Dian Wulan Hastuti [a], Adhi Harmoko Saputro [a,*], Cuk Imawan [a]

[a] *Department of Physics, Faculty of Science and Mathematics, Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia*
Corresponding author: [*]adhi@sci.ui.ac.id

***Abstract*— Prediction of glucose concentration on android smartphones and colorimetric paper-based using the Ensemble learning regression model has been successfully developed. Several successful developments in our research include automatic image segmentation, image correction using the RPCC method, and the development of a regression model for urine glucose predictions. Furthermore, the model was successfully validated for best performance in the respondent's urine susceptible to color change. We used artificial urine at a 0–2000 mg/dl concentration to create a regression model based on Ensemble learning with the boosting optimization method. In addition, we also compared the Ensemble Bagging regression model and the single learner model, Decision Tree. Server-based applications were also developed using RESTful API communication with two servers: an upload server using Node.js and a computing server using the MATLAB Production Server. The testing process results using artificial urine samples showed that the performance of R^2 and RRMSE were 0.98 and 0.05 for the Decision Tree and Ensemble Bagging regression models, respectively. While for the Ensemble Boosting regression model, R^2 and RRMSE at the testing process are 0.98 and 0.04. The best validation results using respondents' urine samples are shown in the Ensemble Boosting regression model with R^2 and RRMSE performance values of 0.97 and 0.06, respectively. The success rate of the application was 100% on both the Samsung Galaxy A51 and Huawei Nova 5T. This research estimated the glucose concentration reasonably well for health monitoring applications.**

***Keywords*— Ensemble; glucose; MATLAB production server; RESTful API; RPCC; urine.**

Manuscript received 20 Jul. 2021; revised 24 Aug. 2021; accepted 11 Oct. 2021. Date of publication 30 Jun. 2022.

I. INTRODUCTION

Automated colorimetric analysis using mobile platforms can represent environmental parameters, health monitoring status, food safety, crop harvests, and forensic analysis. Several studies using colorimetric methods are in the food safety field as smart packaging [1], [2] and in the environmental field as a quality monitoring system, and water [3]–[5] and toxic gas monitoring systems [6]–[8]. In the health sector, colorimetry is used in analytical papers (Microfluidic Paper-based Analytical Devices (μ-PADs)) [9]–[13], lab on chip (LOC) devices [14]–[16], and test strips [17], [18].

The test strip is the oldest and has been the most used analytical tool for more than 60 years [19]. Test strips are commonly used in several hospitals in Indonesia, especially for urinalysis. A colorimetric response occurs when the reagent strip comes in contact with the specific analyte. The reaction appears relatively short after reacting with the urine analyte [18], [19]. The color change is visually compared with the manufacturer's reference color standard as a class of a measured analyte concentration [20]. In addition, the strip test method makes it possible to detect multiple analytes at once with just one immersion into the sample [21].

However, this device is considered to have several weaknesses, such as limitations related to the manual interpretation of the displayed results [19]. People with eye disorders, color blindness, and the elderly make visual analysis with the naked eye more difficult [22]. The subjectivity factor in color perception, especially concerning ambient lighting conditions, causes a high level of error in readings of strip tests [23].

One application of smartphone-based colorimetry applications is in controlling diabetes mellitus [23], [24]. Based on the 2016 WHO data release, the number of diabetes cases has continued to increase over the last few decades. Globally, it is estimated that 422 million adults over 18 lived with diabetes in 2014. These cases represent a prevalence rate of 8.5%, which has increased substantially from 4.7% in 1980. The most significant number of people with diabetes is estimated to be in Southeast Asia and the Western Pacific,

accounting for about half of the world's diabetes cases [25], [26].

An attempt to provide automatic and error interpretation of colorimetric results is developing commercial reader tools such as Konica Minolta, an easy-to-use colorimetric instrument for analyzing color consistency without needing a computer [27]. A commercial reading device in the medical field since the 1990s is the Clinitek Atlas (Siemens Healthcare Diagnostics, Eschborn, Germany). Clinitek Atlas is a spectroscopic method-based tool specifically designed to analyze urine [28]. However, these two devices are considered relatively expensive as home-based personal care tools and are also not practical to carry around. In addition, for analysis needs in hospitals, this tool is also considered less effective in data transmission and communication between health workers and doctors [19]. Expensive equipment maintenance costs have caused health facilities in Indonesia to return to using test strip-based colorimetric devices with visual readings without an automatic reader.

Also, studies related to urinalysis using the supervised learning method with smartphones have been widely reported in previous studies. Furthermore, detection of other urinary analytes using artificial urine was also carried out, such as analytes of hemoglobin, red blood cells, protein, nitrite, leukocytes, ketones, vitamin C, urobilinogen bilirubin, and specific gravity [29]–[31]. These studies still focused on a person's health quality detection model but did not test real urine. Direct testing of the patient's urine can be an opportunity to research health problems in the community.

The supervised learning method in the machine learning model of the SVM classification produces 100% accuracy in applying pH test strips using a pH buffer solution [32]. Furthermore, the regression model in machine learning ANN produces an MAE value of 0.071 in the urinalysis application using a test strip with 11 analytes [19]. Our preliminary study used a colorimetric method to predict Brassica Oleracea (red cabbage) pH using Ann and KNN supervised machine learning regression. The system performance results were expressed by a correlation coefficient of 99.83, and a minimum RMSE value of 0.11 was obtained [33]. The detection of sheep serum sugar levels using Support Vector Regression, Random Forest, and PLSR yielded a performance of 0.96 on the correlation coefficient. Another machine learning model for predicting Crypto-Currency price is a decision tree that yields 96% accuracy [34]. The following classifier Ensemble method is a Random Forest used to detect TB antigen-specific antibodies with a smartphone, resulting in an accuracy of 98.4% [35].

In addition to using a single base learner algorithm described, the use of multiple learners for numerical analysis with more complex data and is considered capable of providing more accurate and more stable prediction results has also been widely reported [36]. Short-term traffic prediction uses a gradient boosting machine, which produces an MAE performance and a MAPE of 16.99 and 0.07458, respectively [37]. Ensemble boosting and bagging based machine learning models for potential groundwater prediction with Random Forest model in performance with accuracy = 0.86, Kappa = 0.67, Precision = 0.85, and Recall = 0.91 [38].

This work aimed to develop a system for quantitative analysis of urine glucose concentration using a smartphone in real-time. The main development of the research was that the image segmentation method is carried out automatically to obtain a specific Region of Interest (ROI) from the test strip, glucose pad. Color correction using the Root Polynomial Color Correction (RPCC) method with color standards adopted from the X-Rite ColorChecker was applied to compensate for the different image features from more than one smartphone [39], [40]. Ensemble learning regression with Boosting optimization with sequentially multiple learners fitting will be used. It will be compared with Bagging optimization and a single learner, Decision tree as training and testing data on an artificial urine sample. Furthermore, the model was then used as data validation in the respondent's urine at the Tk. 1 Said Sukanto Hospital, Jakarta, Indonesia.

Multiple learners' automatic prediction of urine glucose concentrations creates a major contribution to research that can be used as a personal monitoring test and medical treatment tool in hospitals [41], [42]. In the future, this research could be an alternative technology for preventing diabetes cases in Indonesia.

II. Materials and Methods

This section presents the material used, namely the urine sample and the respondents' urine sample. Furthermore, the research methods included image acquisition design, prediction system design, and the implementation system.

A. Sample Preparation

Sample preparation included the preparation of artificial urine and patient urine samples. Artificial urine was intended to be used as regression model training data, while respondents' urine was used as validation data. Artificial urine was made by dissolving the chemical standard of the analyte according to the needs of the specimen under study. The chemical used to make artificial urine was dextrose powder (Merck, Germany). Deionized water (DI) was then used to obtain 0–2000 mg/dl glucose concentrations.

The respondents' urine samples used for data validation were obtained from Tk. 1 Said Sukanto Hospital, Jakarta, Indonesia. Sampling was carried out under the standard operating procedures by collecting the middle urine obtained after the patient's initial and final urination. This process aimed to obtain urine that represents the patient's health condition. This study collected urine samples from 20 patients with various health conditions for further analysis. The collected urine was then taken to the Department of Physics, Faculty of Mathematics and Natural Sciences at Universitas Indonesia to examine its analyte content.

B. Image Acquisition Design

The image acquisition process was carried out using strip housing, which controls environmental conditions to produce reproducible images and reduces interference caused by illumination factors, sensor sensitivity, and reflectance. The strip housing design consisted of three main parts: the head, body, and test strip holder. The image acquisition process used the head strip housing as a smartphone holder. There is a light source at the head, namely a hard-strip LED type. The last part of the strip housing is the test strip holder located at

the bottom of the strip housing body. A color board on the strip housing holder functions as a correction factor for image results.

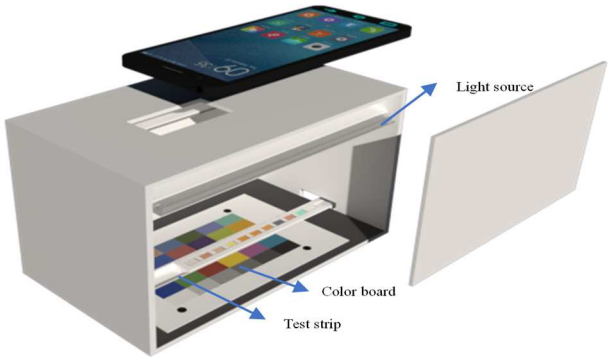


Fig. 1 Strip housing

The use of color boards was intended to facilitate the segmentation process after the image was acquired. In addition, the color board was used as a color correction board in the image under study. The color board design consisted of 24 color reference segments adopted from the X-Rite Color Checker. The reference standard printing process was carried out on a white acrylic board, which was then placed at the bottom of the strip housing in the image acquisition process. The design of the color board set is shown in Fig. 2. There was a holder for placing the test strip on the color board design, namely the stripboard.

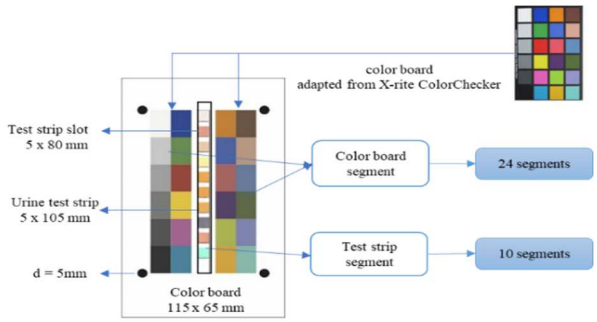


Fig. 2 Color board design

The urine test strips used were those of the Verify™ manufacturer, which are standard test strips commonly used to monitor the patient's health condition and are available on the market in Indonesia. The glucose analyte was in the 10th position on the Verify™ urine test strip in this study.

(a) (b)

Fig. 3 Urine strip test: (a) reference standard Verify paper sheet, and (b) test stripboard

This study used two smartphones with equivalent specifications and prices: the Samsung A51 (SM51) and the Huawei Nova 5T (HN5T). The camera sensors used in both smartphones come from the SONY manufacturer, with the types being IMX582 and IMX586, respectively. The primary camera used on both smartphones is 48MP with f/2.0 on the SM51 and f/1.8 on the HN5T. Furthermore, the Open Camera application is also used to take images with standard mode settings. AUTO settings determine the ISO, and white balance and flashlight are turned off.

C. Prediction System Design

The design of the prediction system in this study was divided into two following methods: the design of the concentration prediction modeling algorithm carried out on a computer and the implementation of the system on a smartphone. In this case, the computer was used to conduct training and testing at each stage of the process. The determination of the correction steps, the selection of segmentation methods, and the classification model parameters were carried out appropriately to determine which actions produced the best performance. Fig. 4 is an algorithm design for a glucose concentration prediction system based on urine test strip images.

The design of the glucose concentration prediction system was carried out in two processing stages: pre-processing and processing. Data pre-processing is an important step in the data mining process to prepare the primary processing or further analysis. The initial stage of pre-processing was filtering and image segmentation. The processing stage included image correction, feature extraction, and modelling to obtain the best performance prediction results of urine analyte concentration. These stages and processes were implemented using MATLAB 2020b and a MATLAB Production Server to deploy prediction algorithms in the next step.

A Gaussian filter was applied to remove noise in the image and improve the quality of the target image. Next, the filtered image was segmented. The segmentation process aimed to divide the image into several regions of interest segments (ROI). Segmentation was done automatically based on the coordinate position of each color block using the template matching method, which was supposed to make it easier to determine the ROI positioning. The matching stage was used to detect two segments: 24 ROI of color boards and 10 ROI test strips.

The segmented image then entered the correction stage, an image quality improvement process [43]. In this study, the image correction process applied the Root Polynomial Color Correction (RPCC) method, which was done by optimizing the image correction matrix. The smallest value of color difference of ΔE (Euclidean distance) was transformed on the input image features so that the corrected image results were close to the reference value. Image correction consisted of three main stages: the training, application, and validation stages.

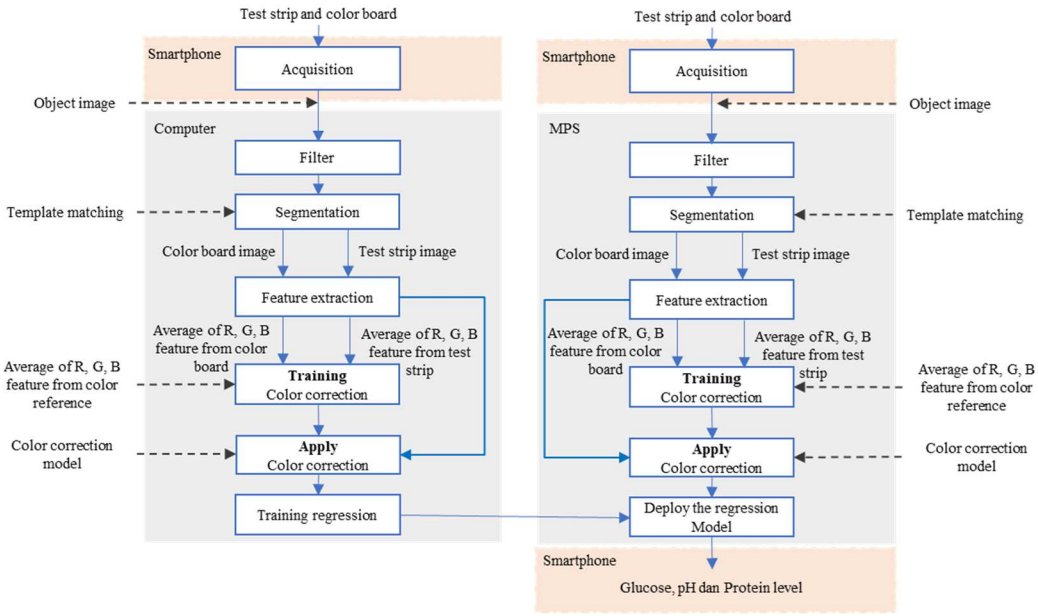


Fig. 4 Block diagram of diabetic urine concentration prediction system

The training phase was carried out by finding the optimal value of the color correction matrix in each reference segment and test strip segment. At this stage, the sRGB image was transformed in the XYZ 1931 color space by using the D65 white reference, whose standard luminance was the color standard, at 6504 K color temperature. The training process was carried out to obtain a matrix with a minimal loss function value as a color correction matrix. At this stage, the components of lightness (l), chroma (c), and hue (h) were optimized to obtain a correction matrix with a color close to the reference value. Furthermore, the correction matrix was applied to the input image to obtain the optimally corrected image.

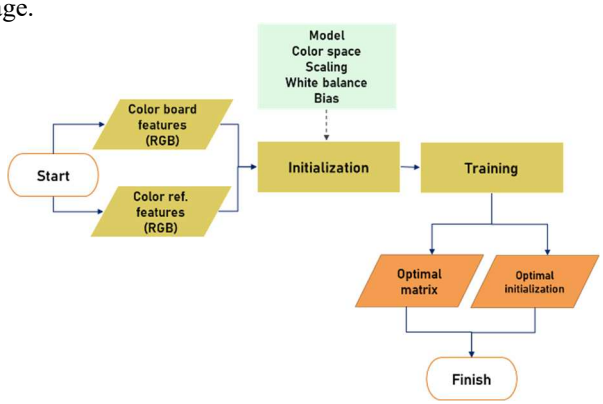


Fig. 5 Color correction method

The validation process was carried out by calculating the value of ΔE in the image before and after correction. In this study, CIEDE2000 was used to calculate that the value of ΔE = 0–3 is in the best category, and it can be said that the color correction was close to the actual value. Meanwhile, the value of ΔE = 3–6 was excellent and sufficient when ΔE = 6–10. In addition, the loss function value was taken into account so that the optimally corrected image was obtained for further processing [44].

After the acquisition, segmentation, and correction stages were completed, image feature extraction was carried out in the color space consisting of image building pixels, and the Red (R), Green (G), and Blue (B) channels. In feature extraction, the averaging method was used. After the data features were extracted, the data were divided into training and testing data. K-fold cross-validation was used to produce an optimal result. Furthermore, the data was processed with the Ensemble regression model using boosting optimization.

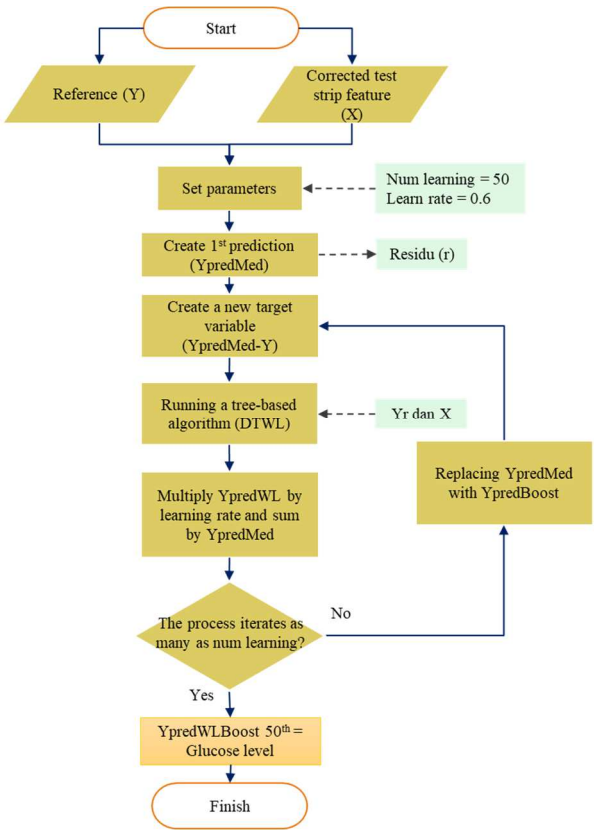


Fig. 6 Diagram flow of regression model

This method begins by determining the initial predictive value (YpredMed) using the median value of the target

variable (Y) as a function of the predictor variable (X). Furthermore, the value of YpredMed was then used to calculate the residual (r) against the target variable (Y). The residue (r) then replaces the target variable (Yr). Using a tree regression-based model (weak learners) was predicted as a predictor variable (X) function to produce YpredWL. Next, the value of YpredWL is multiplied by the learning rate and summed by the Value of YpredMed. The result obtained is YpredWLBoost, the final prediction result of LSBoost, whose RMSE value can be calculated. Fig. 6is, a block diagram of the regression model, used. Furthermore, boosting optimization method was also compared with the regression model using Ensemble learning with bagging optimization and a single learner, Decision Tree.

D. Implementation System

The prediction application design on a smartphone consisted of two tasks: image processing and graphical user interface (GUI). The regression model that was successfully tested and produced the best performance was compiled using the MATLAB Compiler SDK to produce a file in the *.ctf format. Furthermore, the application was made both on the client- and server-side. In this research, we used two servers, namely Node.js and the MATLAB Production server. The Node.js-based server was used for the image file upload process, and a MATLAB-based server to process image files that were already at the location of the uploaded image into a predicted total chlorine value. The client-side was built using Android Studio.

III. Results and Discussion

A. Sample Description

In this study, the samples used came from artificial urine samples and respondents' urine samples taken from Tk. 1 Said Sukanto Hospital, Jakarta, Indonesia. Furthermore, the samples were divided into low concentration glucose samples at 0–700 mg/dl and high concentration glucose samples at 700–2000 mg/dl. The group division is intended to see the response of each concentration group. It is based on observations with the naked eye that the color change response appears random when the test strip is dipped in samples with concentrations ranging from 700 mg/dl.

Artificial urine samples were made to vary with an increase of every 50 mg/dl in the concentration range of 0–700 mg/dl. In the glucose concentration range of 700–2000 mg/dl, an artificial urine sample was prepared with an increase in every 100 mg/dl of variation. There was a total of 70 sample variations for both smartphones. In addition to artificial urine samples, as many as 20 urine samples from respondents from Tk. 1 Said Sukanto Hospital, Jakarta, Indonesia, were used for the validation process.

TABLE I
URINE ARTIFICIAL SAMPLE

Concentration	Amount of data
0–700 mg/dl	1,200
701–2000 mg/dl	1,112

After all the samples were prepared, the next step was to dip the urine test strip into the sample solution. Three test strips were dipped into each artificial urine sample and respondent's urine. A total of 2,312 data sets were used as a regression model for both smartphones (see TABLE I for details).

B. Image acquisition and Pre-processing

The initial stage in image processing was acquiring sample test strips using cameras from two types of smartphones: the Samsung Galaxy A51 (SM51) and Huawei Nova 5T (HN5T). Image results were obtained based on the maximum resolution setting at 4,000 × 3,000 pixels on both smartphones with the standard autofocus setting mode. After the test strip showed a stable color change, the image acquisition process was carried out by the reference time response rules on the included urine test strip datasheet, which was 30 seconds. Fig. 7 shows the image acquisition results on the HN5T and SM51 smartphones.

(a) (b)

Fig. 7 Image result was taken using a smartphone camera: (a) HN5T and (b) SM51

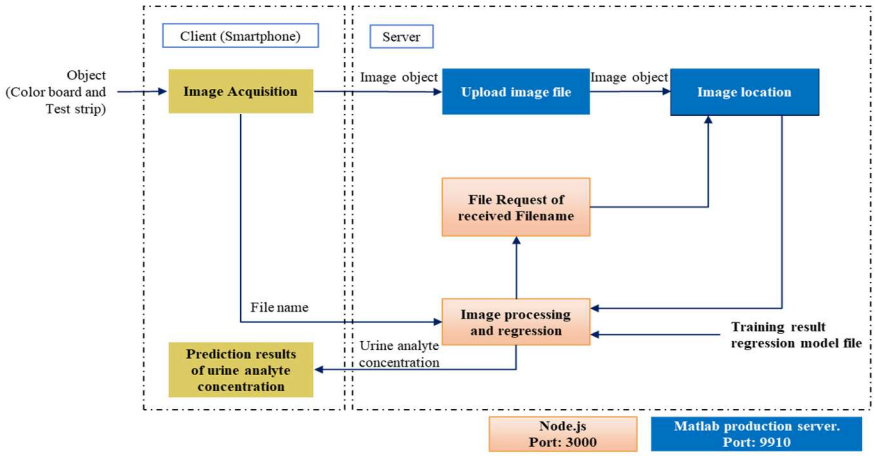
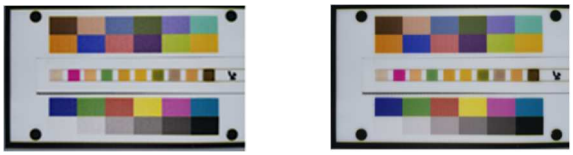


Fig. 8 Block diagram of implementation system

The filtering stage with a Gaussian filter produced a blurry effect on the image to reduce noise in the input image. The value of $\sigma = 2$ on the applied Gaussian filter successfully processed the image. Fig. 9 shows the result of the image before and after being given a Gaussian filter. The next step was image segmentation.



(a) (b)

Fig. 9 The image was taken using a smartphone camera: (a) before being given a Gaussian filter and (b) after being given a Gaussian filter

The initial stage of segmentation was done by removing the background limited by black areas so that only the color board and urine test strips were obtained. The next stage was to take the area under four black circles to obtain the ROI used in the next stage, where the masking process implemented on the image produced ROI, which was used as the research object. At this segmentation stage, 34 ROIs were studied, consisting of 24 ROI color boards with a size of 96×60 pixels and 10 ROI test strips sized 41×41 pixels dyed to each of the prepared analytes. Fig. 10 shows each stage of the segmentation processes carried out.

TABLE II
AVERAGE OF EUCLIDEAN DISTANCE (ΔE)

Concentration	Before correction	After correction
HN5T	4.01	1.37
SM51	5.34	1.39

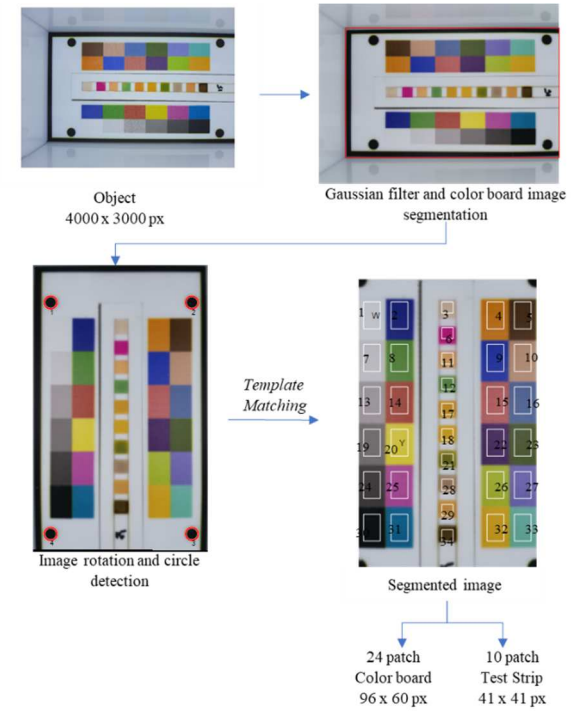


Fig. 10 Image segmentation process

After the image was successfully segmented, the next step was image correction. Fig. 11 shows the image obtained after image correction on both the HN5T and SM51 smartphones. Color correction was done by comparing the Euclidean distance values in the image before and after correction using the RPCC method. The results of image correction on both mobile phones are shown in TABLE II. The average result of ΔE obtained after the correction was in the value of 1.37-1.39 or ΔE <3. The ΔE proved that the image color correction algorithm successfully produced the best performance value in the urine glucose prediction system.

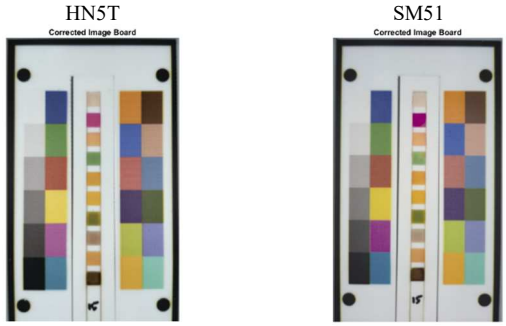


Fig. 11 Image correction

C. Prediction Result

Prediction of urine glucose concentration was made by comparing the regression model with the Decision Tree and Ensemble methods with bagging and boosting optimization. All datasets obtained from the HN5T and SM51 smartphones were then carried out by training and testing the prediction system with datasets in the low range (0–700 mg/dl), high range (701–2000 mg/dl), and combined concentrations (0–2000 mg/dl). Furthermore, the validation process was carried out on the most optimal regression model using the respondent's urine dataset. Five-fold cross-validation was applied to the system, while the system was evaluated using the value of Root Means Square Error (RMSE), Relative Root Means Square Error (RRMSE), and the Correlation Coefficient (R^2).

1) Regression model at low concentration (0–700 mg/dl): The error value for the three training models was quite good at 0.01 in the regression model with decision tree and Ensemble-Boosting. The relative error value when using Ensemble-Bagging produced a value of 0.03 in the training data. In addition, the R^2 value was also excellent in the three regression models used. The RMSE value was also quite good when using the Decision Tree model and Ensemble-boosting: 8.74 and 8.16 in the training data, respectively, while the RMSE in testing was 20.45 and 21.85, respectively. The cross-validation at low concentrations for each regression model can be seen in TABLE III.

TABLE III
CROSS VALIDATION RESULTS FOR CONCENTRATION 0–700 MG/DL

Model Reg	RRMSE Tr	RRMSE Tt	RMSE Tr	RMSE Tt	R^2 Tr	R^2 Tt
DT	0.01	0.03	8.74	20.45	1.00	0.99
En - Bag	0.03	0.04	19.36	27.18	1.00	0.99
En - Boost	0.01	0.03	8.16	21.85	1.00	0.99

2) Regression model at high concentration (700–2000 mg/dl): Different results are shown in the regression model on sample data at high concentration. The error value in the Decision Tree and Ensemble-Bagging testing data is quite large at 0.12. The relatively high RMSE testing values reinforce this for both Decision Tree and Ensemble-Bagging regressors, at 125.97 and 134.17, respectively.

TABLE IV
CROSS VALIDATION RESULTS FOR CONCENTRATION 701–2000 MG/DL

Model Reg	RRMSE Tr	RRMSE Tt	RMSE Tr	RMSE Tt	R^2 Tr	R^2 Tt
DT	0.05	0.12	52.35	125.97	0.99	0.92
En - Bag	0.08	0.12	90.32	134.17	0.96	0.91
En - Boost	0.01	0.03	8.16	21.85	1.00	0.99

In addition, the value of R^2 looks quite low, which is 0.92 for the Decision Tree and 0.91 for the Ensemble-Bagging. The distribution of data points in samples with high concentrations shows a wide distribution. There are many data points located quite far from the main diagonal. When applied using boosting optimization at high concentrations, the Ensemble regression model obtained a minimal error of 0.03 in the testing data. The RMSE value was also small at 21.85, resulting in an optimal R^2 value of 0.99

3) Regression model on combined concentration (0–2000 mg/dl): The combined data from low and high concentrations showed good results for the three types of regressors used. The minimum error value was obtained using the regressor Ensemble with a boosting optimization model of 0.04 on the training data. The RMSE value was also small at 52.94 and 79.70 on the training and testing data. It can be said that the regression model using an ensemble with boosting optimization can provide the best value in the urine glucose prediction system.

The response graph shows that the data points spread from the main diagonal line using the Decision Tree, but these are different when the regression model uses regressor Ensemble bagging and boosting (see Fig. 12). The data points appear closer to the main diagonal line, indicating that the predicted and target values correlate.

TABLE V
CROSS VALIDATION RESULTS FOR CONCENTRATION 0–2000 MG/DL

Model Reg	RRMSE Tr	RRMSE Tt	RMSE Tr	RMSE Tt	R^2 Tr	R^2 Tt
DT	0.02	0.05	29.71	91.75	1.00	0.98
En - Bag	0.03	0.05	52.94	79.70	0.99	0.98
En - Boost	0.02	0.04	27.64	75.19	1.00	0.98

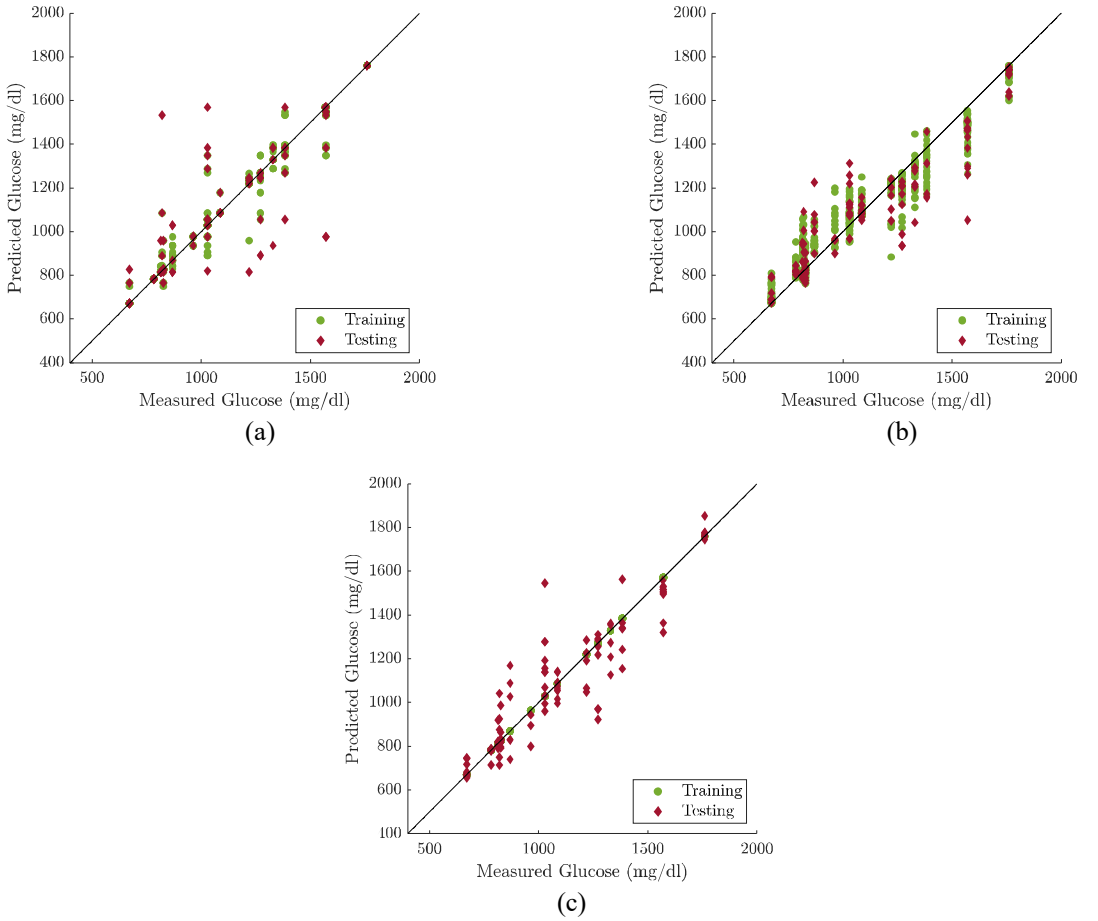


Fig. 12 The graph of the response of the measurement value to the predicted value of glucose in the combined sample (HN5T and SM51) at a concentration of 701–2000 mg/dL: (a) DT, (b) En-Bag dan, and (c) En-Boost

When all stages of the process on the artificial urine sample were completed, the next step was to enter the validation process, which was carried out using the respondent's urine sample at all low and high concentrations. In TABLE VI, the

most significant error value when using the Decision Tree is 0.08. The predictive value close to the target value is also expressed by the R^2 value of 0.96. The comparison graph of all modeling results at low, high, and a combination of both concentrations can be seen in Fig. 13.

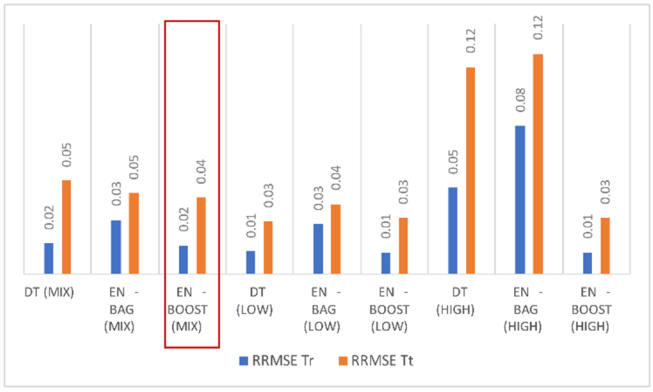


Fig. 13 Comparison of modelling results

The validation value was quite good when using a regressor Ensemble with bagging optimization, which produced an error value of 0.06 and an RMSE of 90.13. The best validation value with an error value of 0.06 and RMSE 88.18 used a regressor Ensemble with boosting optimization, which also produced a reasonably optimal R^2 value of 0.98. Based on these data, the Ensemble–Boosting learning model showed the best results so that the model was chosen to be applied to the implementation system.

TABLE VI
VALIDATION RESULTS OF URINE RESPONDENTS IN CONCENTRATION OF 0–2000 MG/DL

Model Reg	RRMSE	RMSE	R^2
DT	0.08	130.22	0.96
En – Bag	0.06	90.13	0.97
En – Boost	0.06	88.18	0.97

D. System Implementation Result

There are three buttons on the application interface: a button to select the image, upload, and result. The successful application chose whether the prediction takes the image directly using the camera or from the photo gallery. Image capture using the camera can use the default camera application from the smartphone or third-party applications. In this case, we took an image using the Open Camera application according to the training data that was previously fed to the model.

The next stage was to upload the image to the application to run on the upload and computing servers. If the image was already available on the preview screen, the user could directly press the upload button. This upload process could be monitored in real-time with the help of the progress bar in the application. When the progress bar reached 100%, the image had been successfully uploaded, and the user could immediately press the result button. However, if the progress bar had not reached 100% and keep pressing the result button, the "Upload Image First" notification appeared.

When the computing server successfully processed the image sent, a text display appeared on the interface layer in urine glucose prediction results from the processing carried out by the computing server. Additionally, a "Data Error" notification appears when the app fails to process the rendered image.

Fig. 14 Glucose prediction application

IV. CONCLUSION

This research developed a system for predicting the glucose concentrations in urine at 0–2000 mg/dl. The regression model was made using Decision Tree, Ensemble–bagging, and Ensemble–boosting with RGB color space as its features. Based on the third model used, Ensemble–boosting showed the best performance at R^2 and RRMSE at 0.98 and 0.04, respectively. Further research made it possible to develop a system without housing to make the equipment more portable. In addition, another development was the analysis of 10 analytes, so that urine analyte prediction was not only for one analyte, and the application could be widely developed for use on all types of smartphones in the future.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support from the Universitas Indonesia on International Indexed Publications (PUTI) SAINTEKES No. NKB-2394/UN2.RST/HKP.05.00/2020. We would like to thank Tk. 1 Said Sukanto Hospital, Jakarta, Indonesia, for providing urine samples.

REFERENCES

[1] A. Listyarini, W. Sholihah, and C. Imawan, "A paper-based Colorimetric Indicator Label using Natural Dye for Monitoring Shrimp Spoilage," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 367, no. 1. doi: 10.1088/1757-899X/367/1/012045. url: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85049373519&doi=10.1088%2F1757-899X%2F367%2F1%2F012045&partnerID=40&md5=21cf99e920aca2161e81b91cd87ff65e.

[2] S. Singh, K. K. Gaikwad, M. Lee, and Y. S. Lee, "Temperature sensitive smart packaging for monitoring the shelf life of fresh beef," 2018, doi: 10.1016/j.jfoodeng.2018.04.014. url: https://doi.org/10.1016/j.jfoodeng.2018.04.014.

[3] I. Hussain, M. Das, K. U. Ahamad, and P. Nath, "Water salinity detection using a smartphone," *Sensors Actuators, B Chem.*, vol. 239, pp. 1042–1050, 2017, doi: 10.1016/j.snb.2016.08.102. url: http://dx.doi.org/10.1016/j.snb.2016.08.102.

[4] J.-C. Yan, J. Ren, L.-L. Ren, Y. Yang, S.-F. Yang, and T.-L. Ren, "A novel structure design and fabrication method for low liquid consumption and high precision device of colorimeter in water quality detection," *Sensors Actuators A*, vol. 289, pp. 1–10, 2019, doi: 10.1016/j.sna.2019.02.016. url: https://doi.org/10.1016/j.sna.2019.02.016.

[5] J.-C. Yan *et al.*, "Development of a portable setup using a miniaturized and high precision colorimeter for the estimation of phosphate in natural water," 2019, doi: 10.1016/j.aca.2019.01.030. url: https://doi.org/10.1016/j.aca.2019.01.030.

[6] E. H. K. Alkamil, S. Al-Dabooni, A. K. Abbas, R. Flori, and D. C. Wunsch, "Learning from experience: An automatic pH neutralization system using hybrid fuzzy system and neural network," *Procedia Comput. Sci.*, vol. 140, pp. 206–215, 2018, doi: 10.1016/j.procs.2018.10.330. url: https://doi.org/10.1016/j.procs.2018.10.330.

[7] N. Asthana and R. Bahl, "IoT Device for Sewage Gas Monitoring and Alert System," *Proc. 1st Int. Conf. Innov. Inf. Commun. Technol. ICIICT 2019*, 2019, doi: 10.1109/ICIICT1.2019.8741423.

[8] M. Carrozzo *et al.*, "UAV intelligent chemical multisensor payload for networked and impromptu gas monitoring tasks," *5th IEEE Int. Work. Metrol. AeroSpace, Metroaerosp. 2018 - Proc.*, pp. 112–116, 2018, doi: 10.1109/MetroAeroSpace.2018.8453543.

[9] C. Liu, F. A. Gomez, Y. Miao, P. Cui, and W. Lee, "A colorimetric assay system for dopamine using microfluidic paper-based analytical devices," *Talanta*, vol. 194. pp. 171–176, 2019. doi: 10.1016/j.talanta.2018.10.039.

[10] K. R. Mallires, D. Wang, P. Wiktor, and N. Tao, "A Microdroplet-Based Colorimetric Sensing Platform on a CMOS Imager Chip," *Anal. Chem.*, vol. 92, no. 13, pp. 9362–9369, Jul. 2020, doi: 10.1021/acs.analchem.0c01751. url: https://doi.org/10.1021/acs.analchem.0c01751.

[11] T. Akyazi, L. Basabe-Desmonts, and F. Benito-Lopez, "Review on microfluidic paper-based analytical devices towards commercialisation," *Anal. Chim. Acta*, vol. 1001, pp. 1–17, Feb. 2018, doi: 10.1016/J.ACA.2017.11.010.

[12] L. M. Fu and Y. N. Wang, "Detection methods and applications of microfluidic paper-based analytical devices," *TrAC - Trends in Analytical Chemistry*, vol. 107. pp. 196–211, 2018. doi: 10.1016/j.trac.2018.08.018.

[13] M. I. G. S. Almeida, B. M. Jayawardane, S. D. Kolev, and I. D. McKelvie, "Developments of microfluidic paper-based analytical devices (μPADs) for water analysis: A review," *Talanta*, vol. 177, pp. 176–190, Jan. 2018, doi: 10.1016/J.TALANTA.2017.08.072.

[14] G. M. Fernandes *et al.*, "Novel approaches for colorimetric measurements in analytical chemistry – A review," *Analytica Chimica Acta*, vol. 1135. pp. 187–203, 2020. doi: 10.1016/j.aca.2020.07.030.

[15] J. Wu, M. Dong, C. Rigatto, Y. Liu, and F. Lin, "Lab-on-chip technology for chronic disease diagnosis," *npj Digit. Med.*, vol. 1, no. 1, pp. 1–11, 2018, doi: 10.1038/s41746-017-0014-0.

[16] M. Salve, A. Wadafale, G. Dindorkar, and J. Kalambe, "Quantifying colorimetric assays in μPAD for milk adulterants detection using colorimetric android application," *Micro Nano Lett.*, vol. 13, no. 11, pp. 1520–1524, 2018, doi: 10.1049/mnl.2018.5334.

[17] M. Ra, M. S. Muhammad, C. Lim, S. Han, C. Jung, and W. Y. Kim, "Smartphone-Based Point-of-Care Urinalysis under Variable Illumination," *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, no. December 2017, pp. 1–11, 2018, doi: 10.1109/JTEHM.2017.2765631.

[18] M. Oyaert and J. R. Delanghe, "Semiquantitative, fully automated urine test strip analysis," *J. Clin. Lab. Anal.*, vol. 33, no. 5, pp. 1–7, 2019, doi: 10.1002/jcla.22870.

[19] M. Anthimopoulos, S. Gupta, S. Arampatzis, and S. Mougiakakou, "Smartphone-based urine strip analysis," *IST 2016 - 2016 IEEE Int. Conf. Imaging Syst. Tech. Proc.*, pp. 368–372, 2016, doi: 10.1109/IST.2016.7738253.

[20] M. Jalal Uddin, G. J. Jin, and J. S. Shim, "Paper-Plastic Hybrid Microfluidic Device for Smartphone-Based Colorimetric Analysis of Urine," *Anal. Chem.*, vol. 89, no. 24, pp. 13160–13166, 2017, doi: 10.1021/acs.analchem.7b02612.

[21] J. Guo, X. Huang, and X. Ma, "Clinical identification of diabetic ketosis/diabetic ketoacidosis acid by electrochemical dual channel test strip with medical smartphone," *Sensors Actuators, B Chem.*, vol. 275, no. August, pp. 446–450, 2018, doi: 10.1016/j.snb.2018.08.042. url: https://doi.org/10.1016/j.snb.2018.08.042.

[22] S. Schaefer, "Colorimetric water quality sensing with mobile smart phones," no. April, p. 88, 2014, doi: 10.14288/1.0074338.

[23] M. Spaanderman, R. L. Smeets, P. J. F. Lucas, M. Velikova, and J. T. van Scheltinga, "Smartphone-based analysis of biochemical tests for health monitoring support at home," *Healthc. Technol. Lett.*, vol. 1, no. 3, pp. 92–97, 2014, doi: 10.1049/htl.2014.0059.

[24] J. Doupis, G. Festas, C. Tsilivigos, V. Efthymiou, and A. Kokkinos, "Smartphone-Based Technology in Diabetes Management," *Diabetes Ther.*, vol. 11, no. 3, pp. 607–619, 2020, doi: 10.1007/s13300-020-00768-3. url: https://doi.org/10.1007/s13300-020-00768-3.

[25] World Health Organization. (2016). Global report on diabetes. World Health Organization. https://apps.who.int/iris/handle/10665/204871.

[26] Infodatin, "Hari Diabetes Sedunia Tahun 2018," *Pus. Data dan Inf. Kementrian Kesehat. RI*, pp. 1–8, 2019.

[27] "Konica minolta." https://sensing.konicaminolta.asia/color-measurement/ (accessed Jul. 21, 2020). url: https://sensing.konicaminolta.asia/color-measurement/.

[28] "clinitek atlas auto urine chem analyzer rack." https://www.siemens-healthineers.com/it/urinalysis/systems/clinitek-atlas-auto-urine-chem-analyzer-rack (accessed Feb. 16, 2020). url: https://www.siemens-healthineers.com/it/urinalysis/systems/clinitek-atlas-auto-urine-chem-analyzer-rack.

[29] J. Il Hong and B. Y. Chang, "Development of the smartphone-based colorimetry for multi-analyte sensing arrays," *Lab Chip*, vol. 14, no. 10, pp. 1725–1732, 2014, doi: 10.1039/c3lc51451j.

[30] P. C. Lin *et al.*, "A machine learning approach for predicting urine output after fluid administration," *Computer Methods and Programs in Biomedicine*, vol. 177. pp. 155–159, 2019. doi: 10.1016/j.cmpb.2019.05.009.

[31] A. A. H. Gadalla *et al.*, "Identification of clinical and urine biomarkers for uncomplicated urinary tract infection using machine learning algorithms," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, 2019, doi: 10.1038/s41598-019-55523-x.

[32] A. Y. Mutlu, V. Kiliç, G. K. Özdemir, A. Bayram, N. Horzum, and M. E. Solmaz, "Smartphone-based colorimetric detection: Via machine learning," *Analyst*, vol. 142, no. 13, pp. 2434–2441, 2017, doi: 10.1039/c7an00741h.

[33] D. Wulan Hastuti, M. Harahap, F. Adila Ferdiansyah, A. Harmoko Saputro, and C. Imawan, "Prediction system for pH measurement on Brassica oleraceae (Red Cabbage) using machine learning regression," *J. Phys. Conf. Ser.*, vol. 1528, no. 1, 2020, doi: 10.1088/1742-6596/1528/1/012050.

[34] K. Rathan, S. V Sai, and T. Manikanta, "Crypto-Currency price prediction using Decision Tree and Regression techniques," *2019 3rd Int. Conf. Trends Electron. Informatics*, pp. 190–194, 2019.

[35] A. M. Shabut *et al.*, "An intelligent mobile-enabled expert system for tuberculosis disease diagnosis in real time," *Expert Syst. Appl.*, vol. 114, pp. 65–77, 2018, doi: 10.1016/j.eswa.2018.07.014. url: https://doi.org/10.1016/j.eswa.2018.07.014.

[36] D. O. Oyewola, E. G. Dada, O. T. Omotehinwa, and I. A. Ibrahim, "Comparative Analysis of Linear, Non Linear and Ensemble Machine Learning Comparative Analysis of Linear , Non Linear and Ensemble Machine Learning Algorithms for Credit Worthiness of Consumers," no. September, 2019.

[37] S. Yang, J. Wu, Y. Du, Y. He, and X. Chen, "Ensemble Learning for Short-Term Traffic Prediction Based on Gradient Boosting Machine," *J. Sensors*, vol. 2017, 2017, doi: 10.1155/2017/7074143.

[38] A. Mosavi, F. Sajedi Hosseini, B. Choubin, M. Goodarzi, A. A. Dineva, and E. Rafiei Sardooi, "Ensemble Boosting and Bagging Based Machine Learning Models for Groundwater Potential Prediction,"

Water Resour. Manag., vol. 35, no. 1, pp. 23–37, 2021, doi: 10.1007/s11269-020-02704-3.

[39] Y. Lu, X. Li, Z. Gong, L. Zhuo, and H. Zhang, "TDCCN: A two-phase deep color correction network for Traditional Chinese Medicine tongue images," *Appl. Sci.*, vol. 10, no. 5, pp. 1–21, 2020, doi: 10.3390/app10051784.

[40] P. D. Marrero Fernández, F. A. Guerrero Peña, T. Ing Ren, and J. J. G. Leandro, "Fast and robust multiple ColorChecker detection using deep convolutional neural networks," *Image Vis. Comput.*, vol. 81, pp. 15–24, Jan. 2019, doi: 10.1016/J.IMAVIS.2018.11.001.

[41] N. Promphet *et al.*, "Non-invasive textile based colorimetric sensor for the simultaneous detection of sweat pH and lactate," *Talanta*, vol. 192, no. September 2018, pp. 424–430, 2019, doi: 10.1016/j.talanta.2018.09.086. url: https://doi.org/10.1016/j.talanta.2018.09.086.

[42] P. T. Wang, J. J. Chou, and C. W. Tseng, "Colorimetric characterization of color image sensors based on convolutional neural network modeling," *Sensors Mater.*, vol. 31, no. 5, pp. 1513–1522, 2019, doi: 10.18494/SAM.2019.2271.

[43] S. Perumal and T. Velmurugan, "Preprocessing by Contrast Enhancement Techniques for Medical Images," *Int. J. Pure Appl. Math.*, vol. 118, no. 18, pp. 3681–3688, 2018.

[44] C. S. Wang *et al.*, "Development of a novel mobile application to detect urine protein for nephrotic syndrome disease monitoring," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, pp. 1–8, 2019, doi: 10.1186/s12911-019-0822-z.