Original

Verification of individual differences in visual judgment of urine test strips and proposal of objective evaluation method

†Keigo Inagaki*[1], Katsumasa Nakamura*[2]

†Correspondence: Cooperative Major in Medical Photonics, Hamamatsu University School of Medicine, Shizuoka, Japan.
Institute for Medical Photonics Research, Hamamatsu University School of Medicine, 1-20-1 Handayama, Higashi-ku, Hamamatsu City, Shizuoka, Japan.
E-mail: kinagaki@hama-med.ac.jp
Received August 10, 2023; accepted March 25, 2024
*[1] Institute for Medical Photonics Research, Hamamatsu University School of Medicine, Shizuoka, Japan
*[2] Faculty of Medicine, Hamamatsu University School of Medicine, Shizuoka, Japan

ABSTRACT

Objective: Practitioners at small-scale medical check-up facilities lacking automatic urine analyzers visually analyze and interpret urine test strips. However, visual judgment is liable to human errors in estimating parameters and variations in judgment time. Therefore, to objectively judge the color of urine test strips, we developed an automated urine test strip colorimetric program using images taken with a smartphone camera.
Materials and Methods: For urinalysis, 40 urine samples were randomly selected, and six nurses were included as evaluators. The effectiveness of the program was confirmed by determining the differences in visual judgment among evaluators. A color chart was used as a judgment reference table for the comparison of visual judgment of urinalysis test strips. Two automatic urine analyzers, US-3500 (Eiken Chemical Co., Ltd.) and LABOSPECT 006 (Hitachi High-Tech Co., Ltd.), were used for the comparative evaluation.
Results: The study showed slight inter-rater differences in the evaluation of most parameters, including protein, glucose, ketone, specific gravity, occult blood, leukocytes, and nitrite; however, individual differences were observed for urobilinogen, bilirubin, pH, creatinine, and albumin. Moreover, we compared visual judgment to the new program and observed that the agreement rate of the automated urine test strip colorimetric program (81.9%) was higher than that of visual judgment (75.1%).
Conclusion: Based on our findings, we propose an inexpensive and simple method for the stable evaluation of urine test strips for urinalysis that is unaffected by human error during visual judgment.

〔Lab Med Int 2024; 3(2): 42-49〕

Key Words
urine test strip, visual judgment, objective evaluation, colorimetry

I. Introduction

Urinalysis is one of the most frequently performed tests in healthcare settings; it is inexpensive and used to detect risk signals for key metabolic functions using urine samples [1]. Test papers are used in a simple urine test. Because this method is inexpensive and non-invasive, it has been widely used as an effective tool for diagnosing several diseases, including urinary tract infections, renal disorders, cardiovascular diseases, and metabolic dysfunctions, such as diabetes and liver diseases [2) 3)]. In general hospitals, automatic urine analyzers are introduced to evaluate urine test strips, thereby avoiding human errors that may arise due to the systemization of examination rooms, especially in such large facilities. However, introducing an expensive automatic urine analyzer is not feasible for general practitioners at small-scale examination facilities. Therefore, visual judgment is performed in such facilities. However, the judgment time of each parameter differs in urinalysis using test strips, and determining all parameters within the set time is difficult. In addition, color evaluation is affected by environmental factors, such as the color and intensity of illumination during observation, making accurate determination challenging. In a previous study, Winkens et al. observed a lack of agreement between interobserver and intraobserver urinalysis results using multiple test strips [4]. In contrast, Bekhof et al. reported that interobserver agreement was highly effective, and the interobserver scores of urinalysis test strips did not deviate for multiple categories [5]. In addition, a previous study in Japan reported that the subjective judgments of nurses resulted in differences in color expression and opinions [6]. Determining the color of the test strip for urinalysis provides important information regarding the early detection of abnormalities; therefore, performing an accurate evaluation is necessary. In particular, inexperienced nurses may find it difficult to accurately judge and communicate color information [7]. In addition, the number of male nurses has increased in recent years, and 1 in 20 men have congenital color blindness [8].

Numerous previous reports have focused on the objective evaluation of urinalysis test strips using smartphones [9)–11]. Studies on judgments by machine learning based on hue have been performed thus far, but few have focused on color differences in judging urinalysis test strips. If the strips can be judged solely based on color difference, machine learning may not be required. In addition, regardless of the manufacturer of the test paper, the standard color chart data for urinalysis test papers can be used for evaluation. In addition to smartphones, urine test strips have been evaluated using scanners [12], and ongoing research is investigating methods that can be consistently and universally used by all practitioners. However, to date, only a few, inadequate validated studies exist on the simultaneous evaluation of interindividual visual assessment differences and program effectiveness using urine specimens from patients. In addition, there is no consensus regarding the presence or absence of individual differences in visual judgment.

The present study aimed to address these issues by developing an automated urine test strip colorimetric program for automatic analysis of urine test strips from images taken with a smartphone [13]. This study included six nurses to investigate the individual differences in visual judgment of urinalysis test strips. Using the results of an automatic urine analyzer as the gold standard, we investigated the rate of agreement between the visual judgment of urine test strips and the automatic urine test strip colorimetric program.

II. Materials and methods

Figure 1 demonstrates the experimental protocol used in this study. A photography box (PU5025B: PULUZ) was used to provide constant lighting. The smartphone used was the Xperia 5II (XQ-AS42: SONY) equipped with the standard lens of a triple-lens camera (ZEISS lens). Other parameters were as follows: sensor size, 1/1.7-inch large format; number of effective pixels, approximately 12.2 million; focal length, 24 mm; F value, 1.7; dual photodiode; and hybrid image stabilization (optical + electronic). The urinalysis test strip (Uropaper® III 'Eiken' (E-UR97): Eiken Chemical Co. Ltd.) used was capable of testing 12 parameters. The color chart shown in **Figure 1**, which is commonly used in clinical practice, was used as a judgment reference table for comparison of visual judgment of urinalysis test strips. For urinalysis, 40 samples were randomly selected from the laboratory

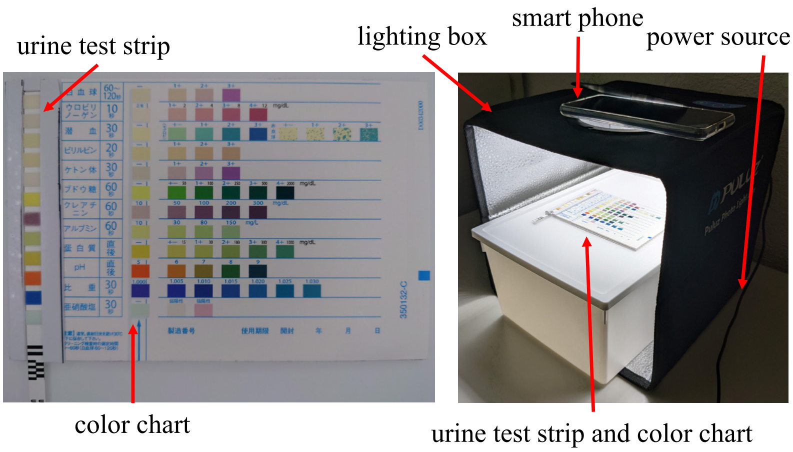

Figure 1 Experiment apparatus

departments of hospitals in H city. Basic information on the specimens is provided in **Table 1**. Two automatic urine analyzers, US-3500 (Eiken Chemical Co., Ltd.) and LABOSPECT 006 (Hitachi High-Tech Co., Ltd.), were used for comparative evaluation. US-3500 was used to compare urobilinogen, occult blood, protein, glucose, ketone bodies, bilirubin, nitrite, specific gravity, white blood cells, and pH. LABOSPECT 006 was used to compare creatine and albumin levels.

Table 1 Basic information on urine specimens

Contents	Number of people (%)
Sex	
Female	21 (52.5)
Male	19 (47.5)
Average age, years	
Female	45.86 ± 18.2
Male	66.89 ± 15.7
Clinical department	
General surgery	1 (2.5)
Lower gastrointestinal surgery	1 (2.5)
Liver internal medicine	1 (2.5)
Ophthalmology	1 (2.5)
Hematology	1 (2.5)
Vascular surgery	1 (2.5)
Obstetrics and gynecology	8 (20.0)
Cardiology	1 (2.5)
Pediatrics	1 (2.5)
Upper gastrointestinal surgery	3 (7.5)
Nephrology	3 (7.5)
Orthopedics	5 (12.5)
Endocrine	2 (5.0)
Urology	10 (25.0)
Immunology	1 (2.5)

We acquired data on urobilinogen, occult blood, protein, glucose, ketone bodies, bilirubin, nitrite, specific gravity, white blood cells, and pH for all 40 samples using automatic analyzers. However, we could only obtain data on creatine levels from five samples and albumin from one sample using automatic analyzers. The automatic urine test strip colorimetric program developed in this study was created on a desktop computer (HP EliteDesk 800 G4 TWR: HP). The development language was Python, and the OpenCV library was used for image processing.

Experimental methods

Visual judgment

Visual judgment was performed by six individuals (four male and two female individuals) who had no eye disease and had nursing qualifications. During the visual judgment of samples, the individuals were adequately spaced to prevent any potential interference with each other's results. The specific procedure followed during judgment is detailed below:

(1) Turn on the shooting box (leave for 30 min for light stabilization).

(2) Enter the patient ID on the entry form.

(3) Start the survey after 30 min of turning on the power.

(4) Soak urinalysis test strips completely in urine samples for approximately 1–2 s.

(5) Gently tap the side of the test paper onto a tissue paper to remove excess urine.

(6) Place the color chart and test paper into the photography box.

(7) Within 30–60 s of soaking the sample in urine and removing it from the spitz, determine the closest color

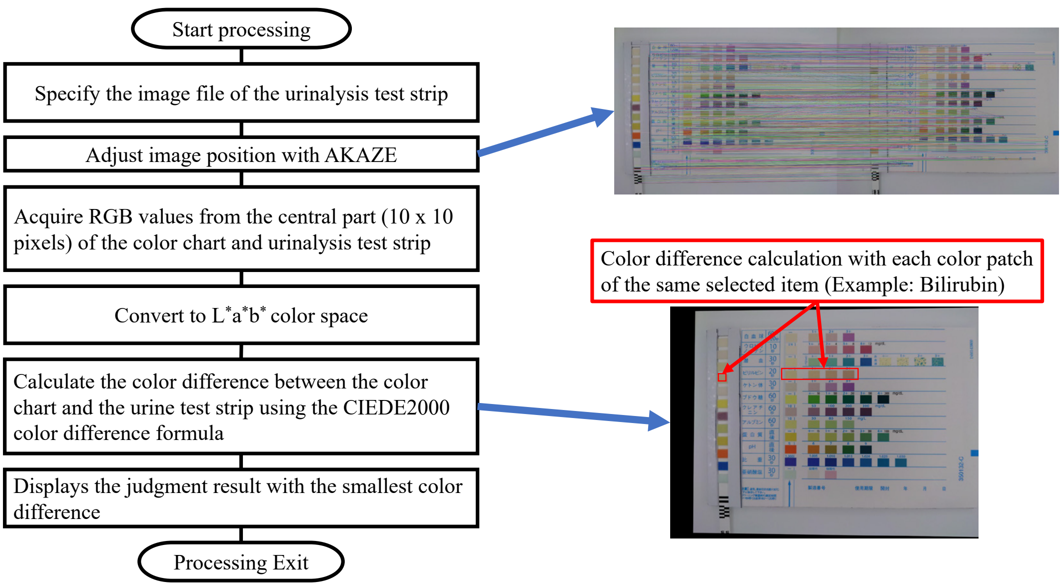

Figure 2 Processing flow

on the color chart and check it on the entry form for visual judgment.
(8) Repeat steps (4) to (7) for all samples.

Judgment by an automated urine test strip colorimetric program

・Automated urine test strip colorimetric program

Figure 2 demonstrates the process flow of an automated urine test strip colorimetric program. The colors of the urinalysis test papers and those of the color chart were compared as per the following procedure, and each parameter was judged accordingly.
(1) Specify the image file of the urinalysis test strip.
(2) Adjust the position of the image with AKAZE.
(3) Acquire RGB values from the central part (10 × 10 pixels) of the color chart and urinalysis test strip.
(4) Convert to L*a*b* color space.
(5) Calculate the color difference between the color chart and urine test strip using the CIEDE2000 color difference formula (color difference calculation with each color patch of the same selected item).
(6) Record the judgment result with the smallest color difference.

・Filming environment and procedures

Similar to visual evaluation, color evaluation was performed using the photography box (PU5025B: PULUZ) under the same environment.
(1) Turn on the shooting box (leave for 30 min for light stabilization).
(2) Start shooting after 30 min of turning on the power.
(3) Install the smartphone with an activated camera on the top of the shooting box (set the color temperature of the camera to sunlight, which is the same as the color temperature of lighting inside the shooting box).
(4) Soak urine test strips completely in urine for approximately 1–2 s.
(5) Gently tap the side of the test paper onto a tissue paper to remove excess urine.
(6) Place the test paper in the shooting box.
(7) Take a picture of the sample 50 s after removing it from the spitz.
(8) Judge the photographed images using the automatic urine dipstick colorimetric program.
(9) Obtain data regarding age, sex, clinical department, urinalysis results, and oral medication status (presence or absence of sodium-glucose co-transporter 2 inhibitors) from electronic medical records.
(10) Confirm whether the results match those obtained from the automatic urine analyzer.

Statistical analysis

The target sample size for intraclass correlation coefficient (ICC) (2.1) was set at 40 with reference to a study by Doros et al [14]. (assuming six examiners, an ICC of 0.8, a significance level of 0.05, and a confidence interval of 0.2, the required number of cases is 32). ICC (2.1) was used for inter-rater reliability. The statistical software IBM SPSS Statistics 25.0 (IBM Corp., Armonk, NY, USA) was used for ICC calculations. The rate of agreement was calculated as the rate of agreement with the results obtained from the automatic urine analyzer (number of matched urine specimens/number of specimens investigated).

Ethical considerations

This research was reviewed and approved by the ethics committee of Hamamatsu University School of Medicine (approval number: 21-239).

III. Results

Table 2 lists the ICC (2.1) for each parameter of urinalysis using test strips. The reference ICC criteria reported by Landis et al. were used [15]. Glucose, occult blood, leukocytes, and nitrite were within the Almost Perfect criteria. Protein, ketone, specific gravity, and albumin were within the Substantial criteria. Urobilinogen, pH, and creatinine were within the Moderate criteria. Bilirubin was within the Fair criteria. Cronbach's alpha was 0.8 or higher for all parameters.

Table 3 shows the comparison of the results of the automatic urine analyzer, visual judgment, and automated urine test strip colorimetric program. The values in the table indicate the concordance rate of judgment results with that of the automatic urine analyzer; A–F show the results of visual judgment by the six examiners, examin-

Table 2 Intraclass correlation coefficient of inter-examiner reliability

Parameter	ICC (2.1) (95% CI)	Cronbach's alpha	p-value
Protein	0.783 (0.690–0.863)	0.959	<0.001
Glucose	0.941 (0.909–0.965)	0.990	<0.001
Urobilinogen	0.422 (0.289–0.575)	0.821	<0.001
Bilirubin	0.377 (0.246–0.533)	0.801	<0.001
Ketone	0.794 (0.705–0.870)	0.958	<0.001
Specific gravity	0.742 (0.616–0.842)	0.958	<0.001
Occult blood	0.864 (0.799–0.916)	0.976	<0.001
pH	0.529 (0.347–0.692)	0.916	<0.001
Leukocytes	0.883 (0.825–0.928)	0.979	<0.001
Nitrite	0.907 (0.860–0.943)	0.983	<0.001
Creatinine	0.516 (0.349–0.675)	0.904	<0.001
Albumin	0.661 (0.538–0.776)	0.930	<0.001

ICC, intraclass correlation coefficient; CI, confidence interval er avg. indicates the average of the A–F results, and the program indicates the results from the automatic urine test strip colorimetric program. Urobilinogen, occult blood, protein, glucose, ketone bodies, bilirubin, nitrite, specific gravity, leukocytes, and pH were evaluated for concordance across 40 samples (N = 40). Data for creatinine and albumin could not be obtained from the automated urine analyzer due to the small number of test requests per day. Therefore, the concordance rate was evaluated using data obtained for five specimens (N = 5) of creatine and one specimen (N = 1) of albumin from the automatic urine analyzer.

A comparison of the results of the average visual judgment (Examiner avg.) and the automated urine test strip colorimetric program (Program) revealed that the agreement rate of judgment by the colorimetric program was high. Moreover, the average match rate of all 12 parameters was 75.1% for visual judgment and 81.9% for the automated urine test strip colorimetric program. Concordance was better using the automated urine test strip colorimetric program. The lowest agreement rate of judgment results of specific gravity was observed with the automated urine analyzer for both visual determination and the automated urine test strip colorimetric program.

Table 4 shows the agreement rate of ± 1 rank with the automatic urine analyzer. The average ± 1 rank agreement rate for the 12 items was 92.8% for the examiners and 92.9% for the program.

IV. Discussion

In the present study, we recruited six nurses to investigate individual differences in visual judgment of urinalysis test strips. Using the results from the automatic urine analyzer as the gold standard, we investigated the agreement rate between the results of visual judgment of urine

Table 3 Concordance rate with the automated urine analyzer

Parameter	A	B	C	D	E	F	Examiner avg.	Program
Protein	72.5	85.0	82.5	82.5	87.5	87.5	82.9	92.5
Glucose	97.5	97.5	95.0	100.0	90.0	97.5	96.3	95.0
Urobilinogen	82.5	92.5	77.5	95.0	90.0	90.0	87.9	92.5
Bilirubin	67.5	90.0	62.5	95.0	82.5	80.0	79.6	75.0
Ketone	95.0	97.5	87.5	97.5	92.5	95.0	94.2	100.0
Specific gravity	10.0	7.5	10.0	37.5	27.5	10.0	17.1	7.5
Occult blood	90.0	90.0	92.5	90.0	92.5	90.0	90.8	90.0
pH	67.5	82.5	70.0	35.0	52.5	62.5	61.7	82.5
Leukocytes	85.0	85.0	85.0	77.5	82.5	82.5	82.9	90.0
Nitrite	97.5	97.5	97.5	97.5	97.5	100.0	97.9	97.5
Creatinine	60.0	60.0	80.0	60.0	20.0	80.0	60.0	60.0
Albumin	100.0	100.0	0.0	0.0	0.0	100.0	50.0	100.0
Average of 12 parameters	77.1	82.1	70.0	72.3	67.9	81.3	75.1	81.9

Examiner avg., average visual judgment; Program, automated urine test strip colorimetric program

Table 4 The ± 1 rank agreement rate with the automatic urine analyzer

Parameter	A	B	C	D	E	F	Examiner avg.	Program
Protein	100.0	100.0	100.0	97.5	100.0	100.0	99.6	100.0
Glucose	100.0	100.0	100.0	100.0	97.5	100.0	99.6	100.0
Urobilinogen	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Bilirubin	100.0	97.5	95.0	100.0	100.0	97.5	98.3	100.0
Ketone	100.0	100.0	97.5	100.0	100.0	100.0	99.6	100.0
Specific gravity	30.0	25.0	30.0	27.5	42.5	20.0	29.2	17.5
Occult blood	97.5	97.5	97.5	97.5	97.5	97.5	97.5	97.5
pH	97.5	100.0	97.5	65.0	97.5	100.0	92.9	100.0
Leukocytes	100.0	97.5	97.5	92.5	97.5	97.5	97.1	100.0
Nitrite	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Creatinine	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Albumin	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Average of 12 parameters	93.8	93.1	92.9	90.0	94.4	92.7	92.8	92.9

test strips and the automatic urine test strip colorimetric program.

ICC (2.1) was calculated as the inter-examiner reliability of the six nurses to investigate individual differences in visual judgment of urinalysis test strips. As Cronbach's alpha was 0.8 or higher for all parameters, it was assumed that the internal consistency among raters is maintained; an ICC (2.1) value of 0.7 or higher generally indicates high reliability [16]. The parameters with an ICC (2.1) value of 0.7 or higher were protein, glucose, ketone, specific gravity, occult blood, leukocytes, and nitrite. Urobilinogen, bilirubin, pH, creatinine, and albumin had ICC (2.1) values less than 0.7. Although there were slight inter-rater differences for the parameters of protein, glucose, ketone, specific gravity, occult blood, leukocytes, and nitrite, differences were observed for urobilinogen, bilirubin, pH, creatinine, and albumin. These discrepancies may be attributed to the fact that the color changes caused by the application of urine may be different from the colors in the chart or may fall between adjacent colors. One of the evaluators recorded cases of difficult judgment when the color did not match any one of the color charts or fell between adjacent colors. In addition, the measurement time specified for urinalysis test strips was 30–60 s. As 12 parameters need to be judged within 30 s, it is difficult to judge similar colors, and there is a possibility of errors in judgment. In the future, it will be necessary to use instruments, such as a spectrophotometer or color difference meter, to measure the color difference between adjacent colors on the color chart, determine the color of urine test paper after adding urine, and verify the underlying cause for the color similarity.

In the present study, a comparison of the visual judgment and automated urine test strip colorimetric program revealed that the agreement rate of ±1 rank for each item, other than specific gravity, for both the examiner and the program, was over 90%. Furthermore, the agreement rate between the average value of the results for all 12 items and the average value of visual judgment was 75.1%. For the automated urine test strip colorimetric program, the average concordance rate for all 12 parameters was 81.9%, indicating a higher concordance rate. Urobilinogen, bilirubin, pH, creatinine, and albumin, which showed an ICC (2.1) value of less than 0.7, had higher than the average agreement rates using the automated urine test strip colorimetric program for all parameters except bilirubin. Considering the rate of concordance for each evaluator in visual judgment, some parameters differ depending on the evaluator. Regarding specific gravity, the ICC (2.1) yielded good results within the range of substantial criteria, but the concordance rate was significantly low, ranging from 7.5% to 37.5%. All judges noted that none of the colors on the color chart matched for specific gravity. Therefore, the matching rate is low because no matching color is displayed on the color chart after staining with urine. In the future, objective evaluation of colors using a spectrophotometer or color difference meter will be necessary. In addition, creatine and albumin had ICC (2.1) values of 0.516 and 0.661, respectively, indicating low inter-examiner reliability. As a small number of data points were obtained from the automated urine analyzer, the concordance rate is only for reference, and this method is considered effective.

Although several studies have evaluated program reliability using artificial urine and reagents [17)–19)], only a few studies evaluating program reliability using actual urine specimens are available. Rahmat et al. validated scanner color comparison using real urine specimens and reported an accuracy of 95.45% using a high-resolution scanner [12]. A low-resolution scanner reported an accuracy of 83%, which is comparable to the results of the automated urine test strip colorimetric program in this study. In addition, Africa et al. verified the accuracy of color comparison using 45 samples from an accredited urinalysis laboratory and reported an accuracy of 96.51% [20]; in this study, classification was performed using not only color comparison but also machine learning, resulting in improved accuracy. Although the automated urine test strip colorimetric program developed in the present study does not require data for learning, it is less accurate than the program developed by Africa et al [20].

Imaging using smartphones to analyze urinalysis test strips may support point-of-care testing in home settings [18) 21)]. An FDA-approved product that uses a smartphone application to test the parameters of urinalysis test strips for early detection of chronic kidney disease and urinary tract infections is available in the market [22] (Healthy.io, https://healthy.io/), and smartphone applications, such as "Vivoo" [23] and "Uchek" [24], have been developed. Moreover, attention is being paid to the use of point-of-care tests for health management at home because people are refraining from medical examinations due to the coronavirus pandemic.

This study has several limitations. First, the sample size is limited to 40. Furthermore, as this method requires many items and the steps are complicated, it may lead to measurement errors depending on the technique of the individual performing the evaluation. Therefore, it is important to reduce the number of required items and simplify the procedure. Moreover, the current automated urine test strip colorimetric program requires comparison with the standard color chart of each urine test strip manufacturer. Therefore, we are aiming to add a feature that will allow users to save the standard color charts from manufacturers that are mainly used in clinical practice and select the urine test strips to be used within the app. As part of the internal and external quality control surveys for the app, we intend to conduct multiple measurements of the same sample, having multiple people operate the program, and acquiring samples at multiple facilities.

V. Conclusion

The automatic urine test strip colorimetric program developed in the present study is a program that focuses only on color differences; therefore, it does not require a large amount of prior data for machine learning. In addition, if the standard color chart data of each manufacturer is available, processing urinalysis test strips produced by various manufacturers worldwide without the need for calibration is possible. However, because only 40 samples were analyzed in this study, concordance rates need to be verified in additional samples to determine the effectiveness of the automated urine test strip colorimetric program. Therefore, in the future, we plan to validate the automated urine test strip colorimetric program using a large number of urine specimens.

Conflicts of interest

The authors declare no conflict of interest.

Funding source

This study was supported by TERUMO LIFE SCIENCE FOUNDATION [grant number 21-III 7009] and the HUSM Grant-in-Aid for manuscript writing.

Acknowledgement

We would like to thank Takuya Kanamori, Ami Kinpara, Kouhei Sugiura, Toshinobu Moriyama, and Mitsue Tanaka for their cooperation in this research.

Authorship contributions

All authors were involved in the preparation of the manuscript and reviewed the final manuscript.

References

1) Prah JK, Amoah S, Ocansey DW, et al. Evaluation of urinalysis parameters and antimicrobial susceptibility of uropathogens among out-patients at University of Cape Coast Hospital. Ghana Med J 2019; 53(1): 44-51. doi: 10.4314/gmj.v53i1.7.

2) Simerville JA, Maxted WC, Pahira JJ. Urinalysis: A comprehensive review. Am Fam Phys 2005; 71(6): 1153-62.

3) Devillé WLJM, Yzermans JC, van Duijn NP, et al. The urine dipstick test useful to rule out infections. A meta-analysis of the accuracy. BMC Urol 2004; 4: 4. doi: 10.1186/1471-2490-4-4.

4) Winkens RA, Leffers P, Degenaar CP, et al. The reproducibility of urinalysis using multiple reagent test strips. Eur J Clin Chem Clin Biochem 1991; 29(12): 813-8. doi: 10.1515/cclm.1991.29.12.813.

5) Bekhof J, Kollen BJ, Groot-Jebbink LJM, et al. Validity and interobserver agreement of reagent strips for measurement of glucosuria. Scand J Clin Lab Investig. 2011; 71(3): 248-52. doi: 10.3109/00365513.2011.558109.

6) Saito Y, Koike J. Color expression by the nursing profession viewed from the perspective of color science. Kitakanto Med J 2001; 51(1): 35-41. doi: 10.2974/kmj.51.35.

7) Sakaguchi M, Saito Y. Examination of the color representation of bloody drainage in a test tube Comparison by years of clinical experience and observation opportunities. J Jpn Soc Nurs Res 1998; 21: 230. doi: 10.15065/jjsnr.19980630153.

8) Japanese Ophthalmological Society. Search by disease name. <https://www.nichigan.or.jp/public/disease/name.html?pdid=33> [cited 2023 Jan 5].

9) Guo D, Li G, Miao JQ, et al. A smartphone-based calibration-free portable urinalysis device. J Central S Univ 2021; 28: 3829-37. doi: 10.1007/s11771-021-4883-7.

10) Woodburn EV, Long KD, Cunningham BT. Analysis of paper-based colorimetric assays with a smartphone spectrometer. IEEE Sens J 2019; 19(2): 508-14. doi: 10.1109/JSEN.2018.2876631.

11) Balbach S, Jiang N, Moreddu R, et al. Smartphone-based colorimetric detection system for portable health tracking. Anal Methods 2021; 13(38): 4361-9. doi: 10.1039/D1AY01209F.

12) Rahmat RF, Royananda, Muchtar MA, et al. Automated color classification of urine dipstick image in urine examination. J Phys Conf Ser 2018; 978: 012008 doi: 10.1088/1742-6596/978/1/012008.

13) Inagaki K. Comparison of judgment results between the color comparison method of urine analysis test papers using the CIE DE2000 color-difference formula and the automatic urine analyzer. J Nurs Sci Eng 2021; 8: 170-6. doi: 10.24462/jnse.8.0_170.

14) Doros G, Lew R. Design based on intra-class correlation coefficients. Am J Biostat 2010; 1(1): 1-8. doi: 10.3844/amjbsp.2010.1.8.

15) Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33(1):

159-74.
16) Tsushima E. Medical data analysis learned in SPSS. Tokyo, Japan: TokyoTosho Co., Ltd. 2016: 212-3.
17) Kim S-C, Cho Y-S. Predictive system implementation to improve the accuracy of urine self-diagnosis with smartphones: Application of a confusion matrix-based learning model through RGB semiquantitative analysis. Sensors 2022; 22(14): 5445. doi: 10.3390/s22145445.
18) Ra M, Muhammad MS, Lim C, et al. Smartphone-based point-of-care urinalysis under variable illumination. IEEE J Transl Eng Health Med 2018; 6: 2800111. doi: 10.1109/JTEHM.2017.2765631.
19) Thakur R, Maheshwari P, Datta SK, et al. Machine learning-based rapid diagnostic-test reader for albuminuria using smartphone. IEEE Sens J 2021; 21(13): 14011-26. doi: 10.1109/JSEN.2020.3034904.
20) Africa ADM, Velasco JS. Development of a urine strip analyzer using artificial neural network using an android phone. ARPN J Eng Appl Sci 2017; 12(6): 1706-13.
21) Siu VS, Lu M, Hsieh KY, et al. Toward a quantitative colorimeter for point-of-care nitrite detection. ACS Omega 2022; 7(13): 11126-34. doi: 10.1021/acsomega.1c07205.
22) Healthy.io.
<https://healthy.io/> [cited 2023 Jan 5].
23) Vivoo.
<https://vivoo.io/products/vivoo-urine-test>[cited 2023 Jan 5].
24) Uchek.
<https://www.biosense.in/ucheck.php> [cited 2023 Jan 5].