SYSTR : A Machine Learning based Smart Urine Analysis Tool

K. Anusha
Department of *Biomedical engineering*
ACS college of engineering
Bangalore, INDIA
vanianusha143@gmail.com

K. Sharath
Department of *Biomedical engineering*
ACS college of engineering
Bangalore, INDIA
sharathkarnam24@gmail.com

GUIDE:
Dr.Anitha.S
Proffesor
ACS college of engineering
Bangalore, INDIA
dranithasammilan@gmail.com

Deepthi B
Department of *Biomedical engineering*
ACS college of engineering
Bangalore, INDIA
deepthib2024@gmail.com

Prathiksha .K
Department of *Biomedical engineering*
ACS college of engineering
Bangalore, INDIA
prathikshakatriki@gmail.com

***Abstract*— The proposed system presents a Computer Aided Diagnosis (CAD) for the early detection of diseases such as diabetes, urinary tract infections, kidney disorders using urine sample by adopting the principle of photoelectric comparison to test the number of biochemical components in the urine. Color changes in accordance with the reaction of the biochemical components in the dipstick by employing Colorimetric technique. This Technique is used to determine the concentration of compounds in fluids by the application of Beer-Lambert's Law. Each bud in the test strip has its principle of value, on which they react to the components present in the fluid. The parameters of the urine are analyzed by using a tool with test strips affixed to the lid through the colorimetric method or reflectance method. The steps that are performed by the proposed system are image acquisition, grid identification, RGB color feature extraction, data mapping followed by machine learning algorithm which compares the proportions of various parameters. Hence abnormalities in color are determined and disease is predicted at the earliest to provide a point of care (POC) diagnosis**

Keywords—Computer aided diagnosis, urinalysis, test strips, image processing, machine learning.

I. INTRODUCTION

One of the essential essences of human life is health. Early prediction of the disease by the symptoms of an illness that appears is necessary to receive proper medical treatment from the doctors. The result of advance examination of the symptoms of a diseas considerably affects the successive doctor action against a patient. Therefore, precise early examination result is necessary in order to obtain the correct result of early diagnosis or screening. The symptoms of an illness can be detected through various methods performed on the human body or from the results of human excretion such as urinalysis. Urinalysis is an vital part of the early examination of kidney disorders and urinary tract infections. Urine dipstick examination is performed to determine some chemical substances present in the urine like proteins, ketones, hemoglobin, bilirubin, glucose, urobilinogen, nitrites, acetone, and leucocytes, pH testing. Urine Reagent Strips are resistant plastic strips on which several separate reagent areas are appended. The test is for the identification of one or more of the following substances in urine: Leukocytes, Glucose, Ketone (Acetoacetic acid), Specific Gravity, Protein, Urobilinogen, Bilirubin, Blood, Nitrite, Ascorbic Acid, and ph. Urine examination using dipstick urine can be achieved by matching the color change on the reagent paper that has been dipped into the urine sample with the standard color chart of urine examination. This helps to determine the proportion of content contained in the urine. Color similarity between reagent papers and customary color chart of urine examination is still done physically using the eyes, causing differences in color perception that may affect urine examination results using a urine dipstick. here Euclidean Distance is used to resolve differences in color perception that occurs in urine examination using a dipstick urine, i.e., at the time of carrying out color matching between reagent strip with a unique color chart of urine analysis. There will be many changes during the time of body dysfunction before blood composition is changed to a certain extent. Urinalysis is a useful procedure as an indicator of health or disease, and as such, is a part or routine health screening. Urine Reagent Strips can be used in general to keep a check on health, and it also aids in the diagnosis and monitoring of metabolic or systemic disorders that affect the functioning of kidney, endocrine disorders and diseases or disorders of urinary tract. Diabetes is a metabolic problem with high occurrence rate today, and there are globally millions of people suffering from this disease. Although frequent blood-glucose monitoring can bring benefits to the health of diabetic patients, the disinfection process and stabbing pain during blood collection led to very poor compliance. Besides blood glucose, the content of urine glucose is another important parameter for the clinical diagnosis and regular self-monitoring of the patients. As urine glucose detection is a non-invasive procedure, patients' abidance is good, and more emphasis was placed on glucose determination.

A urine dipstick is a diagnostic tool used to identify pathological changes in subjects' urine in standard urinalysis.

A standard urine test strip may consist of up to 10 different chemical pads with reagents which react (change color) when dipped in, and then removed from, a urine sample. The test can be read in as little as 60-120 seconds after dipping, although sometimes it requires longer period. Regular testing of the urine with multiple parameter strips is the initial step in the diagnosis of many diseases. The analysis consists of the test for the presence of ketones, hemoglobin, proteins, glucose, bilirubin, urobilinogen, acetone, nitrite, and leucocytes as well as testing of pH, specific gravity or to test for infection by different pathogens. The test strips comprise of a ribbon like plastic or paper of about 5 mm wide, plastic strips have pads which has chemicals that react with the substances present in urine producing a peculiar color. In the paper strips the reactants are absorbed onto the paper. Paper strips are generally specific to one single reaction (e.g., pH measurement), while the strips with pads allow many determinations parallelly. These strips help in serving different purposes, such as qualitative strips that determine if the sample is positive or negative, or there are semi-quantitative ones that in addition to providing a positive or negative reaction also provide an approximation of the result, in the latter the color reactions are almost proportional to the concentration of the substance being tested. The output of the results is carried out by comparing the strip color with the color scale provided, no additional equipment is required. This type of analysis is very usual in the control and checking of diabetic patients. The time taken for the appearance of the test results on the strip can vary from a few minutes after the test to 30 minutes after immersion of the strip in urine). Semi-quantitative readings are usually reported as: trace, 1+, 2+, 3+ and 4+; although tests can also be estimated as mm per deciliter.

Point of Care diagnosis (POCT) has gained attention for its wide range of applications in the healthcare. Starting from blood glucose measurement to complex immunological evaluations, POC is user friendly even for end users even with feeble technical expertise. This characteristic makes it suitable for disease identification, quality of care, or patient monitoring. Although, discoveries in semiconductor technology have created transformative consumer devices, but they have limited impact on the healthcare industry. Medical devices have rigorous requirements for safety and effectiveness, and they generally iterate on power-hungry designs. One of the leading diagnostic tools of POC tool analysis is strip-based colorimetric diagnostics. When samples such as urine, blood, or other body fluids are deposited on a strip, signals will obtain in the form of colors. analyzing the color changes, which is done by manual human perception, is usually accurate. But the colors may be perceived differently in varying light conditions, resulting in diagnostic errors which are likely to affect the medical decision-making of the patient.

To reduce errors caused by human interference, some technological devices assist the interpretation of the diagnostic results that have been introduced. Digital photography, scanners, and smartphones have been utilized to understand the effects of light and colors on the chemistry of solutions. In healthcare products, device portable colorimetric readers, scanners, video cameras, or high-resolution digital cameras have been used to capture the colorimetric data as digital images. A sample holder is used to fix the paper-based microfluidic plastic substrate and the surrounding light effects are subtracted. The color of the reagent pads on the plastic base is then analyzed under the proper illumination of a computer screen.

II. METHODOLOGY

The uranalysis could be a set of screening tests that will realize some common diseases. It may be accustomed to screening for and/or facilitating the designation of conditions like tract infections, urinary organ disorders, liver problems, diabetes, or different metabolic conditions, to name a few.

A uranalysis, jointly called Routine and Microscopy(R&M), is an associate degree array of tests performed on urine, and one among the foremost common ways of diagnosis.

Principle: The uranalysis here is completed by mistreatment the instrument for the analysis of pee parameters through the quantitative chemical analysis technique or coefficient ways or such with the dipsticks (test strips). Adopting the principle of photoelectric comparison to see the quantity of organic chemistry parts within the urine in keeping with the color amendment within the dipsticks that reacts with the organic chemistry parts in urine. Each bud in the test strip has its principal importance, on that they react to the urine components.

The performance characteristics of the test to be as shown below:

Specific Gravity	1.016 - 1.022
pH	4.8 - 7.4
Leukocytes	-ve - apx 500 LEU/µL
Nitrite	-ve - +ve
Protein	Nil - 500mg/dL(5 g/L)
Glucose	Nil - 1000 mg/dL
Ketone Bodies	-ve - 150 mg/dL
Bilirubin	-ve - apx 6mg/dL
Urobilinogen	Normal - 12 mg/dL

Table: Measurement Table

Early morning first freshly voided midstream urine is always used for the uranalysis because it provides very precise results. ensuring regarding the instrumentation the individual uses, wide-mouth sterile plastic instrumentation with no additives, and no contamination. needed Equipment: Any uranalysis device supported the colorimetry (such as U 411 Cobas, Unisys 1100 Roche Siemens), test strips(dipsticks), Microscope, test tubes, and coverslips. Procedure:

1. Physical Examination: The color and appearance of the urine are observed and recorded.

2. Chemical Examination: The strip used for urine analysis for chemical examination is employed for the detection of specific gravity, pH, glucose, ketone bodies, protein, urobilinogen, nitrite, and bilirubin in urine. In semi-automated urine analyzers, the procedure is somewhat totally different from the manual technique. Verify the science lab -ID and patient with the date are labeled on the container. Scan the sample barcode and enter the color and clarity of the urine. Test strip from the ampoule and dip the strip into the well-mixed urine sample for some seconds(1-2sec). Dap the strip on the tissue to get rid of the excess urine from the strip. Place the sample strip onto the loading zone within the pee instrument and therefore the print of the result comes off the instrument.

3. Microscopy: The urine is to be centrifuged for 10-15 minutes at 2500-3000rpm and therefore the sediment of that sample is to be taken on a slide and lined with a coverslip. Then it's ascertained below the magnifier for additional examination of bacteria, pus cells, casts, animal tissue cells, RBCs, crystals, etc.

Colorimetric Analysis

A colorimetry could be a technique of crucial the concentration of a substance or compound in a very answer with the help of a color chemical agent. There are two elementary laws of absorption that are extremely vital in quantitative chemical analysis estimation. These are Lambert's law and Beer's law. Lambert's law states that once monochromatic light passes through a solution of constant concentration, the absorption by the solution is directly proportional to the concentration of the solution. In distinction, Beer's law states that once monochromatic light passes through a solution of constant concentration, the absorption by the solution is directly proportional to the concentration of the solution.

Thus, the variation of color of the reaction mixture (or system) with the amendment of substrate concentration forms the premise of colorimetry. The formation of color is because of the reaction between substances and reagents within the applicable proportions. The intensity of color ascertained is then compared therewith of a reaction mixture that contains a acknowledged quantity of substrate.

Photoelectric comparison

The photoelectrical effect could be a phenomenon within which electrically charged particles are released from or within a material when it absorbs an electromagnetic wave. The impact is commonly outlined because of the ejection of electrons from a metal plate once light falls on it. in a very broader definition, the energy is also infrared, visible, or ultraviolet, X-rays, or gamma rays; the fabric is also a solid, liquid, or gas; and therefore, the free particles is also ions (electrically charged atoms or molecules) moreover as electrons.

The development was basically vital within the development of recent physics because of the puzzling queries it raised regarding the character of light—particle versus rippled behavior—that was finally resolved by Albert Einstein in 1905. The effect remains vital for analysis in areas from materials science to astronomy, moreover, as forming the premise for a range of helpful devices.

SYSTR could be a urine-scan tool for early unwellness prediction, right from the comfort of home.

Procedure:

- SYSTR could be a uranalysis tool that comes with a sterile clear instrumentation.
- The uranalysis tool consists of urine parameters placed below the lid of the instrumentation. A urine sample is collected within the instrumentation, and it's flipped the wrong way up to let the reagents take a look at the strip reacting with the components of urine therefore, scanning it yields the readings of the individual parameter within the SYSTRs package. This work supported the RGB analysis.

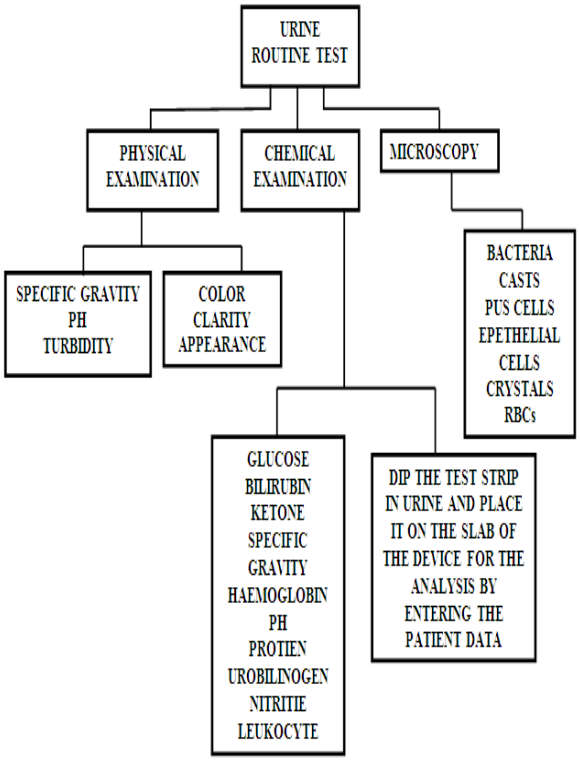
Then the analyzed knowledge is matched with the corresponding organic chemistry parameters and their concentration. this is often achieved through mistreatment knowledge mapping.

Figure: 3.1: operating setup

- The Readings are then analyzed in the software and a prediction is made regarding the different diseases/conditions one may be prone to or the risk of contracting a disease.

Software

Urinalysis tool The objective of the software is two-fold. First, to efficiently obtain the grid squares corresponding to the reagent as required and perform classification-based analysis to extrapolate the values used for the prediction of ailments. Second, a machine learning model will be used to convert the inferred values from the reagent matrix to a comprehensive report of possible diseases that the user may have.



Block diagram of SYSTR

III. SOFTWARE

IMAGE PROCESSING

Images outline the globe, every image has its own story, it contains a great deal of crucial information which will be useful in some ways. This information can be obtained with the assistance of the technique referred to as Image Processing. It is the core part of computer vision which plays an important role in several real-world examples like robotics, self-driving cars, and object detection. Image processing permits us to remodel and manipulate thousands of images at a time and extract necessary insights from them. It has a wide range of applications in nearly every field.

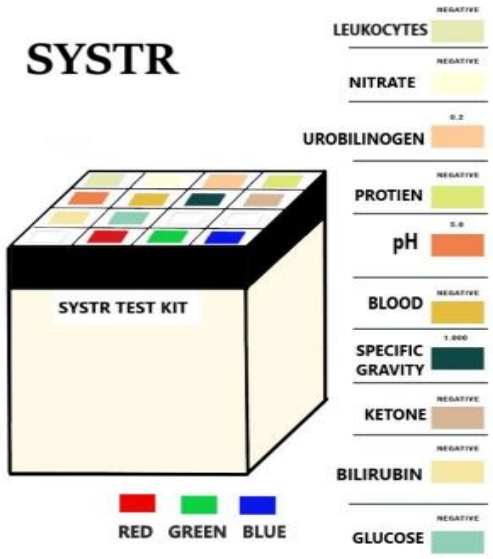
OPENCV

OpenCV is a huge open-source library for computer vision, machine learning and image processing. OpenCV supports a wide range of programming languages. It can process images and videos to spot objects, faces etc. OpenCV is integrated with numerous libraries such as NumPy, which is a highly optimized and extremely sophisticated library for numerical operations, i.e., multiple operations are often done at a stretch when NumPy combined with OpenCV.

NoSQL

NoSQL, additionally referred to as “not only SQL”, “non-SQL”, is an approach to database design that enables the storage and querying of data outside the traditional structures found in relational databases. Whereas it will still store information found inside relational database management systems (RDBMS), it just stores it differently compared to an RDBMS. The decision to use a relational database versus a non-relational database is basically discourse, and it varies depending on the use case.

IV. HARDWARE



Test strip:
Dipstick or test strip is considered as one of the basic diagnostic tools to determine pathological changes in a patient's urine. A standard test strip contains 10 different chemical pads that react (change color) with urine samples when made to immerse completely. The test can be read in as little as 60 to 120 seconds after dipping. Routine testing of the urine with parameter strips helps to diagnose a wide range of diseases.
The analysis includes testing for the presence of proteins, glucose, ketones, hemoglobin, bilirubin, urobilinogen, acetone, nitrite, and leukocytes as well as testing of pH and specific graveness or to test for infection by different pathogens. The test strips correspond to a strip made of plastic or paper of about 5 millimeters wide, plastic strips have pads saturated with chemicals that reply with the composites present in urine producing a characteristic color. Paper strips are frequently specific to a single response (e.g., pH dimension), while the strips with pads allow several determinations contemporaneously.
There are test strips that serve different purposes, similar as qualitative strips that only determine if the sample is positive or negative, or there are semi-quantitative ones that provide a positive or negative response that also give an estimation of a quantitative result, in the response the colour responses are roughly commensurable to the attention of the substance being tested for in the sample. The comparison results are carried out by comparing the pad colors with a colour scale handed by the manufacturer. This type of analysis is veritably common in the control and monitoring of diabetic cases.
SYSTR with its flagship urinalysis tool that comes with a sterile transparent container with its lid attached with 10 parameters in 4X4 matrix. 4 in each row filled in 3 different rows like shown in the figure. Last row is attached with red, green, and blue colors for RGB analysis. RGB analysis is done to obtain the exact values of colour information.

Camera:
The image is captured using camera with minimum of 10X resolution that has capacity to emphasize and process the image. It is just important to know that the choice of camera must be such that it does not distort the appearance of the test strips.

Container:
A sterile plastic container is is used such that the test strips are attached to the container lid in 4X4 matrix format as shown in the figure.

V. ML ALGORITHMS

Naive Bayes Algorithm: It is a classification technique based on Bayes’ Theorem with a presumption of independence. In plain terms, a Naive Bayes classifier presumes the presence of a convenient feature in a class which is not related to the presence of any other feature. It is simple and agile to foresee the class of a test data set. It also executes well in multi class prediction.

K- Nearest Neighbors: K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm which is used for classification problems. K-nearest neighbors (KNN) algorithm uses feature similarity to anticipate the values of new data points which further means that the new data point will be defined for a value based on how sharply it meets the points in the training set.

Decision tree: Decision Tree is a decision-making tool is a model of decisions and all their achievable results, including outcomes. It is a tree-structured classifier, where internal nodes serve as the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome

Support Vector Machine: Support Vector Machine (SVM) is an algorithm used for classification. The objective of the SVM algorithm is to look for a hyperplane in an N-dimensional space that classifies the data points. SVM chooses the extreme points/vectors that help in generating the hyperplane. These are called support vectors, and hence the algorithm is termed as Support Vector Machine

VI. ADVANTAGES AND DISADVANTAGES

1. A Hardware system consisting of grids of test strips rather than linear strip to test the presence of parameters
2. Real time monitoring that gives results instantly rather than conventional time taking processes
3. Absence of reagent to stop bacteria growth since we don't require the preservation of urine for a longer duration of time.
4. Confusion matrix will be used to check the most efficient algorithm among different algorithms considered under machine learning.
5. No human interference in analyzing the concentration of parameters diagnosing the disease.
6. There is no need for medical expertise since this can be operated even by a naive person.
7. Early intervention and treatment, which may either cure the disease or improve the outcome and quality of life of patient.
8. Non-invasive method of disease prediction.

VII. EXPECTED RESULT

As discussed above, The urine sample of the target user is collected in hardware container and the container is flipped over to let the reagents of the test strip to react with bio chemical components of the urine which would merely take 30-60 seconds. The compatible camera lens is used to capture the grid matrix under proper lighting condition and this data is fed to SYSTR software. The software initially identifies the grids and analyses the RGB values, compares it with the pre-fed standard values and determines the concentration of the biochemical components present in the urine sample using the principle of Euclidiean's distance.

Then machine learning model is used to predict the possible diseases with these obtained proportions of components. We test the model using four different algorithms (Naive Bayes, Decision tree, Support vector machine (SVM), K-Nearest neighbors) and choose the most accurate one using confusion matrix and train the model using the same.

Accurate results are anticipated to obtain the number of different parameters like ketones, protein, glucose, leukocytes, pH, bilirubin, etc. and the possibilities of occurrence of various diseases corresponding to these parameters in near future.

VIII. REFERENCES

[1] Prasad S, Tyagi AK, Aggarwal BB. Detection of inflammatory biomarkers in saliva and urine: potential in diagnosis, prevention and treatment for chronic diseases. 2016;241(8):783799

[2] Kamakoti V. Evaluation of molybdenum as an electrode material for affinity-based urine dipstick biosensing. 2017.

[3] Chen Z, Martin MT, Xinyu L. A microfluidic paper-based electrochemical biosensor array for multiplexed detection of metabolic biomarkers. 2013;14(5):054402.

[4] Link N, Weber W, Fussenegger M. A novel generic dipstick-based technology for rapid and precise detection of tetracycline, streptogramin and macrolide antibiotics in food samples. 2007;128(3):668680.

[5] Cho, N.H.; Shaw, J.E.; Karuranga, S.; Huang, Y.; da Rocha Fernandes, J.D.; Ohlrogge, A.W.; Malanda, B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res. Clin. Pract. 2018, 138, 271281.

[6] Karim, M.N.; Anderson, S.R.; Singh, S.; Ramanathan, R.; Bansal, V. Nanostructured silver fabric as a free-standing NanoZyme for colorimetric detection of glucose in urine. Biosens. Bioelectron. 2018, 110, 815.

[7] Kanchi, S.; Sabela, M.I.; Mdluli, P.S.; Inamuddin; Bisetty, K. Smartphone based bioanalytical and diagnosis applications: A review. Biosens. Bioelectron. 2018, 102, 136149.

[8] Nguyen, T.T.; Vandevoorde, K.; Wouters, N.; Kayacan, E.; De Baerdemaeker, J.G.; Saeys, W. Detection of red and bicoloured apples on tree with an RGB-D camera. Biosyst. Eng. 2016, 146, 3344.

[9] M. Webster, V. Kumar, Automated doctors: cell phones as diagnostic tools, Clin. Chem. 58 (2012) 16071609.

[10] L. Shen, J.A. Hagen, I. Papautsky, Point-of-care colorimetric detection with a smartphone, Lab Chip 12 (2012) 42404243.

[11] Sheerin N. 2011. Urinary Tract Infection. Medicine. 39(7): 384-389.

[12] McPherson R. and Pincus M. 2016. Henry's Clinical Diagnosis and Management by Laboratory Methods.

[13] Bush L. and Vasquez-Pertejo M. 2017. The Unintended Deleterious Consequences of the Routine Urinalysis. The American Journal of Medicine. 130(1): 3-4.

[14] Guo X., Li H., Woo S., Dong H., Lu F., Lange A. and Wu C. 2015. Glycolysis in the control of blood glucose homeostasis. Acta Pharmaceutica Sinica. (2)4: 358-367.

[15] Keopke J. To Sharpening of Color Image in Clinical Laboratory Diagnosis. 2013. IEEE International Conference on Systems, Man and Cybernetics.

[16] Lammers L., Gibson S., Kovacs D., Sears W. and Stratchan G. 2001. Comparison of test characteristics of urine dipstick and urinalysis at various test cutoff points. Annals of Emergency Medicine. (38)5: 505- 512.

[17] Strasinger S K and Marjorie S D L 2008 Urinalysis and Body Fluids Fifth Edition (Florida: F A Davis Company).

[18] Shanty R N T 2014 Color Interpretation Model of Urine Dipstick Test Results Qualitatively and Quantitatively for Mobile Devices (Surabaya: Institut Teknologi Sepuluh November).

[19] Estridge B H, Reynolds A and Walters N J 2000 Basic Medical Laboratory Techniques Fourth Edition (US: Delmar).

[20] Rahmat R F, Pulungan A F, Faza S and Budiarto R 2017 Image classification of ribbed smoke.