Bachelor of Science in Computer Science & Engineering

# A Smartphone Based Solution for Urinalysis

by

Abir Ebna Harun

ID: 1504011

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

April, 2021

# A Smartphone Based Solution for Urinalysis

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

Abir Ebna Harun
ID: 1504011

Supervised by

Dr. A.H.M Ashfak Habib
Professor
Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)
Chattogram-4349, Bangladesh.

April, 2021

The thesis titled '**A Smartphone Based Solution for Urinalysis'** submitted by ID: 1504011, Session 2019-2020 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

## Board of Examiners

________________________________ Chairman

Dr. A.H.M Ashfak Habib
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

________________________________ Member (Ex-Officio)

Dr. Asaduzzaman
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

________________________________ Member (External)

Dr. Pranab Kumar Dhar
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

# Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

**Signature of the candidate**

**Date:**

# Acknowledgements

The success and the final outcome of my work required a lot of guidance and assistance from many people and to them I am truly grateful. Above all, I want to thank The Almighty Allah for providing me the good health and patience to complete this thesis.

This thesis topic was provided by my honorable supervisor Dr. A.H.M Ashfak Habib, Professor, Department of Computer Science and Engineering, Chittagong University of Engineering & Technology (CUET). When I was working on the project, I wanted it to be done in researching approaches to find more and more possible solutions to get an optimal result. Dear Ashfak sir, I am truly honored and grateful for your tremendous support and appreciation toward my work. I am thankful for your guidance, advice and encouragement throughout the period of my thesis.

I also want to pledge my gratitude to Dr. Kaushik Deb, Professor, Department of Computer Science and Engineering, Chittagong University of Engineering & Technology (CUET) for introducing me to the world of computer vision and digital image processing which has played a great role throughout the entire work. I am also thankful to all the teaching staff of my department and some of my seniors from previous batches for the constant support, encouragement and guidance.

At the end, I want to thank my parents for their support, love and contribution throughout my entire life and academic career. Keep me in your prayer for this life and afterwards.

# Abstract

Urinalysis testing at medical facilities can be costly and inconvenient. According to recent reports, the demand for urinalysis tests will be worth around $2.21 billion by 2027. Our aim in this research is to develop a system that uses urinalysis reagent strip and smartphone as a point-of-care(POC) test device for performing laboratory-free urinalysis. The smartphone-based solution contains several image pre-processing steps to crop the image of a urine test strip from the background image. As a result, it can be used as a sample in an artificial neural network mapping model to map the position of each color pixel of the urine test strip, and then compare it to a reference color chart to provide an accurate computer vision integrated reading.

**Keywords:** Urinalysis, urinalysis reagent strip, urinalysis reagent reference color chart,Point-of-care (POC), Butter-worth filter, Gaussian filter, homomorphic filter, threshold, OTSU, K-means, SOM, cosine similarities, mean square error(MSE)

# Table of Contents


**Acknowledgements** **iii**

**Abstract** **iv**

**List of Figures** **vii**

**List of Tables** **viii**

**1 Introduction** **1**

1.1 Introduction . . . . . . . . . . 1

1.1.1 Urinalysis . . . . . . . . . . 1

1.1.1.1 Urinalysis Reagent Strip . . . . . . . . . . 2

1.1.1.2 Urinalysis Test Strip Reference Color Chart . . . 2

1.1.1.3 Preparing the Test . . . . . . . . . . 3

1.2 Framework/Design Overview . . . . . . . . . . 4

1.3 Difficulties . . . . . . . . . . 5

1.4 Applications . . . . . . . . . . 5

1.5 Motivation . . . . . . . . . . 6

1.6 Contribution of the thesis . . . . . . . . . . 7

1.7 Thesis Organization . . . . . . . . . . 7

1.8 Conclusion . . . . . . . . . . 8

**2 Literature Review** **9**

2.1 Introduction . . . . . . . . . . 9

2.2 Related Literature Review on Smartphone Based Solution for Urinalysis . . . . . . . . . . 10

2.2.1 Smartphone as a POC Platform . . . . . . . . . . 10

2.2.2 Use of Urine Test Strip . . . . . . . . . . 10

2.2.3 Illumination and Device Independence . . . . . . . . . . 11

2.2.4 ANN Model for Colorimetric Mapping . . . . . . . . . . 11

2.2.5 Similarity Algorithm for Labeling Color Pixels . . . . . . . 13

2.3 Conclusion . . . . . . . . . . 14

2.3.1 Implementation Challenges . . . . . . . . . . 14

**3 Methodology** **15**

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
3.2 Steps of the Proposed Methodology . . . . . . . . . . . . . . . . . . 15
3.2.1 Image Pre-processing . . . . . . . . . . . . . . . . . . . . . . 16
3.2.1.1 Homomorphic Filter . . . . . . . . . . . . . . . . . 16
3.2.1.2 Gaussian Blur . . . . . . . . . . . . . . . . . . . . . 17
3.2.1.3 Thresholding . . . . . . . . . . . . . . . . . . . . . 18
3.2.2 Applying ANN Coloriemtric Mapping Model . . . . . . . . 18
3.2.2.1 K-means Clustering . . . . . . . . . . . . . . . . . 18
3.2.2.2 Self-Organizing Mapping . . . . . . . . . . . . . . 19
3.2.3 Similarity Algorithm . . . . . . . . . . . . . . . . . . . . . . 19
3.2.3.1 Cosine Similarity . . . . . . . . . . . . . . . . . . . 20
3.2.3.2 Mean Square Error . . . . . . . . . . . . . . . . . . 20
3.2.4 Implementation . . . . . . . . . . . . . . . . . . . . . . . . . 20
3.3 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21

**4 Results and Discussions** **22**
4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
4.2 Dataset Description . . . . . . . . . . . . . . . . . . . . . . . . . . 22
4.3 Evaluation of the Steps of Proposed Smartphone Based Urinalysis System . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
4.3.1 Image Pre-processing . . . . . . . . . . . . . . . . . . . . . . 23
4.3.2 Applying ANN Colorimetric Mapping Model . . . . . . . . 24
4.3.2.1 K-means Clustering . . . . . . . . . . . . . . . . . 24
4.3.2.2 Applying Self-Organizing Mapping (SOM) . . . . 25
4.3.3 Similarity Algorithm . . . . . . . . . . . . . . . . . . . . . . 26
4.3.3.1 Cosine Similarity Approach . . . . . . . . . . . . 26
4.3.3.2 Mean Square Error Approach . . . . . . . . . . . 27
4.4 Evaluation of Performance . . . . . . . . . . . . . . . . . . . . . . . 27
4.5 Impact Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
4.5.1 Social and Environmental Impact . . . . . . . . . . . . . . 28
4.5.2 Ethical Impact . . . . . . . . . . . . . . . . . . . . . . . . . 28
4.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29

**5 Conclusion** **30**
5.1 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
5.2 Future Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

# List of Figures


1.1 Components of Urinalysis Reagent Strip . . . . . . . . . . . . . . . 2
1.2 Urinalysis test strip color chart . . . . . . . . . . . . . . . . . . . . . 3
1.3 Preparing sample and taking photo of the strip . . . . . . . . . . 4
1.4 Block diagram of a smartphone based urinalysis system framework 5
1.5 Accustrip® Urine Reagent Test Strip Reader . . . . . . . . . . . . 6

3.1 Steps of the proposed methodology . . . . . . . . . . . . . . . . . . 15
3.2 Block diagram of image pre-processing section . . . . . . . . . . . 16
3.3 Homomorphic filtration . . . . . . . . . . . . . . . . . . . . . . . . . 17

4.1 Dataset created from universal urinalysis color chart . . . . . . . . 23
4.2 Image pre-processing evaluation . . . . . . . . . . . . . . . . . . . . 24
4.3 Applying K-means clustering method . . . . . . . . . . . . . . . . . 25
4.4 Applying SOM using Numpy (CPU) . . . . . . . . . . . . . . . . . 25
4.5 Applying SOM using TensorFlow(GPU) . . . . . . . . . . . . . . . 26
4.6 Cosine similarities result . . . . . . . . . . . . . . . . . . . . . . . . 26
4.7 Mean square error approach result . . . . . . . . . . . . . . . . . . . 27

# List of Tables

4.1 Performance analysis using 2 test subjects . . . . . . . . . . . . . . 28

# Chapter 1

# Introduction

## 1.1 Introduction

Whenever we visit a doctor for a routine or uncommon disease, the doctor may recommend us to visit diagnostic centers for certain tests, one of which is "Urinalysis." According to Meticulous Research, urinalysis tests will reach a market value of approximately $2.21 billion by 2027 [1]. It shows the rapid growth of urinalysis test in the medical cases. The aim of this thesis study is to make urinalysis test cheaper and hassle free.

### 1.1.1 Urinalysis

Urinalysis means analyzing the components of urine with the help of some reagent and certain procedures [2]. This is a medical analysing procedure to get the data from the patient's urine for understanding the internal issues and finding the problem in order to cure them. By urinalysis test a patient can know about [3]:

- Leukocytes : Level of leukocytes detects UTI (urinary tract infection)
- Nitrite : Analysing nitrite also detects UTI.
- Urobilinogen : High urobilinogen means liver diseases such as hepatitis or cirrhosis.
- Protein : Finding proteinuria (can be caused by kidney diseases)
- pH : High pH indicates probability of having kidney stones.
- Blood : Hematuria (kidney dysfunction caused by bacteria)
- Specific Gravity : Kidney diseases, UTI, brain injuries.

- Ketone : High ketone in urine indicates diabetic ketoacidosis (DKA) that can lead to coma / death.
- Bilirubin : Level of red blood cell, indicates liver damage/disease , UTI.
- Glucose : Glucose level defines diabetes.

Urinalysis can be done by certain ways. One of the popular way of urinalysis is the microscopic way. The most conventional way is to perform urinalysis with urinalysis reagent paper [4].

#### 1.1.1.1 Urinalysis Reagent Strip

Urinalysis reagent strip or commonly known as urine test strip or dipstick is a diagnostic tool that is used to determine pathological change in a patients body [5]. It is usually a made by plastic with regent paper pallets on it. A standard urine test strip contains 10 individual reagent paper pallets.

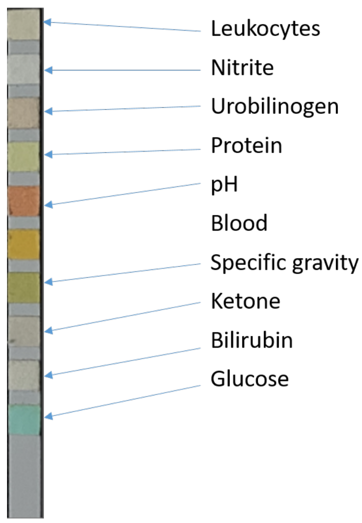



Figure 1.1: Components of Urinalysis Reagent Strip

In Figure 1.1 the components of a standard urinalysis reagent strip is shown. Each pallet of the reagent paper holds an unique color pattern and on the contrary of the reaction with urine causes a significant color change of the strip [6].

#### 1.1.1.2 Urinalysis Test Strip Reference Color Chart

The reagent pallets of urinalysis test strip follows a color changing pattern upon contacting the urine for certain times. The change of color follows a universal reference color chart.

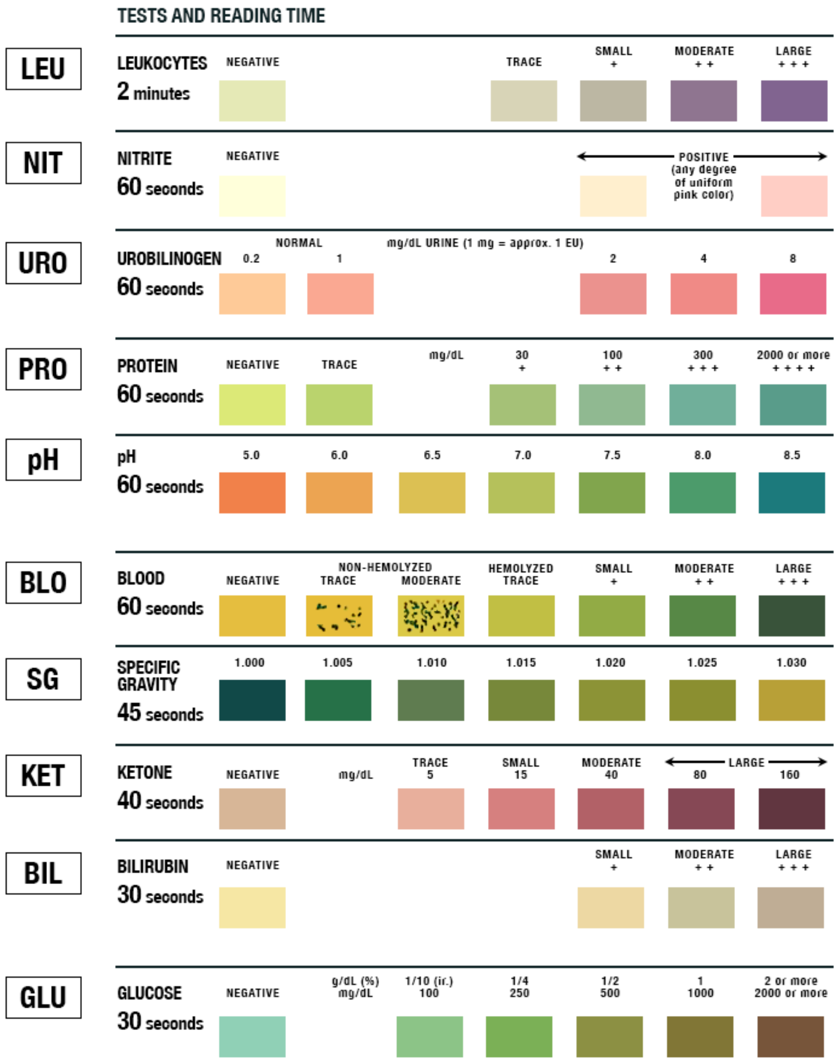



Figure 1.2: Urinalysis test strip color chart

The Figure 1.2 shows the urinalysis test strip reference color chart [7]. Each of the color plates contain an unique color combination to determine the result. The reference chart is initially the output of the test strip.

#### 1.1.1.3 Preparing the Test

To perform a urinalysis test using the urine test strip, it needs to be dipped in a test tube which is filled with the fresh urine of the patient. Then the test performer should carefully check that each of the color pallets of the strip has wet with urine and wait for 60-120 seconds.

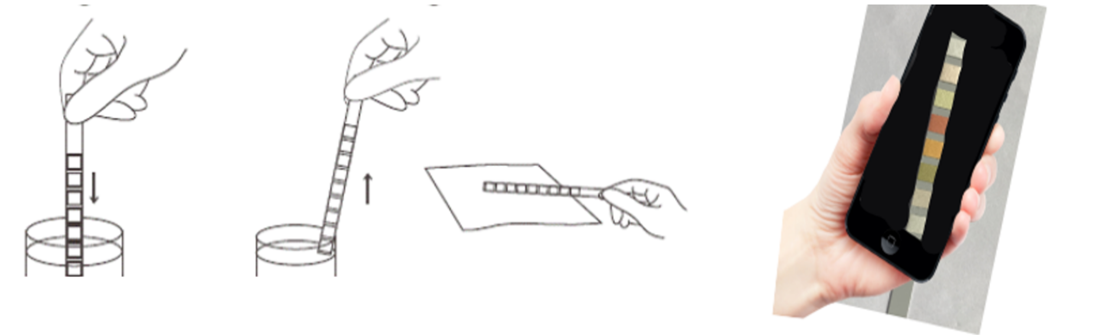



Figure 1.3: Preparing sample and taking photo of the strip

As shown in Figure 1.3, the strip need to be placed on a black paper or cloth after 60-120 seconds of dipping into the sample urine. The user would take a photo of the strip with the help of a smartphone in natural lighting condition.

## 1.2 Framework/Design Overview

The smartphone based urinalysis system contains an image pre-processing unit, an artificial neural network(ANN) model, a similarity algorithm and a data set containing the reference color sets. After capturing the image of the strip, the image get sent to the pre-processing unit. The main steps are summarized as -

1. The image pre-processing unit filters and crops the strip from the black background.
2. The processed image is used as a sample image in an artificial neural network(ANN) model.
3. The ANN model clusters and plots a map of the color pixels from the strip image.
4. A similarity algorithm determines the final reading of the strip by matching it with the reference color.

Block diagram of a smartphone based urinalysis system framework is shown below :

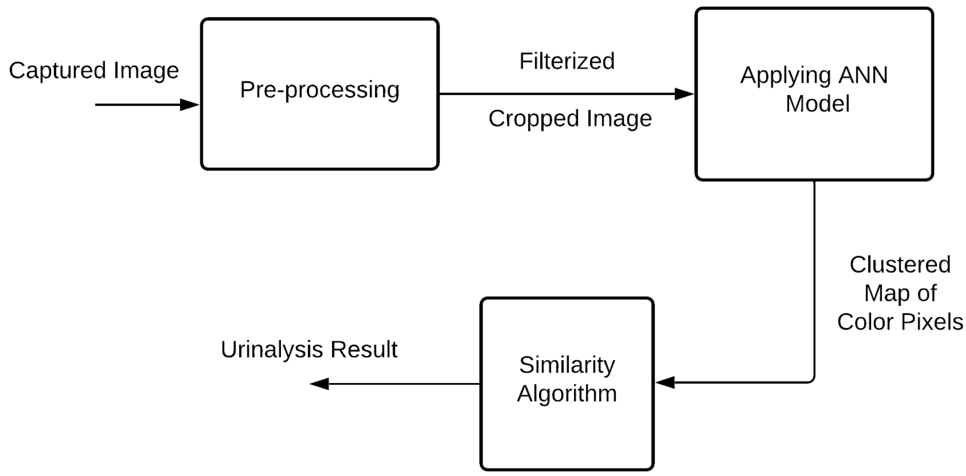



Figure 1.4: Block diagram of a smartphone based urinalysis system framework

## 1.3 Difficulties

To design a urinalysis via smartphone system, there are some certain difficulties have encountered. The major difficulties faced in this problem are :

- The number of urinalysis test strip is limited.
- Due to the global pandemic COVID-19, the enough real life test data could not be found.
- The image pre-processing unit can only work accurately if the image of the strip is taken on a black background.

## 1.4 Applications

According to the estimates of the global burden of disease, kidney and urinary tract disorders cause approximately 830,000 deaths and 18,467,000 disability-adjusted life per years [8]. The application of a smartphone based solution for urinalysis can create a global impact to change these numbers by giving the patient frequent,faster and accurate test reports. The major applications of the system are:

1. Anyone without the knowledge of medical terminology/experience can perform the test.

2. A great application for home based/frequent monitored patients.
3. Affordable and gives accurate result than the convenient method of testing.

## 1.5 Motivation

In 1964 - The company Boehringer Mannheim, today Roche, launched the first multicolored reagent strips [9]. The urinalysis reagent strip is now widely available on the market and generally used by personnel, medical facilities, and laboratories to collect urinalysis data from urine more easily. The strip is read by bare eyes in convenient use. In medical facilities and laboratories, an automated colorimetric analyzer machine is used to read the data from the strip.


Figure 1.5: Accustrip® Urine Reagent Test Strip Reader

In figure 1.5 the Accustrip® Urine Reagent Test Strip Reader machine is shown. This machine reads the data from the urinalysis reagent strip and shows the output. The approximate price of this machine is USD $1121.79. On the other hand, the standard reagent strip costs 11$ per box(contains 100 strips/box), which is capable of performing 100 individual tests.
Therefore, a smartphone based solution for urinalysis can provide more accuracy in results within a short period of time and also an affordable solution for the frequent check up on patients.

## 1.6 Contribution of the thesis

The aim of a thesis or research project is to achieve a particular set of objectives, such as defining a new approach or improving an existing one. In this thesis, every approaches and studies are played for achieving the following primary contributions.

1. To design a smartphone app that takes image and analyses the color blocks of the urinalysis reagent strips.
2. To analyze the processed data and matched it to the pre-fetched database and show the diagnosis results.
3. To gain the maximum user friendliness (since no medical knowledge is required to perform the test)
4. To reduce the time and cost consumption in urinalysis.

## 1.7 Thesis Organization

The rest of the thesis report is structured as follows:

- The chapter 2 provides a brief overview of previous research works in the fields of computer vision and urinalysis.
- The chapter 3 describes the proposed methodology for a smartphone based solution for urinalysis. In this chapter, several conventional approaches have been discussed. In detailed explanation sector, the use of Butter-worth filter, Gaussian filter, homomorphic filter, threshold, OTSU method, K-means clustering, self organizing mapping(SOM), cosine similarities, mean square error(MSE) in the system have been explained.
- The chapter 4 provides the possible outcomes and experimental results of each conventional approach of the proposed framework.
- At the end, Chapter 5 includes an overall summary of the thesis work as well as some future recommendations.

## 1.8 Conclusion

In this chapter an overview of urinalysis test, medical terminology and also test procedures is explained. This chapter also talked about the difficulties, the overview of the proposed framework, the motivation behind the work and also the contribution of the thesis. In the next chapter the literature review is provided, where the background and present state of the problem is described.

# Chapter 2

# Literature Review

## 2.1 Introduction

The point-of-care (POC) test is an alternative solution to traditional laboratory-based diagnostics due to reduced turnaround times, portability, and no need for highly trained laboratory staff. The smartphone is the perfect device for performing POC tests because of the capabilities of taking images and processing them to get the desired result. The most challenging factor in a smartphone-based urinalysis system is the illumination and device independence. The random lighting condition can affect the urinalysis biomarker's color. So several researches had done in this contrary. Another issue is selecting ANN/clustering methods which were used in the previous research to identify the biomarkers. Here in this chapter, we will discuss the various approaches and feature extraction techniques applied by different researchers, along with their performance and aspects of the system.
A brief discussion on smartphone based solution for urinalysis is divided into 4 basic parts. Smartphone as a POC platform, use of urine test strip, illumination and device independence, ANN model for colorimetric mapping and similarity algorithm for labeling color pixels are discussed in this chapter.

## 2.2 Related Literature Review on Smartphone Based Solution for Urinalysis

### 2.2.1 Smartphone as a POC Platform

Using a smartphone as a point-of-care(POC) test device is a modern life solution. Smartphones are often integrated into POC platforms due to their multifunctionality, enabled by high-quality digital cameras, computer processors, touchscreen interface and wireless data transfer. It is predicted that by 2020 about 80% of the planet's population will use smartphones [10].

There are several researches done by Haakon Karisen, Tao Dong on the cases of smartphone based rapid screening urinary biomarkers. A smartphone with a camera is a convenient tool for the analysis of colorimetric assays, and a software application has been developed for smartphones to photograph the colorimetric assay and classify colorimetric reactions according to the reference colors. To facilitate the detection of multiple biomarkers, e.g., 12 biomarkers with 2-7 references per biomarker, automatic localization of the test/reference pads has been implemented through recognition of corner alignment marks and projective coordinate transformation for perspective removal [11].

### 2.2.2 Use of Urine Test Strip

The use of urine test strip can be seen in the work of T. Smith et al. [12]. They introduced a completely unique manifold and companion software for dipstick urinalysis that eliminates many of the aspects that are traditionally suffering from user error and precise sample delivery, accurate readout timing, and controlled lighting conditions. Their results are obtained by capturing videos employing mobile and by analyzing them using custom-designed software. They showed that the results obtained with their proposed device are quite satisfying and consistent as a properly executed dip-and-wipe method whereas the special manifold is quite a limitation for additional tools free urinalyisis.

In another work with urine test strip by R.Yang et al. [13], they proposed a smartphone algorithm for colorimetric analysis of varied urine markers. A CIEDE2000 formula in CIELab color space was applied for the evaluation of color difference, which can greatly improve the analytical performances of urine strips. They implemented their system through MATLAB color library for color space transformation and the limitation is it was based on human vision.

### 2.2.3 Illumination and Device Independence

Karisen and Dong also faced the issue of illumination and device independence while doing the research on integrating smartphones as POC platforms. The main challenge is to achieve illumination and device independence for an outsized range of colorimetric biomarkers without using any standardization, like attachments, special lamps, or boxes. They achieved illumination and device-independent semi-quantitative detection by using discriminant analysis and classification of simultaneously photographed colorimetric test results and reference colors to confirm that any variation in image conditions applies approximately equal for reference and test [14]. This needs retraining of classifiers on each analysis but appears to be the foremost viable solution to solving the challenges while maintaining user-friendliness.

### 2.2.4 ANN Model for Colorimetric Mapping

Several previous researchers have worked on defining a high accuracy approach in color segmentation and mapping colors with an ANN model. A new technique of the dominant color palette extraction by K-means clustering algorithm and make it more efficient by using K-implies bunching calculation in some specific manner has explained by Kumar and Gopal in recent years [15].

An image segmentation system is proposed by Dong et al. for the segmentation of color image-supported neural networks [16]. So as to live the color difference properly, image colors are represented during a modified L/sup */u/sup */v/sup */ color space. The segmentation system includes a set of unsupervised and supervised segmentation modules. The unsupervised segmentation can be achieved by a two-level approach, i.e., color reduction and color clustering. In color reduction, image colors can be projected into a little set of prototypes using self-organizing map (SOM) learning. In color clustering portion, simulated annealing (SA) identifies the optimal clusters from the SOM prototypes. This two-level approach takes the benefits of SOM and SA, which may achieve the near-optimal segmentation with a coffee computational cost. The supervised segmentation contains both color learning and pixel classification. In color learning portion, color prototype is defined to be represented as a spherical region in color space. A procedure of hierarchical prototype learning (HPL) is used to obtain the various sizes of color prototypes from the sample of object colors. These color prototypes provide an honest estimate for object colors. The image pixels can be classified by the matching of color prototypes. The experimental results show that the system has the specified ability for the segmentation of color images during a sort of visual task.

In this thesis, the final outcome followed a SOM model to cluster the colors from the sample image.The self-organizing map (SOM) was explained by Xiao-Yu Zhang, Jiu-Sheng Chen, Jian-Kang Dong [17] which is considered as a powerful tool for exploratory data analysis which has been employed during a wide selection of color clustering. SOM, which is an unsupervised neural network mapping a group of n-dimensional vectors to a two-dimensional topographic map, are able to do the near-optimal segmentation with low computational cost. We means that the amount of output units utilized in a SOM influences its applicability for clustering. By proposing a clustering method that efficiently classifies image objects with an unknown probability distribution, without requiring the determination of complicated parameters, they demonstrated that SOM are often used for clustering. To make sure that, this clustering method is efficient and highly reliable, they defined a hierarchical SOM and used it to construct the clustering method. Their experimental results showed that the system has the specified ability for the clustering of color a spread of vision tasks.

An improved SOM algorithm and its application to color feature extraction were described by LP Chen, YG Liu, ZX Huang, YT Shi [18]. The finest method to cluster a color-induced map. They simply reduce the redundancy of the dominant color features in a colored image and while preserving the diversity and quality of the extracted colors is of importance in applications such as image analysis and compression. In their research, they represented an improved self-organization map (SOM) algorithm namely MFD-SOM, and its application to color feature extraction from images. In addition, they applied a linear neighborhood function within the proposed algorithm getting to improve its performance on color feature extraction. Their results of feature extraction on image datasets demonstrate the characteristics of the MFD-SOM algorithm.

### 2.2.5 Similarity Algorithm for Labeling Color Pixels

The extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis was explained by Mojsilovic, Hu and Soljanin in their research [19]. In their study, they present a replacement algorithm for color matching that models the behavior of the human sensory system in capturing the colour appearance of a picture. They first developed a replacement method for color code-book design within the Lab space. The tactic is compatible for creating small fixed color code-books; for image analysis, matching, and retrieval. Then they introduced a statistical technique to extract perceptually relevant colors. They also proposed a replacement color distance measure that's supported the optimal mapping between two sets of color components representing two images.

On the contrary of using cosine similarity method; Bora, Gupta and Khan published research [20] on comparing the performance of L*A*B* and HSV color spaces with respect to color image segmentation. Due to color image segmentation is a very conventional topic for image processing research, the LAB and HSV are the two frequently chosen color spaces. In their research, they provided a comparative analysis to performed between these two color spaces with respect to color image segmentation. For measuring their performance, they considered the parameters: mse and psnr . At the end they found that HSV color space is performing better than LAB.

A new image segmentation method was explained by Mohsen, Hadhoud, Moustafa and Ameen based on particle swarm optimization [21]. This method focusing on the mean square error (MSE) that can also be used to refine the result of similarity condition which is the smallest difference between means. .

## 2.3 Conclusion

In this chapter, a detailed literature review of the thesis is discussed. For better understanding, the discussion was split up into 4 parts. Different approaches and researches used by different researchers are described here for each part. After the discussion of the implementation challenge, the next chapter will describe the proposed methodology and its detailed explanations.

### 2.3.1 Implementation Challenges

The implementation challenges for this thesis according to the previous researchers are as follows :

- There is not enough data set for performing a robust solution according to some researchers that had described in this chapter.
- Not enough device is available to test the device independence and illumination issue.
- The test cases were risky to find due to the global pandemic COVID-19.

# Chapter 3

# Methodology

## 3.1 Introduction

Creating a smart phone-based urinalysis solution is a computer vision-integrated task. Several traditional methods were used in this thesis work. Each method produce different results. The study's aim was to find the best solution with the fewest complications with each method. The proposed method and its detailed explanation are presented in this chapter.

## 3.2 Steps of the Proposed Methodology

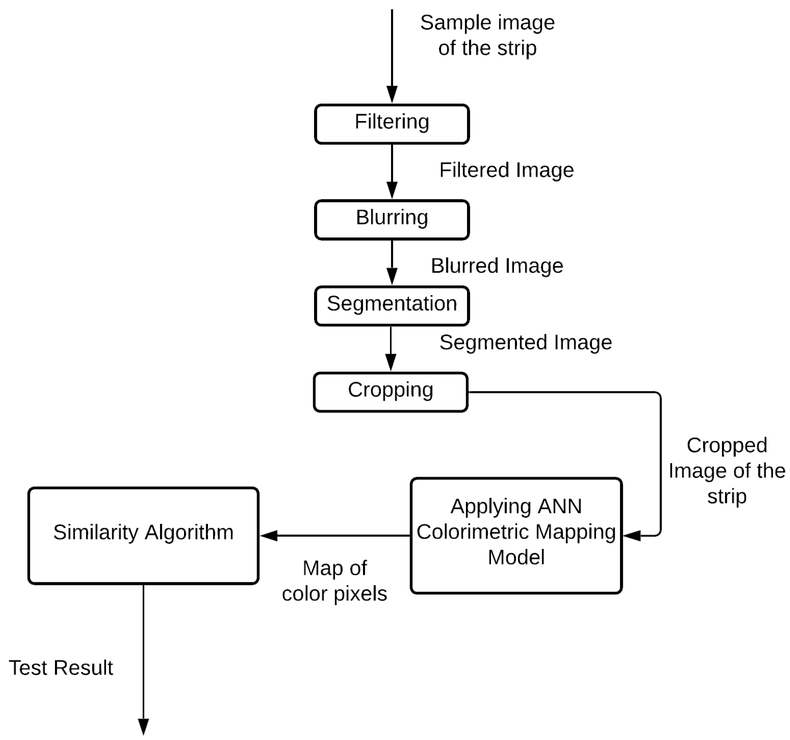



Figure 3.1: Steps of the proposed methodology

Figure 3.1 shows the steps of the proposed methodology for smartphone based solution for urinalysis. The system is split into three parts. Image pre-processing, ANN models for colorimetric mapping and similarity algorithm approaches. They are described below:

### 3.2.1 Image Pre-processing

After capturing the image, the image is sent to the pre-processing section. The image is now processing with the homomorphic filter, Gaussian blur and thresholding method to become segmented in order to get cropped from the black background.

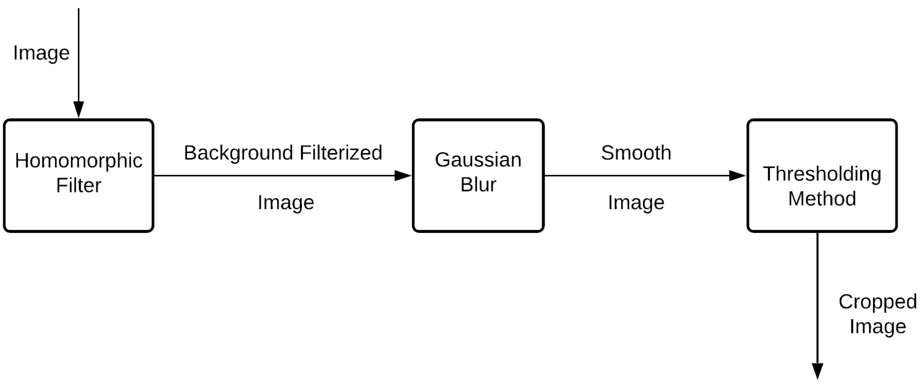



Figure 3.2: Block diagram of image pre-processing section

In Figure 3.2 the overview of pre-processing unit is described. The detailed explanation of each process is given below :

#### 3.2.1.1 Homomorphic Filter

Homomorphic filtering is a generalized image processing technique that involves a nonlinear mapping to a special domain, followed by a mapping back to the first domain using linear filter techniques [22] .

Figure 3.3 shows how homomorphic filter processes an image [23]. After applying homomorphic filter, it frequently normalizes the brightness of an image and increases the contrast. To remove multiplicative noise, homomorphic filtering is used. Although illumination and reflectance cannot be separated, their approximate locations in the frequency domain can be determined [24]. Since illumination

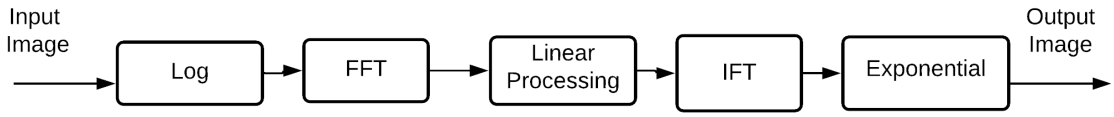



Figure 3.3: Homomorphic filtration

and reflectance combine in a multiplicative manner, the components are made additive by taking the logarithm of the image intensity, allowing these multiplicative image components to be separated linearly in the frequency domain. Illumination variations can be thought of as multiplicative noise that can be reduced by log domain filtering.Since the high-frequency components are assumed to represent more the reflectance in the image of the strip, and the low-frequency components are assumed to represent mostly the illumination in the scene, the high-frequency components are increased and the low-frequency components are decreased to make the illumination of an image more equal .

#### 3.2.1.2 Gaussian Blur

Gaussian blur, also known as Gaussian smoothing, is a blurring technique used in image processing. Using a Gaussian blur on an image is equivalent to mathematically convolving the image with a Gaussian function [25].

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{2\sigma^2}{2\sigma^2}} \tag{3.1}$$

Equation 3.1, where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and is the Gaussian distribution's standard deviation [26]. When applied in two dimensions, this formula yields a surface with concentric circles and a Gaussian distribution from the center point. This distribution's values are used to construct a convolution matrix, which is then applied to the original image. The new value for each pixel is a weighted average of the pixels in its near area. The original pixel's value receives the most weight (due to its highest Gaussian value), while neighboring pixels receive less weight as their distance from the original pixel increases. This produces a blur that preserves edges and boundaries better than other, more uniform blurring filters. Which is a necessary step before the approach of thesholding method.

#### 3.2.1.3 Thresholding

In thresholding approach, OTSU method is used to minimize intra-class intensity variance, or equivalently, by maximizing inter-class variance. The OTSU algorithm looks for a threshold that minimizes the intra-class variance, which is defined as the weighted sum of the two classes' variances.

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \tag{3.2}$$

Equation 3.2 derived the weights $\omega_0$ and $\omega_1$ are the probabilities of the two classes separated by a threshold $t$, and $\sigma_0^2$ and $\sigma_1^2$ are variances of these two classes [27]. By applying OTSU method, max contour of the image can be segmented. The system can finally crop the image of the strip from the background image.

### 3.2.2 Applying ANN Coloriemtric Mapping Model

After cropping the image, the next step is to use an ANN based colorimetric mapping model. For researching purposes, K-means clustering and Self-organizing mapping (SOM) both ANN models have been used in individual cases in this thesis.

#### 3.2.2.1 K-means Clustering

K-means clustering is a technique for partitioning a series of data points into many disjoint subsets with the points in each subset considered to be "close" to one another (according to some metric). The standard Euclidean distance function is a common metric, at least when the points can be geometrically represented. This method can be purely automated though it is very difficult to achieve a consistent and accurate image segmentation. Current image segmentation research using clustering algorithms shows that the K-means clustering algorithm provides the best results so far, although there are some changes that can be made to boost the results [28].

Color annotation using K-means clustering is also considered a valid approach in this case. The use of the K-means clustering algorithm to automatically annotate colors with the necessary terms using predefined colors for automatic image annotation [29]. Experiments are carried out to determine the optimal number of centroids, distance measures, and initialization mode for clustering. But in this case, this approach couldn't annotate enough color pixels.

#### 3.2.2.2 Self-Organizing Mapping

A self-organizing map (SOM) or also known as self-organizing feature map (SOFM) is a type of artificial neural network (ANN) that is trained using unsupervised learning to generate a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, referred to as a map, and is thus a dimensionality reduction technique [30].

$$W_v(s+1) = W_v(s) + \theta(u, v, s) \cdot \alpha(s) \cdot (D(t) - W_v(s)) \tag{3.3}$$

In equation 3.3 , s is the phase index, t is the training sample index, u is the index of the best matching unit for the input vector $D(t)$, $\alpha(s)$ is a monotonically decreasing learning coefficient, and $\theta$(u, v, s) is the neighborhood function that provides the distance between the neuron u and the neuron v each step [31]. Depending on the implementations of SOM in a color image, t can scan the training data set systematically. In this thesis, the both approaches of SOM using numpy (CPU integrated) and TensorFlow (GPU integrated) are applied. The result of both type implementing approach will be discussed in the next chapter.

### 3.2.3 Similarity Algorithm

After applying the colorimetric mapping, a similarity algorithm is needed to match the feature with the urinalysis reference color chart. In this study, several similarity algorithms have been approached and each of them provide different outcomes. Cosine similarities and mean square error (MSE) are the methods used in this section.

#### 3.2.3.1 Cosine Similarity

The similarity between two non-zero vectors in an inner product space is cosine similarity. It's equal to the cosine of the angle between them, which is the same as the inner product of the same vectors normalized to have the same length.

$$\text{similarity } = \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{nA_iB_i}}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{3.4}$$

In equation 3.4, $A_i$ and $B_i$ are components of vector $A$ and $B$ respectively [32]. In this thesis work, $A$ is the vector representation of strip image and $B$ is the vector representation of reference color images.
The limitation of cosine similarity function is, it can only be applied in HSV color space. This entire thesis work was performed in RGB color space due to the implication of SOM. So, cosine similarity was a failed approach in this case.

#### 3.2.3.2 Mean Square Error

The mean square error (MSE) is the average of the squares of the error, or the average squared difference between the predicted and real values.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(e_i)^2 = \frac{1}{n}\mathbf{e}^\top\mathbf{e} \tag{3.5}$$

Equation 3.5 defines that the MSE is the mean $\left(\frac{1}{n}\sum_{i=1}^{n}\right)$ of the squares of the errors in $\mathbf{e_i}$ matrix, where $\mathbf{e}$ is an $n \times 1$ matrix [33]. In this thesis work, the matrix $\mathbf{e}$ contains the difference between the mapped pixels and reference color pixels. The least MSE indicates the most similarity between 2 color pixels.

### 3.2.4 Implementation

All the approaches that has described above is implemented by the help of Python, OpenCV, TensorFlow , Numpy and some conventional tools. The experimental results of these approaches will be described briefly in the Chapter 4.

## 3.3 Conclusion

In this chapter the proposed methodology of smartphone based solution for urinalysis system is described. For better understanding and explaining it is divided into 3 subsection : image pre-processing, applying ANN colorimetric mapping model and similarity finding algorithm for color pixels. On the next chapter the experimental result analysis of the proposed methodology is provided.

# Chapter 4

# Results and Discussions

## 4.1 Introduction

In the previous chapter, a detailed explanation of the proposed smartphone based solution for urinalysis was described. This chapter examines the performance of the proposed methodology. The system was implemented in Python IDE with Intel Core i7 processor and 8 GB RAM. The graphics unit NVIDIA GeForce GTX 1060 is also used in some certain cases. For this thesis work, the universal reference color-chart by Multistix [7] is used as dataset.

## 4.2 Dataset Description

The only dataset for this thesis was the universal urinalysis reference color chart for urine test strip. Which is segmented purposely and categorized into 10 categories according to their labeling. The categories provide the identification of each unique biomarkers and contain unique color combinations. [13] Each category is put into individual folders by carefully labeling them with their primary identities. The dataset can be called by the Python IDE and easily flatten into matrices in order to find similarities between the sample colors.

The dataset is believed to be a standard measurement in urinalysis test. Here in the figure 4.1 the primary dataset is illustrated :

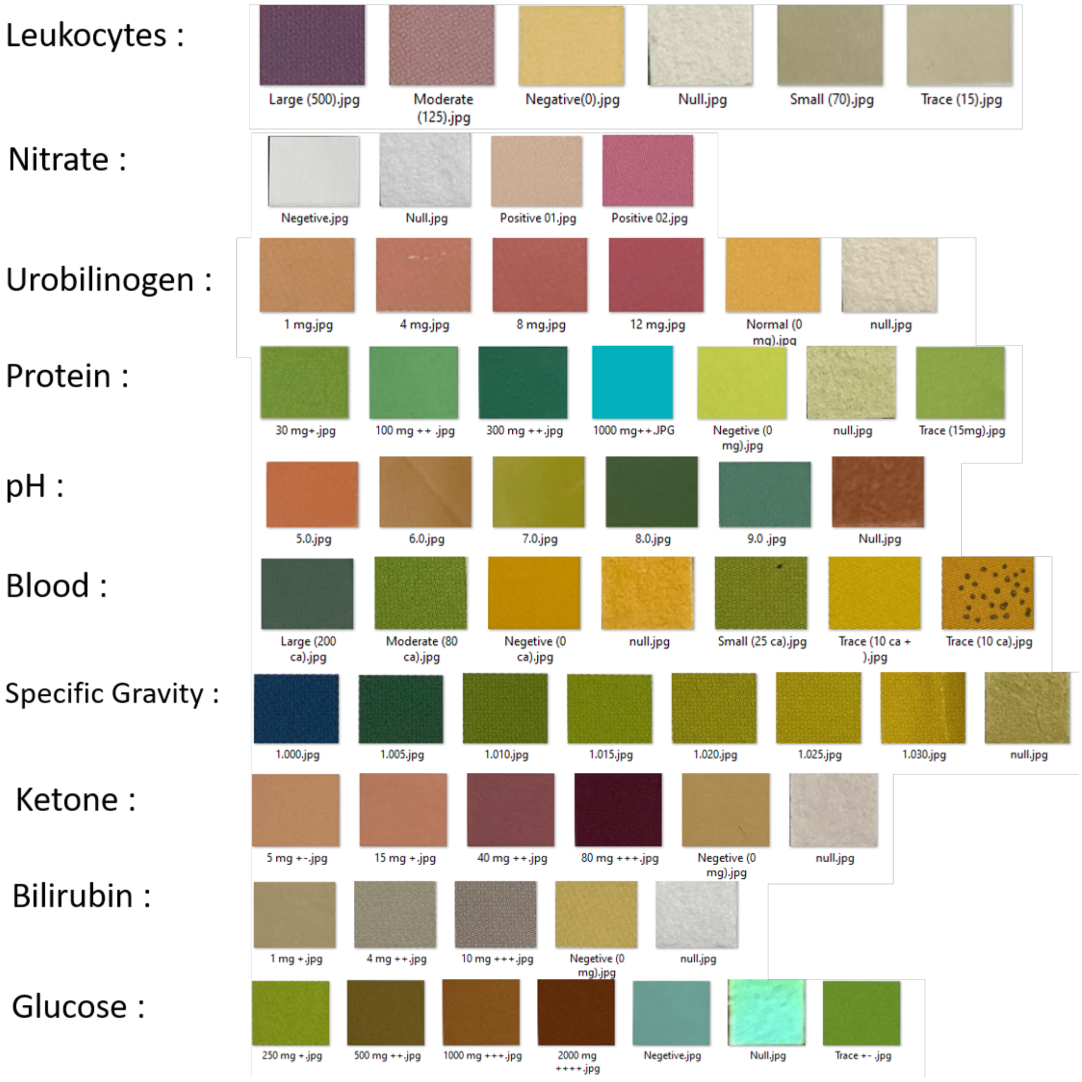



Figure 4.1: Dataset created from universal urinalysis color chart

## 4.3 Evaluation of the Steps of Proposed Smartphone Based Urinalysis System

Evaluation of each step of the proposed methodology is described as follows:

### 4.3.1 Image Pre-processing

The image pre-processing system contains the sequence of homomorphic filter, Gaussian blur and thresholding method. The evaluation of image pre-processing section is given below:

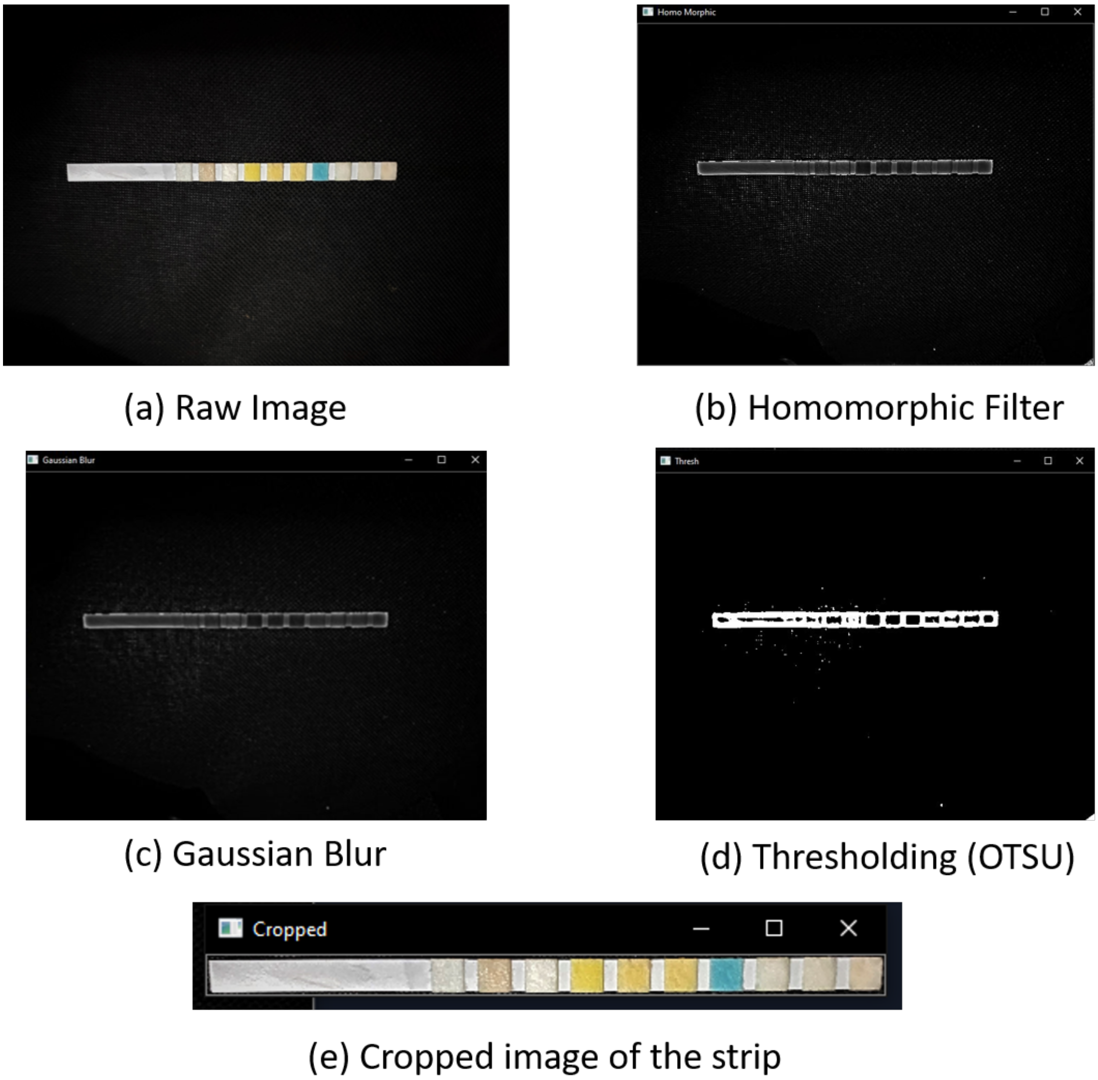



Figure 4.2: Image pre-processing evaluation

In Figure 4.2, the maximum accuracy of cropping the strip from the background image can be observed. In any lighting condition, the system can crop the image of the strip from a black background.

### 4.3.2 Applying ANN Colorimetric Mapping Model

In this part both K-means and self organizing mapping (SOM) has been applied and the results are deeply observed.

#### 4.3.2.1 K-means Clustering

Applying K-means clustering to find and cluster color pixel at this large scale was quite difficult process because of the fix number of centroid. The experimental result after applying K-means is shown in Figure 4.3 below :

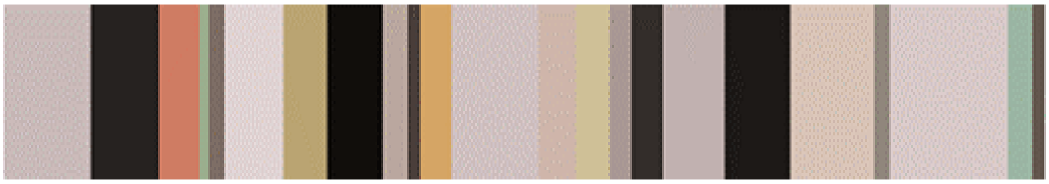
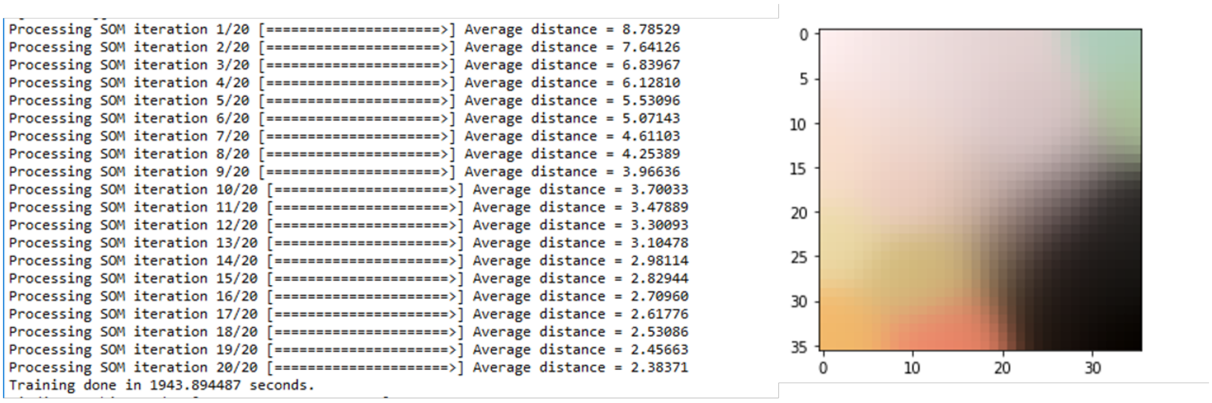


Figure 4.3: Applying K-means clustering method

All the colors of the strips can not be found in the cluster. Hence, the result is not satisfied.

#### 4.3.2.2 Applying Self-Organizing Mapping (SOM)

In this thesis work, At first SOM was applied by Numpy library in Python IDE. The ANN runs the testing processes through the CPU which takes a lot of time.




(a) Long training session of SOM (b) SOM output

Figure 4.4: Applying SOM using Numpy (CPU)

As shown in Figure 4.4, the SOM output is quite satisfying but the training time is so long (approx 32 minutes for our mentioned CPU). So, we look for an alternative solutions for this problem. Since self-orgainizing mapping (SOM) is a graphic integrated ANN model [34], it can be processed through the GPU (Graphics processing unit). Hence, TensorFlow library is used [35] to process SOM with GPU unit.

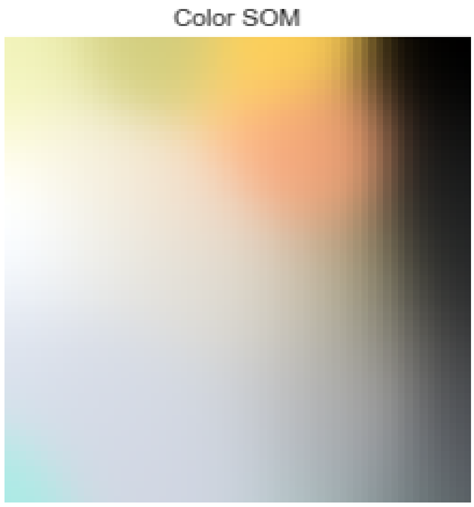



Figure 4.5: Applying SOM using TensorFlow(GPU)

In Figure 4.5, we can see the satisfying result by applying SOM using GPU. The approximate time to get this output is also very minimal.

### 4.3.3 Similarity Algorithm

In this section, both cosine similarty and mean square erron (MSE) approaches are evaluated:

#### 4.3.3.1 Cosine Similarity Approach

We call the dataset file in out Python IDE and use the cosine similarity method to find out similarites between the mapped color pixels and the reference chart.

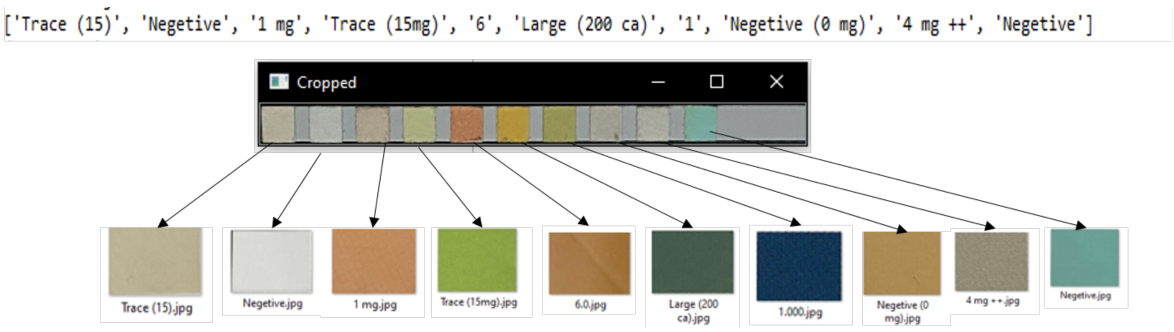



Figure 4.6: Cosine similarities result

As shown in Figure 4.6, a not-dipped (no contact with urine) strip was used to test. The expected result is all "Null" but the cosine similarity approach couldn't give a satisfying result because of the RGB color space.

#### 4.3.3.2 Mean Square Error Approach

Then, the mean square erron (MSE) approach is evaluated. The sole use of this algorithm is to find out the minimum error or dissimilarity between the clustered color pixels and the reference color pixels from the chart. The minimum mean square error defines the similarity between the color pixels.

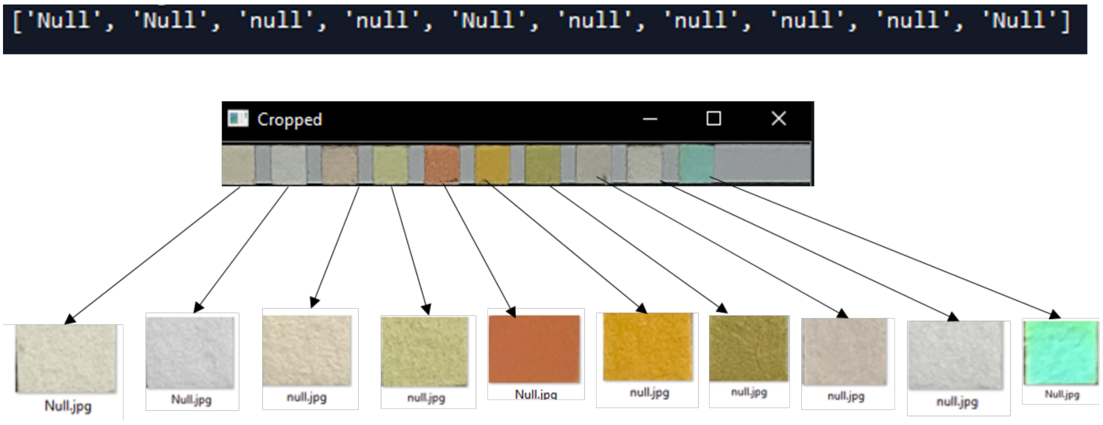



Figure 4.7: Mean square error approach result

As shown in Figure 4.7, a not-dipped (no contact with urine) strip was used to test. Mean square error approach gives a satisfying result by mentioning all the color pallets are "Null"(not dipped in urine).

## 4.4 Evaluation of Performance

Without the "Null" test, we also take samples from 2 volunteered human subjects for experimental testing purposes. Our subject 1 is diabetic, 50 years old and our subject 2 is 25 years old without any health disorders. The comparison of the real-life strip data and the experimental results from our proposed system shows a great deal in the performance evaluation.

In Table 4.1 the performance of the proposed system is compared with the real-data. There are some certain dissimilarities can be seen in some cases of the result because some color pixels have complex matrix structures and difficult to determine with conventional approaches. Overall, the accuracy of the system in this case is much satisfying.

Table 4.1: Performance analysis using 2 test subjects

| Subject: 1 (50 years old) | In real-life data | In Proposed System | Subject: 2 (25 years old) | In real-life data | In Proposed System |
|---|---|---|---|---|---|
| Leukocytes | Trace(15) | Trace(15) | Leukocytes | Negative | Negative |
| Nitrite | Negative | Null | Nitrite | Negative | Negative |
| Urobilinogen | No-read | Normal | Urobilinogen | Normal | Normal |
| Protein | Maximum | 10g+ | Protein | Trace | Trace |
| pH | 7.0 | 7.0 | pH | 6.0 | 6.0 |
| Blood | Negative | Negative | Blood | Negative | Negative |
| Specific gravity | 1.030 | 1.030 | Specific gravity | 1.025 | 1.025 |
| Ketone | No-read | Null | Ketone | Negative | Negative |
| Bilirubin | 1+ | 1+ | Bilirubin | Negative | Negative |
| Glucose | No-read | Negative | Glucose | Negative | Negative |

## 4.5 Impact Analysis

A smartphone based solution for urinalysis can create a massive impact in social platform. Here in this section the impact analysis is described :

### 4.5.1 Social and Environmental Impact

The social and environmental impacts of the proposed system are :

- Urinalysis can be done within the short period and hassle free.
- User can get more accurate results than the traditional method.
- Urinalysis test cost can be minimized.
- No medical knowledge is needed to conduct a smartphone based urinalysis test.

### 4.5.2 Ethical Impact

This thesis work is an experimental approach to do medical diagnosis with the help of smartphones. If anyone uses this medical based system for unethical purposes [36] such as altering medical reports, prescribing wrong medicine, wrong recommendations etc can be harmful/life-threatening for the patient. So, moral obligation is extremely needed.

## 4.6 Conclusion

This chapter describes the results of the proposed system. Performance of the system is also explained. So, the proposed system can perform a better urinalysis test from traditional approaches.The output of SOM and mean square error gives almost an accurate result in our cases. Still we illustrated the result of other approaches in this chapter for future research and development process. In the next chapter, the conclusion of the thesis work is drawn.

# Chapter 5

# Conclusion

## 5.1 Conclusion

A smartphone based system can be a revolutionary solution for urinalysis. Here in this thesis, several conventional computer vision and image processing methods are applied and their experimental results are discussed thoroughly. The use of homomorphic filter, Gaussian blur, threshold, OTSU, histogram analysis, k-means clustering , SOM, cosine similarity, mean square error methods can be found through the thesis work.

The global pandemic COVID-19 was a great problem to gather real-life test data and urinary related patients to conduct more tests. If more data was found, the system could be developed in a robust way. The SOM could create a test library with the acquired test data and could give a fast and more accurate result by using machine learning with any sample data.

Although in this thesis, full work is done by using pure mathematics and conventional approaches. The more scopes in future work are described below :

## 5.2 Future Work

In this thesis, the image pre-processing unit can only segment and crop the image of the strip from a black background. This issue can also be solved in future with more research.

An experimental mobile application can be developed and released for free in order to gather more test data by uploading the result and the image of the strip in an online server. After gathering a good number of data, a machine learning approach along with a robust system that can train itself with massive amounts of test data can be developed. There are some color matrices which are too complex to plot or flatten to get matched with the reference color matrices. These problems can be solved with a robust system by the help of advanced machine learning techniques.

# References


[1] Meticulous Market Research Pvt. Ltd.", *Urinalysis Market Worth $2.21 billion by 2027- Exclusive Report by Meticulous Research®*, Mar. 2020. [Online]. Available: https://www.globenewswire.com/news-release/2020/03/26/2006807/0/en/Urinalysis-Market-Worth-2-21-billion-by-2027-Exclusive-Report-by-Meticulous-Research.html (cit. on p. 1).

[2] D. M. Roxe, 'Urinalysis,' 2011 (cit. on p. 1).

[3] A. Hoberman, E. R. Wald, L. Penchansky, E. A. Reynolds and S. Young, 'Enhanced urinalysis as a screening test for urinary tract infection,' *Pediatrics*, vol. 91, no. 6, pp. 1196–1199, 1993 (cit. on p. 1).

[4] J. A. Simerville, W. C. Maxted and J. J. Pahira, 'Urinalysis: A comprehensive review,' *American family physician*, vol. 71, no. 6, pp. 1153–1162, 2005 (cit. on p. 2).

[5] A. K. Yetisen, M. S. Akram and C. R. Lowe, 'Based microfluidic point-of-care diagnostic devices,' *Lab on a Chip*, vol. 13, no. 12, pp. 2210–2251, 2013 (cit. on p. 2).

[6] J. Penders, T. Fiers and J. R. Delanghe, 'Quantitative evaluation of urinalysis test strips,' *Clinical chemistry*, vol. 48, no. 12, pp. 2236–2241, 2002 (cit. on p. 2).

[7] LEARN.PARALLAX.COM, *Urinalysis test strip color chart*. [Online]. Available: https://learn.parallax.com/support/reference/urinalysis-test-strip-color-chart (cit. on pp. 3, 22).

[8] J. Dirks, G. Remuzzi, S. Horton, A. Schieppati, S. A. H. Rizvi *et al.*, 'Diseases of the kidney and the urinary system,' *Disease control priorities in developing countries*, vol. 2, pp. 695–706, 2006 (cit. on p. 5).

[9] *Celebrating 50 years of combur-test® strip technology from roche*. [Online]. Available: https://diagnostics.roche.com/global/en/article-listing/celebrating-50-years-of-combur-test--strip-technology-from-roche.html (cit. on p. 6).

[10] D. Ong and M. Poljak, 'Smartphones as mobile microbiological laboratories,' *Clinical Microbiology and Infection*, vol. 26, no. 4, pp. 421–424, 2020 (cit. on p. 10).

[11] H. Karlsen and T. Dong, 'Smartphone-based rapid screening of urinary biomarkers,' *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 2, pp. 455–463, 2017 (cit. on p. 10).

[12] G. T. Smith, N. Dwork, S. A. Khan, M. Millet, K. Magar, M. Javanmard and A. K. E. Bowden, ‘Robust dipstick urinalysis using a low-cost, micro-volume slipping manifold and mobile phone platform,’ *Lab on a Chip*, vol. 16, no. 11, pp. 2069–2078, 2016 (cit. on p. 10).

[13] R. Yang, W. Cheng, X. Chen, Q. Qian, Q. Zhang, Y. Pan, P. Duan and P. Miao, ‘Color space transformation-based smartphone algorithm for colorimetric urinalysis,’ *ACS omega*, vol. 3, no. 9, pp. 12 141–12 146, 2018 (cit. on pp. 10, 22).

[14] H. Karisen and T. Dong, ‘Illumination and device independence for colorimetric detection of urinary biomarkers with smartphone,’ in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2016, pp. 5184–5187 (cit. on p. 11).

[15] I. P. Kumar, V. H. Gopal, S. Ramasubbareddy, S. Nalluri and K. Govinda, ‘Dominant color palette extraction by k-means clustering algorithm and reconstruction of image,’ in *Data Engineering and Communication Technology*, Springer, 2020, pp. 921–929 (cit. on p. 11).

[16] G. Dong and M. Xie, ‘Color clustering and learning for image segmentation based on neural networks,’ *IEEE transactions on neural networks*, vol. 16, no. 4, pp. 925–936, 2005 (cit. on p. 11).

[17] X.-Y. Zhang, J.-S. Chen and J.-K. Dong, ‘Color clustering using self-organizing maps,’ in *2007 International Conference on Wavelet Analysis and Pattern Recognition*, IEEE, vol. 3, 2007, pp. 986–989 (cit. on p. 12).

[18] L.-P. Chen, Y.-G. Liu, Z.-X. Huang and Y.-T. Shi, ‘An improved som algorithm and its application to color feature extraction,’ *Neural Computing and Applications*, vol. 24, no. 7, pp. 1759–1770, 2014 (cit. on p. 12).

[19] A. Mojsilovic, H. Hu and E. Soljanin, ‘Extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis,’ *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1238–1248, 2002 (cit. on p. 13).

[20] D. J. Bora, A. K. Gupta and F. A. Khan, ‘Comparing the performance of l* a* b* and hsv color spaces with respect to color image segmentation,’ *arXiv preprint arXiv:1506.01472*, 2015 (cit. on p. 13).

[21] F. Mohsen, M. M. Hadhoud, K. Moustafa and K. Ameen, ‘A new image segmentation method based on particle swarm optimization.,’ *Int. Arab J. Inf. Technol.*, vol. 9, no. 5, pp. 487–493, 2012 (cit. on p. 13).

[22] H. G. Adelmann, ‘Butterworth equations for homomorphic filtering of images,’ *Computers in Biology and Medicine*, vol. 28, no. 2, pp. 169–181, 1998 (cit. on p. 16).

[23] H. Chethan and G. H. Kumar, ‘Illumination reduction for camera captured documents using block processing and homomorphic filtering technique,’

in *2011 International Conference on Communication Systems and Network Technologies*, IEEE, 2011, pp. 350–354 (cit. on p. 16).

[24] L. I. Voicu, H. R. Myler and A. R. Weeks, 'Practical considerations on color image enhancement using homomorphic filtering,' *Journal of Electronic Imaging*, vol. 6, no. 1, pp. 108–113, 1997 (cit. on p. 16).

[25] E. S. Gedraite and M. Hadad, 'Investigation on the effect of a gaussian blur in image filtering and segmentation,' in *Proceedings ELMAR-2011*, IEEE, 2011, pp. 393–396 (cit. on p. 17).

[26] R. A. Haddad, A. N. Akansu *et al.*, 'A class of fast gaussian binomial filters for speech and image processing,' *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 723–727, 1991 (cit. on p. 17).

[27] N. Otsu, 'A threshold selection method from gray-level histograms,' *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979 (cit. on p. 18).

[28] M. R. Hassan, R. R. Ema and T. Islam, 'Color image segmentation using automated k-means clustering with rgb and hsv color spaces,' *Global Journal of Computer Science and Technology*, 2017 (cit. on p. 18).

[29] N. Jamil *et al.*, 'Automatic image annotation using color k-means clustering,' in *International Visual Informatics Conference*, Springer, 2009, pp. 645–652 (cit. on p. 19).

[30] B. Sathish, K. B. Shaik, V. Kalist *et al.*, 'Neural network based som for multispectral image segmentation in rgb and hsv color space,' in *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*, IEEE, 2015, pp. 1–6 (cit. on p. 19).

[31] T. Honkela, 'Kohonen network,' *Scholarpedia*, vol. 2, no. 1, p. 1568, 2007 (cit. on p. 19).

[32] P.-N. Tan, *M. steinbach i v. kumar, introduction to data mining*, 2005 (cit. on p. 20).

[33] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä and L. E. Meester, *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media, 2005 (cit. on p. 20).

[34] P. Jayaraj, K. Mithun, G. Gopakumar and U. Jaleel, 'A gpu based virtual screening tool using som,' *International Journal of Computational Biology and Drug Design*, vol. 14, no. 1, pp. 64–80, 2021 (cit. on p. 25).

[35] A. M. Taqi, A. Awad, F. Al-Azzo and M. Milanova, 'The impact of multi-optimizers and data augmentation on tensorflow convolutional neural network performance,' in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, 2018, pp. 140–145 (cit. on p. 25).

[36] M. L. Eaton and D. Kennedy, *Innovation in medical technology: Ethical issues and challenges.* JHU Press, 2007 (cit. on p. 28).