A Technique of Dataset Generation Using Cancerous Images on MATLAB

[1]Mini Puri, [2]Jyoti Mann
[1,2]South Point Institute of Technology and Management, Sonipat, Haryana, India

Abstract
Cancer is one of the most commonly occurring conditions. India has one of the highest rates of cancer in the world . More than 80,000 new cases reported annually across the country. Early detection of cancer is crucial for treatment. Now a days there are various methods of cancer detection. In this paper dataset generation technique is used to find the cancerous image on MATLAB tool. It can improves the accuracy of the detection method.

Keywords
Cancer, Cancer Detection, Dataset, Cancerous Images

I. Introduction

Cancer has a reputation as a deadly disease. Taken as a whole, about half of people receiving treatment for invasive cancer (excluding carcinoma in situ and non-melanoma skin cancers) die from cancer or its treatment. Survival is worse in the developing world, partly because the types of cancer that are most common there are at present harder to treat than those associated with the lifestyle of developed countries. However, the survival rates vary dramatically by type of cancer, and by the stage at which it is diagnosed, with the range running from the great majority of people surviving to almost no one surviving as long as five years after diagnosis. Once a cancer has metastasized or spread beyond its original site, the prognosis normally becomes much worse.

Those who survive cancer are at increased risk of developing a second primary cancer at about twice the rate of those never diagnosed with cancer. The increased risk is believed to be primarily due to the same risk factors that produced the first cancer, partly due to the treatment for the first cancer, and potentially related to better compliance with screening.

Predicting either short-term or long-term survival is difficult and depends on many factors. The most important factors are the particular kind of cancer and the patient's age and overall health. People who are frail with many other health problems have lower survival rates than otherwise healthy people. A centenarian is unlikely to survive for five years even if the treatment is successful. People who report a higher quality of life tend to survive longer. People with lower quality of life may be affected by major depressive disorder and other complications from cancer treatment and/or disease progression that both impairs their quality of life and reduces their quantity of life. Additionally, patients with worse prognoses may be depressed or report a lower quality of life directly because they correctly perceive that their condition is likely to be fatal.

People with cancer, even those who are walking on their own, have an increased risk of blood clots in veins. The use of heparin appears improve survival and decrease the risk of blood clots.

Leukemia is a disease that affects blood-forming cells in the body. It is a cancerous condition characterized by an abundance of abnormal white blood cells in the body. Leukemia begins in the bone marrow and spreads to other parts of the body. Both children and adults can develop leukemia.

Here are four main types of leukemia:

1. Acute myeloid leukemia (AML)
2. Chronic myeloid leukemia (CML)
3. Acute lymphocytic leukemia (ALL)
4. Chronic lymphocytic leukemia (CLL)

Another type of leukemia, hairy cell leukemia, is a rare subtype of chronic lymphocytic leukemia (CLL). It is caused by an increased number of lymphocytes and progresses very slowly. It is called "hairy" because of the appearance of the leukemia cells under a microscope.

The primary differences between the four main types of leukemia have to do with their rates of progression and where the cancer develops. "Chronic" leukemia cells do not mature all the way, so they are not as capable of defending against infections as normal lymphocytes. "Acute" leukemia cells begin to replicate before any immune functions have developed [5].

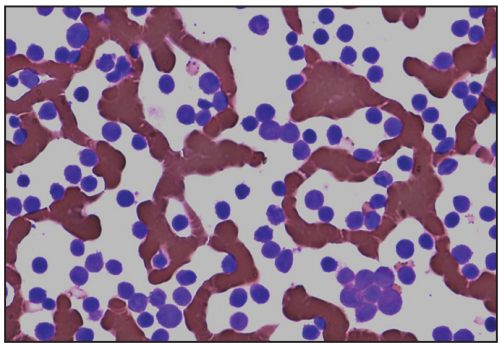

Fig. 1: Cancerous Image

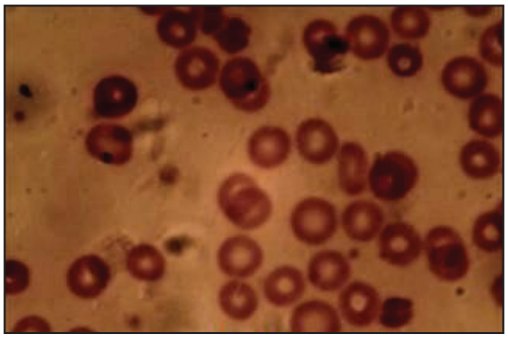

Fig. 2: Cancerous Image

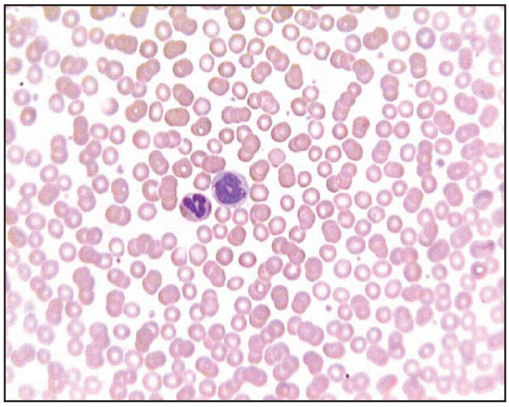

Fig. 3: Non Cancerous Image

II. Methodlogy

The whole implementation is broadly divided into 3 parts:

1. Create the dataset using 4 cancerous images
2. Create the dataset using 8 cancerous images
3. Create the dataset using 15 cancerous images

In that data set there are red, blue and green channel for all three mean, variance and standard deviation. Finally we got the values of MIN and MAX of those three parameters.

III. Tool Used

The MATLAB Language

This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. It allows both "programming in the small" too rapidly create quick and dirty throw-away programs, and "programming in the large" to create complete large and complex application programs. The language features are organized into six directories in the MATLAB Toolbox.

MATLAB software will be used for the whole computation scenario. MATLAB (matrix laboratory) is a numerical computing environment and fourth-generation programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran. Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing capabilities. An additional package, Simulink, adds graphical multi-domain simulation and Model-Based Design for dynamic and embedded systems. In 2004, MATLAB had around one million users across industry and academia. MATLAB users come from various backgrounds of engineering, science, and economics. MATLAB is widely used in academic and research institutions as well as industrial enterprises.

MATLAB is a high-level language and interactive environment for numerical computation, visualization, and programming. Using MATLAB, you can analyze data, develop algorithms, and create models and applications. The language, tools, and built-in math functions enable you to explore multiple approaches and reach a solution faster than with spreadsheets or traditional programming languages, such as C/C++ or Java. We can use MATLAB for a range of applications, including signal processing and communications, image and video processing, control systems, test and measurement, computational finance, and computational biology. More than a million engineers and scientists in industry and academia use MATLAB, the language of technical computing.

It contains hundreds of commands to do mathematics. You can use it to graph functions, solve equations, perform statistical tests, and do much more. It is a high-level programming language that can communicate with its cousins, e.g., FORTRAN and C. You can produce sound and animate graphics. You can do simulations and modeling (especially if you have access not just to basic MATLAB but also to its accessory SIMULINK). You can prepare materials for export to the World Wide Web. In addition, you can use MATLAB, in conjunction with the word processing and desktop publishing features of Microsoft Word, to combine mathematical computations with text and graphics to produce a polished, integrated, and interactive document. A program this sophisticated contains many features and options. There are literally hundreds of useful commands at your disposal. The MATLAB help documentation contains thousands of entries. The standard references, whether the MathWorks User's Guide for the product, or any of our competitors, contain myriad tables describing an endless stream of commands, options, and features that the user might be expected to learn or access.

MATLAB is more than a fancy calculator; it is an extremely useful and versatile tool. Even if you only know a little about MATLAB, you can use it to accomplish wonderful things. The hard part, however, is figuring out which of the hundreds of commands, scores of help pages, and thousands of items of documentation you need to look at to start using it quickly and effectively.

IV. Result and Discussion

Following are the implementation results for the scenario as per given in the previous chapter Methodology.

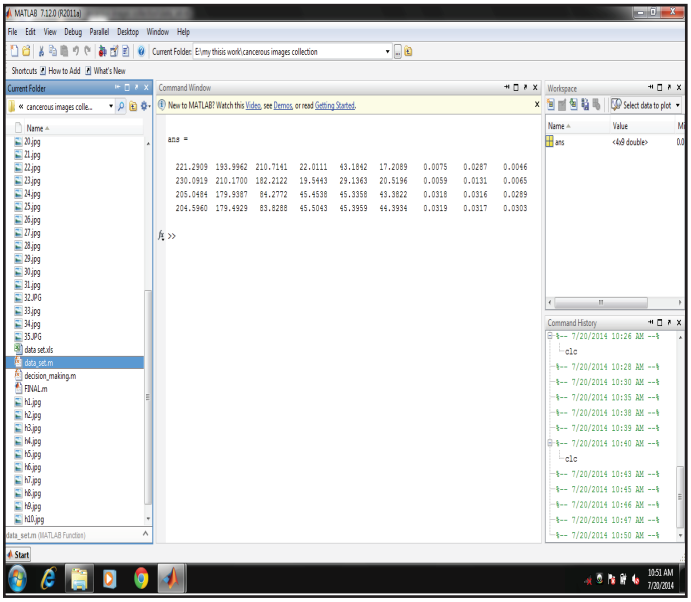

Fig. 4: Data set of 4 images in command window

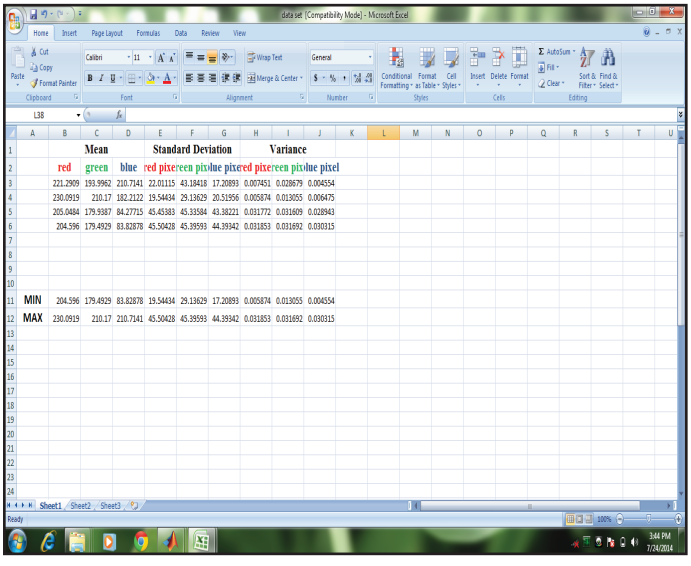

Fig. 5: Data set with MIN and MAX value of parameters

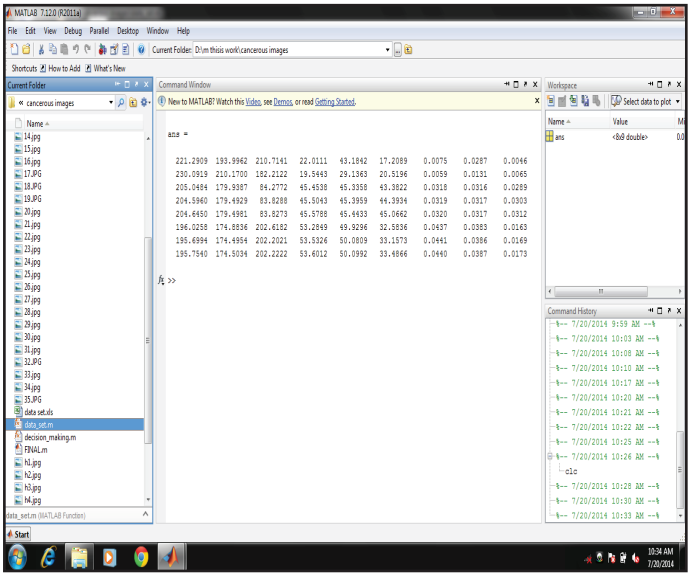

Fig. 6: Data set of 8 images in command window

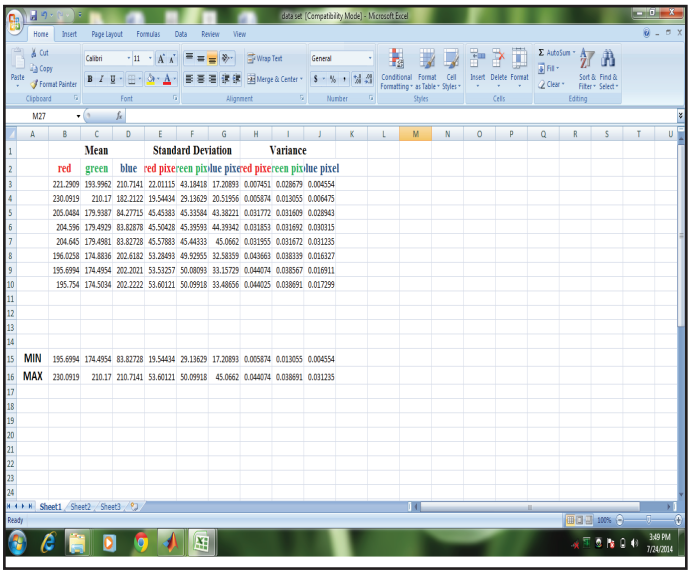

Fig. 7: Data set with MIN and MAX value of parameters

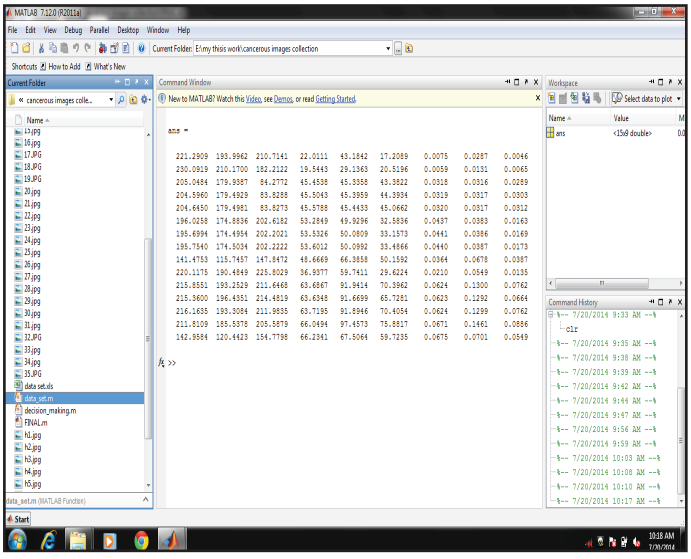

Fig. 8: Data set of 15 images in command window

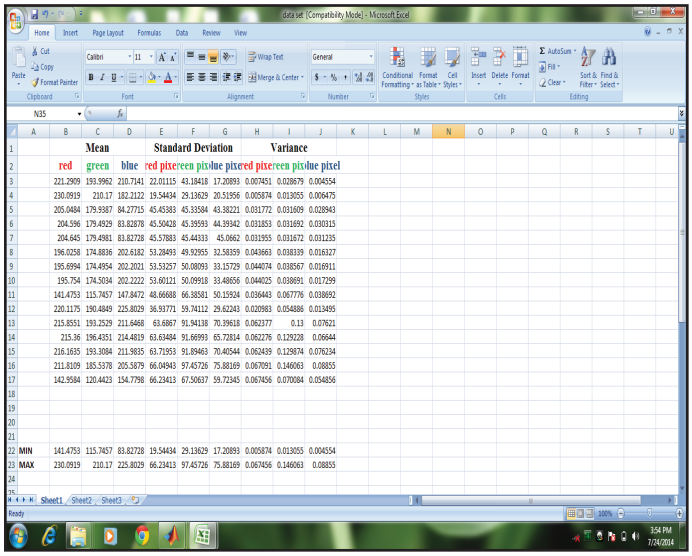

Fig. 9: Data set with MIN and MAX value of parameters

References

[1] Rheingold, Susan; Neugut, Alfred; Meadows, Anna (2003),"156: Secondary Cancers: Incidence, Risk Factors, and Management". In Frei, Emil; Kufe, Donald W.; Holland, James F. Holland-Frei Cancer Medicine (6th ed.). Hamilton, Ont: BC Decker. p. 2399. ISBN 1-55009-213-8. Retrieved 5 November 2009.

[2] Montazeri A (December 2009). "Quality of life data as prognostic indicators of survival in cancer patients: an overview of the literature from 1982 to 2008". Health Qual Life Outcomes 7: 102. doi:10.1186/1477-7525-7-102. PMC 2805623. PMID 20030832

[3] Akl, EA; Kahale, LA; Ballout, RA; Barba, M; Yosuico, VE; van Doormaal, FF; Middeldorp, S; Bryant, A; Schünemann, H (10 December 2014). "Parenteral anticoagulation in ambulatory patients with cancer.". The Cochrane database of systematic

[4] [Online] Available: http://www.en.wikipedia.org/wiki/Leukemia#General_classification

[5] [Online] Available: http://www.cancercenter.com/leukemia/types/

[6] [Online] Available: http://in.mathworks.com/solutions/image-video-processing/

[7] [Online] Available: http://in.mathworks.com/help/images/index.html