



Cancer Detection with an Image Processing Application

Gizem Boyraz¹, Fatma Öztürk², Pınar Kırıcı^{3*}

^{1,2} Istanbul University-Cerrahpasa, Faculty of Engineering, Department of Computer Engineering, İstanbul, Turkey

^{3*} Bursa Uludag University, Faculty of Engineering, Department of Computer Engineering, Bursa, Turkey, pinarkirci@uludag.edu.tr

(1st International Conference on Applied Engineering and Natural Sciences ICAENS 2021, November 1-3, 2021)

(DOI: 10.31590/ejosat.1013525)

ATIF/REFERENCE: Boyraz, G., Öztürk, F., & Kırıcı, P. (2021). Cancer Detection with an Image Processing Application. *European Journal of Science and Technology*, (28), 1442-1446.

Abstract

In this study, image processing algorithms and their performances are examined. In the first part of the study, image processing, color space, functions and frameworks that can be used in C# are studied. As a result of obtaining the necessary information, an algorithm has been written that turns existing images into gray. Then, Median filtering was applied to the image in order to increase the quality of the image. Threshold algorithm was used to clarify the region after filtering. Finally, Sobel edge detection algorithm was used for image processing. Machine learning was used to decide whether the tumor region detected after image processing is cancerous or not. A dataset and NaiveBayes Class are used for machine learning. In summary, the project determines the tumor area with image processing and tells whether the tumor is cancerous with machine learning.

Keywords: Sobel edge detection algorithm, Image processing, Median filtering, Cancer detection

Görüntü İşleme Uygulaması ile Kanser Tespiti

Öz

Bu çalışmada görüntü işleme algoritmaları ve performansları incelenmiştir. Çalışmanın ilk bölümünde görüntü işleme, renk uzayı, C# da kullanılacak fonksiyonlar ve çerçeveler incelenmiştir. Gerekli bilgilerin elde edilmesi sonucunda mevcut görüntüleri griye çeviren bir algoritma yazılmıştır. Daha sonra görüntünün kalitesini artırmak için görüntüye Medyan filtreleme uygulandı. Filtrelemeden sonra bölgeyi netleştirmek için eşik algoritması kullanıldı. Son olarak, görüntü işleme için Sobel kenar algılama algoritması kullanılmıştır. Görüntü işlemeden sonra tespit edilen tümör bölgesinin kanserli olup olmadığına karar vermek için makine öğrenmesi kullanıldı. Makine öğrenimi için bir veri kümesi ve NaiveBayes Sınıfı kullanıldı. Özetle proje, görüntü işleme ile tümör alanını belirler ve makine öğrenmesi ile tümörün kanserli olup olmadığını söyler.

Anahtar Kelimeler: Sobel kenar algılama algoritması, Görüntü işleme, Medyan filtreleme, Kanser algılama

1. Introduction

Along with the developing technology, image processing is used in many fields as well as in medical applications today. With image processing, obtaining information from the image is important in the diagnosis and early diagnosis of diseases. With the increasing incidence of cancer in the world, early diagnosis has become even more important. In this field, image processing enables early diagnosis of diseases by revealing the numerical information in the image that the human eye cannot perceive. In this study, it was aimed to help doctors diagnose cancer with the help of computers.

Image Processing is a method developed to transfer the incoming image to the computer platform and perform some operations, and used to extract some useful information from the obtained image. Today, image processing methods applied by companies in various ways are among the rapidly growing technologies. Image Processing is also the main research area in engineering and computer science departments.

Computer Vision, Optical Sorting, Augmented Reality, Feature Detection, Face Recognition, Lane Departure Warning System, Microscope Image Processing, Morphological Image Processing, Medical Image Processing, Remote Sensing, Non-photorealistic Rendering are image processing applications. Cancer is one of the most common diseases nowadays. Cancer, which catches 14 million people every year and causes 8.2 million deaths in the world; It affects all people regardless of age, gender, language, religion or race. If the similar course of cancer continues, 22 million new cases are expected to emerge in 2030. Estimates have shown that a significant portion of cancer cases that will develop in the coming years will occur in underdeveloped countries [1,13,14,15].

2. Material and Method

2.1. What is Cancer?

All organs in our body are made up of cells. Healthy body cells (except muscle and nerve cells) have the ability to divide. They use these abilities for regeneration of dead cells and repair of injured tissues (inside and outside the body). But these abilities are also limited. they cannot be divided an infinite number. Throughout its life, every cell has a certain number of divisibility. A healthy the cell knows where and how much to divide. On the other hand, cancer cells lose this consciousness, begin to divide uncontrollably and they multiply. Cancer cells accumulate to form tumors (masses). Tumors can compress, infiltrate or destroy normal tissues. If cancer cells separate from the tumor in which they formed, they are transferred to other parts of the body through the blood or lymph. They form tumor colonies where they go and continue to grow. The spread of cancer to other parts of the body in this way is called metastasis.

Benign tumors are not cancer. They do not spread to neighboring areas. Borders is evident. They do not dissolve neighboring tissues. These cells are differentiated into oncogenes (mutation) but it is still possible to predict their origin. They usually do not recur when completely removed.

Malignant tumors are called cancer. They spread to neighboring organs and tissues, and when they encounter bone tissue, they even dissolve it (resorption). Its boundaries are unclear. The cells that make up the malignant tumor are so

differentiated that it is impossible to tell what their origin is. They also spread to distant organs via lymph and blood. Cancers are classified according to the organ in which they begin to form and their appearance under the microscope [6].

2.2. What is Breast Cancer?

Breast cancer means the uncontrolled proliferation of cells that make up the breast tissue. In the nucleus of all cells in the body, there are genes that control the normal growth and health of cells. These genes perform the task of controlling the normal reproduction of cells. These genes can be damaged by various factors such as radiation, environmental factors, microorganisms, etc. If this proliferation does not occur in a certain order, tissue excess occurs, this tissue excess is called tumor. If the cells in this tissue tend to spread and go elsewhere, it is cancer. If they continue to stay where they are, they are called benign tumors [8].

2.3. Symptoms of Breast Cancer

Breast Mass: The most common finding among breast cancer symptoms is palpation of a mass in the breast. Most of these masses are benign tumors. A palpable mass in the breast can be one of 2 different structures. The mass may be a fluid-filled sac. We call this a cyst. Simple cysts are not among the symptoms of breast cancer. Masses that may be a sign of breast cancer are formations filled with a different tissue; We call them solid masses. The best way to distinguish these two different structures is to examine the mass with ultrasound.

Cyst in the breast: Cysts are usually benign tumors of the breast. They range in size from a few millimeters to a few centimeters. It is rare in women under 25 years of age, more common in women approaching menopause.

Solid Masses: Masses that do not contain fluid in the breast and are filled with different cells are called solid masses. Solid masses are more likely to be cancer than cysts. A solid mass detected in the breast of a young woman is not likely to be cancer: The risk of cancer increases with age. The mass detected in the breast is evaluated with mammography and ultrasound, and the possibility of cancer is investigated. It is not possible to make a definitive diagnosis with these methods. However, a definitive diagnosis can be made by examining the part taken from the mass in pathology.

A Palpable Mass Under The Armpit: A palpable mass in the armpit may be a lymph node enlargement due to many reasons; but sometimes it can be the first sign of breast cancer. Therefore, if you have not recently had an infection in your hand or arm, this mass needs to be examined [9].

2.4. Image Processing-Graying the Image

Each pixel value in the image gives the contrast values of that image. All pixel values take a value between 0 and 255. A value of 0 indicates black and a value of 255 indicates white. There are several ways to code in C#. This project was made using Bitmap. The R, B, G values of all pixels of the image are taken and multiplied by the values required to create a gray image. A gray image can also be obtained by adding the R, B, G values obtained in another way and dividing by 3. Turning the image to gray can also be done using the framework. The Apply(Bitmap) method is called using the class Grayscale. This method applies a gray filter to the image .

2.5. Image Thresholding

This function is usually used to create a binary image from a grayscale image. It can also be used with color images. Filters the pixel values of the source image to values that are too large or too small. Mostly used to remove noise in images. There are several threshold types available: Binary, Binary Inverted, To Zero, To Zero Inverted, Truncate [7].

2.6. Image Filtering

Image filtering is often used to remove unwanted signals, called noise, in the image. With filtering, operations such as sharpening the image, revealing certain details, softening the image, sharpening the edge or finding edges are carried out. Filters are usually 3x3 matrices. But its dimensions can be 5x5, 7x7, 9x9, 11x11 [2]. Median filtering was used in the project. In this filtering method, the original sorted pixel is replaced by the median value between its neighbors. This differs from weighted average filters: In weighted average filters, the weighted average of the neighbors is taken, this calculated value is re-centered with the original pixel and the result is found. In the median filter, the neighborhood values are ordered first and then the middle value is taken. This value is considered a direct result [5].

2.7. Edge Finding

Edges in gray-level images can be defined as regions of discontinuities in brightness or as the boundary between two regions where the degrees of brightness differ significantly from each other. This change within the image is obtained using the gradient. Edges can be extracted from these monochrome images using gradient-based algorithms. Another alternative is to linearly combine the edges for each color component for each pixel [3].

C# code has been written for sobel edge detection algorithm, which is one of the edge detection methods. Image is grayed out first so that sobel edge detection algorithm can be used. In two 3x3 matrices, the image is displayed in a pixel matrix. The logic we use here is to create a negative on the left and a positivity on the right and detect the change from left to right. The ProcessImage(Bitmap) method from the FastCornersDetectorclass can be used directly using the Sobel Edge Detection Algorithm framework. It is an algorithm that searches for edges in an image [4].

3. Results and Discussion

Accurate diagnosis is very important in cancer treatment. Suspicious cell sections taken are examined under a microscope by the pathologist or doctor, and the diagnosis of the disease is made definitively. According to the type of disease, the most appropriate treatment type is determined by the doctor and treatment is started.

In this study, an algorithm will be developed for the diagnosis of cancerous region by image processing and this algorithm will be implemented in C# programming language. In this algorithm, steps of graying the image, filtering, image thresholding, edge finding will be followed respectively. The main purpose of the algorithm is to diagnose the cancerous area quickly and accurately.

Machine learning is a system that investigates the study and construction of algorithms that can learn as a structural function and make predictions on data. Such algorithms work by building a model to make data-based predictions and decisions from *e-ISSN: 2148-2683*

sample inputs, rather than strictly following static program instructions [10]. Machine learning falls into three main categories: supervised learning, unsupervised learning, and learning to learn. The learning method used in the study is the classification method under the heading of supervised learning. If the Classification Data is used to predict a category, supervised learning is called classification [12].

The classification method was applied to the study according to the data set. It is divided into 2 different classes as benign and malignant. After classification, 5 different attributes were used to determine the class: age, tumor size, menopause, X-Ray. A data set of 286 different patients, 201 of whom were benign and 85 were malignant, was used.

For breast cancer detection, breast cancer dataset developed by Matjaz Zwitter & Milan Soklic, in University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia, was used. The data set was requested via e-mail on 03/01/2018 from <https://archive.ics.uci.edu/ml/datasets/breast+cancer>. The dataset is composed of ten types of data : class, age, menopause, tumor-size, inv-nodes, node-caps, deg-malig, breast, breast-quad and irradiat. The dataset includes the data of 286 women and 85 of them have cancer. There are 10 attributes in the data set used. These are,

TABLE I. DATA SET

1. Class: no-recurrence-events, recurrence-events
2. age: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99.
3. menopause: lt40, ge40, premeno.
4. tumor-size: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59.
5. inv-nodes: 0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, 27-29, 30-32, 33-35, 36-39.
6. node-caps: yes, no.
7. deg-malig: 1, 2, 3.
8. breast: left, right.
9. breast-quad: left-up, left-low, right-up, right-low, central.
10. irradiat: yes, no.

It consists of 5 different categories that allow us to determine the class: age, tumor size, menopause, X-Ray. There are several options to choose from for each attribute. The data set contains the information of 286 women separately. While 201 samples in the Data Set are non-cancerous, 85 samples are stated to be cancerous in the content.

In the study, Naive Bayes was used as a learning algorithm. For the algorithm to work according to the data set, it is written

according to the model consisting of 2 classes and 5 data. With the learn() method, which is used with the NaiveBayes Class by taking the data from the inputs and outputs arrays, the data is learned by the system.

After the learning process is completed, the decision phase is started. The values entered from the interface are taken as parameters and the decision making process is applied. As a result, it returns 2 results, either benign or malignant, according to the data it learns.

In this study, cancer cell examination was performed using image processing techniques as given in figures 1-7. First a code snippet was written that turns existing color images gray.

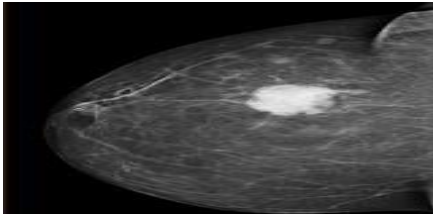


Fig. 1. The grayed out image[16]

Afterwards, Median filtering was used for the quality of the images.

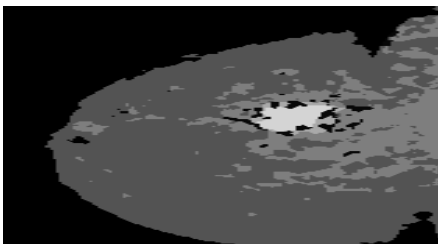


Fig. 2. Filtered image

Threshold algorithm was used to clarify the region after median filtering.



Fig. 3. Thresholding image

Finally, Sobel edge detection algorithm was used for image processing.



Fig. 4. Image with Sobel Edge Detection Algorithm algorithm applied

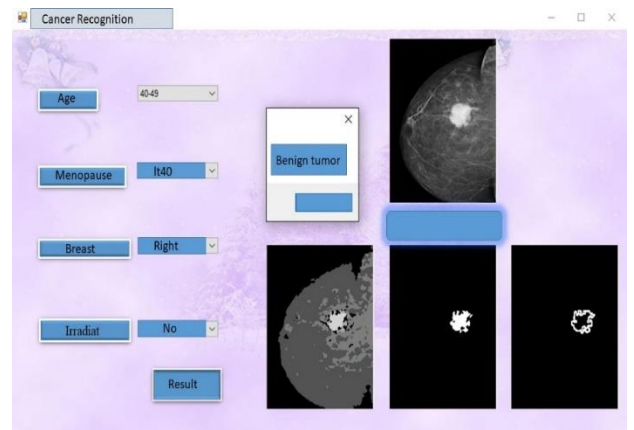


Fig. 5. Screen Output 1

If the algorithm is applied for figure 6, the screen output will be as in figure 7.

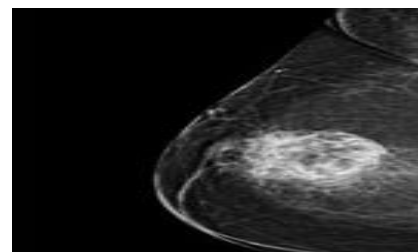


Fig. 6. The grayed out image [17]

Machine learning was used to decide whether the tumor region detected after image processing is cancerous or not. Taught the project using dataset[] and NaiveBayes Class for machine learning. As a result, the project determines the tumor area with image processing method and tells whether the tumor is cancerous or not with machine learning as given in figure 5 and 7.

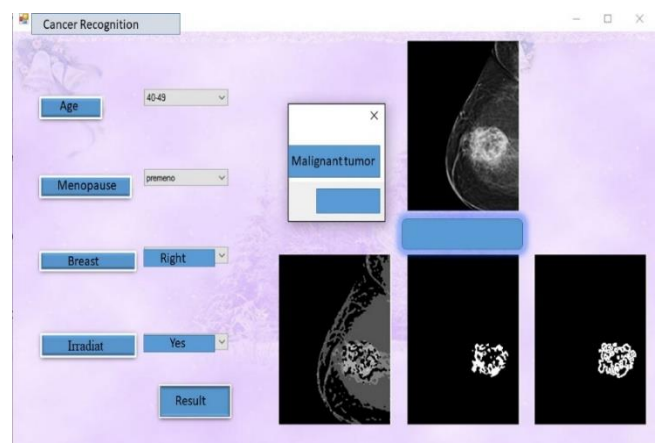


Fig. 7. Screen Output 2

4. Conclusions and Recommendations

Image processing is also used in medical applications today. With image processing, obtaining the expected information from the image is important in the diagnosis and early diagnosis of diseases. With the increasing incidence of cancer in the world, early diagnosis has become even more important. In this field, image processing enables early diagnosis of diseases by revealing the numerical information in the image that the human eye cannot

perceive. In this study, it was aimed to assist doctors in the diagnosis of cancer with the help of computers. In this study, an algorithm was written that turns existing images into gray. Then, Median filtering was applied to the image in order to increase the quality of the image. Threshold algorithm was used to clarify the region after filtering. Finally, Sobel Edge Detection algorithm was used for image processing. Machine learning was used to decide whether the tumor region detected after image processing is cancerous or not. The project determines the tumor area with image processing method and tells whether the tumor is cancerous or not with machine learning.

References

1. Kaya, P.: 2017, Türkiye'de Her Yıl 163 Bin Kişi Kanser Tanısı Alıyor! İşte Kanser İstatistikleri, Medikal Akademi, <https://www.medikalakademi.com.tr/turkiyede-her-yil-163-bin-kisi-kanser-tanisi-aliyor-iste-kanser-istatistikleri/>, last accessed 2017/12/21.
2. Siyah, B.: 2012, "Görüntü Filtreleme Uygulamaları Ve Amaçları Matlab", <http://www.bulentsiyah.com/goruntu-filtreleme-uygulamaları-ve-amaçları-matlab/>, last accessed 2018/1/7.
3. Evans, A. N., Liu, X. U.: A Morphological Gradient Approach To Color Edge Detection. In: Ieee Trans. On Image Processing, 15(6), 1454-1463, (2006)
4. Cesar, S.: 2009, The accord.net image processing and machine learning framework, <http://accord-framework.net/intro.html>, last accessed 2018/1/7.
5. Bulu, H., Alpkoçak, A.: Tıbbi Görüntüler İçin 3 Boyutlu Bölütleme Algoritmalarının Karşılaştırılması. In: 15. Sinyal İşleme Uygulamaları Kurultayı Bildiriler Kitabı, Eskişehir, 2007
6. Kutluk, T., Kars, A.: Kanser Konusunda Genel Bilgiler, Ankara, 1992, 10040932
7. Doğan, T., Sert, E., Taşkın, D.: Araç Destek Sistemleri İçin Kuş Bakışı Görüntü Dönüşümü. In: Akademik Bilişim 2013 – XV. Akademik Bilişim Konferansı Bildirileri, 23-25 January 2013, Antalya
8. Karateke, A.: Meme Kanseri Belirtileri Nelerdir, <https://www.ateskarateke.com/meme-kanseri-belirtileri-nelerdir>, last accessed 2018/1/10
9. Türkiye Meme Vakfı, <http://www.memekanseri.org.tr/meme-sağlığı/meme-kanseri-belirtileri/>, last accessed 2018/1/11
10. Alkan, M. A.: Makine Öğrenimi Nedir?, <https://www.endustri40.com/makine-ogrenimi-nedir/>, last accessed 2018/1/12
11. Makine Öğrenimine Giriş, <https://www.hpe.com/tr/tr/compute/hpc/deep-learning.html?jumpid=ps>, last accessed 2018/1/15
12. Ericson, G.: Microsoft Azure Machine Learning Studio İçin Machinelearning Algoritması Kopya Sayfası, Microsoft, Docs.Microsoft.Com/Tr-Tr/Azure/Machine-Learning/Studio/Algorithm-Cheat-Sheet, last accessed 2018/1/18
13. Kartal, E.: Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri Ve Kardiyolojik Risk Değerlendirmesine İlişkin Bir Uygulama, Thesis, Istanbul University-Cerrahpaşa, Enformatic department, June 2015
14. Sahami, M.: Learning Limited Dependence Bayesian Classifiers. In: KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, (1996)
15. W. Dai Et Al. Transferring Naive Bayes Classifiers For Text Classification. In: Proceedings Of The National Conference On Artificial Intelligence. London; Aaa Press; Mit Press; 1999.
16. Breast care center, <https://www.toplinemd.com/breast-care-center-of-miami/blog/extensive-lymph-node-removal-benefits/>, last accessed 2018/1/15
17. Muzio, D., Jones, B.: Fat Necrosis(Breast), <https://radiopaedia.org/articles/fat-necrosis-breast-2>, last accessed 2018/1/15