

Z. ÖZCAN

M.S. Thesis

2015

OPTIMUM BLEND OF FRACTAL METHODS FOR
AUTOMATIC MALIGNANCY DETERMINATION IN
DERMOSCOPY IMAGES

ZÜHAL ÖZCAN

IŞIK UNIVERSITY

2015

OPTIMUM BLEND OF FRACTAL METHODS FOR AUTOMATIC
MALIGNANCY DETERMINATION IN DERMOSCOPY IMAGES

ZÜHAL ÖZCAN

B.S., Industrial Engineering (IE), 2014

M.S., Industrial Engineering Operations Research (IEOR), 2015

Submitted to the Graduate School of Işık University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Industrial Engineering Operations Research

IŞIK UNIVERSITY

2015

IŞIK UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

OPTIMUM BLEND OF FRACTAL METHODS FOR AUTOMATIC
MALIGNANCY DETERMINATION IN DERMOSCOPY IMAGES

ZÜHAL ÖZCAN

APPROVED BY:

Assist. Prof. Dr. Burak Çavdarođlu
(Işık University-Industrial Engineering)
(Thesis Supervisor)

Assist. Prof. Dr. Tankut Atan
(Işık University-Industrial Engineering)

Assoc. Prof. Dr. Olcay Taner Yıldız
(Işık University-Computer Engineering)

APPROVAL DATE: 10 / 08 / 2015

OPTIMUM BLEND OF FRACTAL METHODS FOR AUTOMATIC MALIGNANCY DETERMINATION IN DERMOSCOPY IMAGES

Abstract

The most important step is early and effective diagnosis in the treatment of melanoma which constitutes the biggest part of skin cancers results in death. At the present time, dermatologists take dermoscopy images, visually examine these images and draw the lesion borders manually for a future reference. With the naked-eye, it is not easy to recognize compartments and tissue structures. Additionally, this procedure is also tedious, biased, and failure rate is high. Moreover, even with the help of dermoscopy, 70% of melanoma claims are still a false-negative diagnosis. This is the motivation for computer assisted diagnosis (CAD) techniques to help dermatologists to reduce possible unlikelihood, to standardize the results and also to speed up the process. The techniques which are developed by Fractal Methods determine irregularities on the lesion borders. Our aim is to designate which Fractal Methods are more effective on determining malignant lesions to minimize false-negative, false-positive and total-false diagnosis. For this purpose, we develop four different mixed integer programming (MIP) classification models, and then applied these models on the dataset of 100 patients. First, we determine the optimum usage rate of Fractal Methods for each classification model using randomly selected 50 patients (training sample). Later, true diagnosis performance of each classification model is evaluated using the remaining 50 patients (testing sample) and the optimum usage rate of Fractal Methods which is already found with the “training sample”. It is observed that the optimum usage rate of fractal methods gives 80% success rate in the best case scenario, and we obtained a success rate of 73.4%, on average, when we perform repeated tests using the optimum usage rate of fractal methods.

DERMOSKOPİDE KÖTÜ HUYLU TÜMÖRLERİN SAPTANMASINDA OPTİMAL ORANSAL KIRILMA METODLARI

Özet

Ölümlerle sonuçlanan deri kanseri vakalarının en büyük kısmını oluşturan melanomun tedavisinde en önemli adım etkin ve erken teşhistir. Dermoskopi, deri kanserinin erken teşhisi amacıyla dermoskop adı verilen cihazlarla deride bulunan koyu renkli lezyonların incelenmesidir. Yöntem, şüpheli lezyonlarda hekimin cerrahi girişim kararına yardımcı olmayı ve lezyon sınırlarındaki düzensizlikleri gözlemleyerek iyi huylu lezyonları kötü huylu lezyonlardan ayırmayı hedefler. Günümüzde dermatologlar, dermoskop görüntülerini çıplak gözle inceler ve lezyon sınırlarını elle çizer. Ancak çıplak gözle doku yapılarındaki renk, yoğunluk, büyüklük farklılıklarını ve sınırlardaki düzensizlikleri ayırt etmek zordur. Ayrıca bu işlem kişiden kişiye değişkenlik gösteren, hata oranı yüksek, tekrarlanamayan sonuçlar doğurur. Dermoskopi ile yapılan melanom teşhislerinde %70'e kadar yanlış-negatif hatası gözlemlenmiştir. Melanom teşhislerinin göreceliğini azaltmak, sonuçları standardize etmek ve yanlış teşhis oranlarını minimize etmek için dermatologlara yardımcı, bilgisayar destekli tanı teknikleri geliştirilmiştir. Oransal Kırılma (Fractal) Metodları kullanılarak geliştirilen bu teknikler, lezyonların sınırlarındaki simetri bozukluklarını ve düzensizlikleri tespit eder. Hedefimiz, lezyon sınırlarının bilgisayar destekli çizimlerine uygulanan 11 Oransal Kırılma Metodundan hangilerinin kötü huylu lezyonların saptanmasında daha etkili olduğunu belirlemek ve bu sayede hatalı teşhis oranlarını minimuma indirmektir. Bu amaçla karışık tam sayı doğrusal programlama kullanarak 4 farklı sınıflandırma modeli geliştirilmiş ve daha sonra bu modeller 100 hastalı bir veri kümesinde uygulanmıştır. İlk önce, sınıflandırma modellerinden her biri için rastgele seçilen 50 hasta (öğretme grubu) kullanılarak Oransal Kırılma Metodlarının optimum kullanım oranları tespit edilmiştir. Daha sonra kalan 50 hasta (test grubu) ve bulduğumuz Oransal Kırılma Metodlarının optimum oranları kullanılarak sınıflandırma metodlarının her birinin doğru teşhis performansları ölçülmüş, en iyi senaryoda doğru teşhis oranı % 80 olarak hesaplanmış ve ortalama doğru teşhis performansı %73.4 olarak elde edilmiştir.

To my family

Acknowledgements

There are many people who helped to make my years at the graduate school most valuable. First and foremost, I would like to express my profound gratitude to my project supervisor, Assist. Prof. Burak avdaroglu, for involving me to his study, his great concern, invaluable support, patience and continuous guidance to prepare this project. Having the opportunity to work with him over the last year was intellectually rewarding and fulfilling. I also would like to thank Sinan Kokara and Tansel Hali for sharing their broad research experience and offering their valuable assistance in the preparation of this study.

Additionally, I am extremely grateful to the foundation, Iık University and especially Industrial Engineering Department Staff, who have always been helpful during my whole master and bachelor education. I would also like to thank to my graduate student colleagues who helped me all through the years full of class work and exams.

Also, I am extremely grateful to my family who has always trusted me and has great role on my personal development against to difficulties of life. I thank my parents Mehmet zcan and Aye zcan and my sister Selenay zcan for their patience and endless encouragement. There were many things to learn from them, and there will always be.

As last one, I thank to TBİTAK for its financial support implicitly during my master education.

Table of Contents

Abstract	ii
Özet	iii
Acknowledgements	v
Table of Contents	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF GRAPHS	x
CHAPTER 1 INTRODUCTION	1
1.1 Problem Statement and Research Objective.....	1
1.2 Research Objective	5
1.3 Limitations of the Research.....	6
1.4 Outline of the Thesis	6
CHAPTER 2 BACKGROUND	7
CHAPTER 3 FRACTAL METHODS	11
3.1 Fractal	11
3.2 Fractal Methods	12
3.2.1 Dilation	13
3.2.2 Euclidean Distance Map	13
3.2.3 Box Counting.....	14
3.2.4 Fast.....	14
3.2.5 Fast (Hybrid).....	15
3.2.6 Parallel Lines	15
3.2.7 Corner (Count).....	16

3.2.8 Corner (Perimeter)	16
3.2.9 Cumulative Intersection	16
3.2.10 Mass Radius (Long)	17
3.2.11 Mass Radius (Short)	17
3.3 Relevance and Impact to Diagnosis & Treatment	18
CHAPTER 4 LESION BORDER IRREGULARITY CLASSIFICATION	
PROBLEM (LBICP)	19
4.1 Methodology	20
4.2 MIP Classification Models	22
4.2.1 Maximizing the Minimum Between-group Distance	23
4.2.1.1 Determining Usage Rate of Fractal Methods Using Training Sample	23
4.2.1.2 Determining Melanoma Using Testing Sample	25
4.2.2 Minimizing the Total Within-group Distances	26
4.2.2.1 Determining Usage Rate of Fractal Methods Using Training Sample	26
4.2.2.2 Determining Melanoma Using Testing Sample	27
4.2.3 Maximizing the Total Between-group Distances	28
4.2.3.1 Determining Usage Rate of Fractal Methods Using Training Sample	28
4.2.3.2 Determining Melanoma Using Testing Sample	29
4.2.4 Minimizing the Maximum Within-group Distance	30
4.2.4.1 Determining Usage Rate of Fractal Methods Using Training Sample	30
4.2.4.2 Determining Melanoma Using Testing Sample	31
CHAPTER 5 DATA ANALYSIS AND RESULTS	33
5.1 Dataset	33
5.2 Computational Experiments	34
CHAPTER 6 CONCLUSION	51
6.1 Summary of the Research	51
6.2 Opportunities for Future Work	53
REFERENCES	54
Curriculum Vitae	57

LIST OF FIGURES

Figure 1.1: Polarized light dermatoscope, Immersion oil dermatoscope, Image taken, respectively	3
Figure 1.2 Dermoscopy image of ROI, Dermatologist drawn border, Computerized border detection, respectively	4
Figure 3.1: Basic fractal shapes	12
Figure 3.2: Dilation method application. A: Original image, B: Dilation method applied.....	13
Figure 3.3: Euclidian Distance Map fractal values of different shapes	14
Figure 3.4: Koch Snowflake Box Counting.....	14
Figure 3.5: Fast FD method	15
Figure 3.6. Parallel Lines FD method	16
Figure 3.7 Cumulative Intersection FD method.....	17
Figure 3.8. Changing centers of Mass Radius (Long) method	17
Figure 3.9. Mass Radius (Short) FD method	18
Figure 4.1. The methodology which is used in models.....	21
Figure 5.1: Sample skin cancer images showing borders delineated by the dermatologist (red) and a CAD based method, BD-DBSCAN (blue)..	33
Figure 5.2 False-negative, false-positive, true-negative, true positive	35

LIST OF TABLES

Table 5.1 False-negative, false-positive and total-false results for 4 models and 4 random groups.....	36
Table 5.2 Model 2 results for 10 random groups.....	38
Table 5.3 Used fractal methods in first 10 random groups.....	40
Table 5.4 20 random scenarios for 15-35 grouping.....	42
Table 5.5 Fractal methods for additional 20 random sets.....	44
Table 5.6 20 random scenarios for 50 grouping.....	45
Table 5.7 Fractal methods for last 20 random groups.....	47
Table 5.8 Fractal methods with their fractions for best random sample.....	48
Table 5.9 20 last random scenarios to determine melanoma.....	49
Table 5.10 Average, min., max. values for last 20 Random Samples.....	50

LIST OF GRAPHS

Graph 5.1. Total false values for first 4 random samples	37
Graph 5.2 Total false values for 10 random samples for model-2.....	39
Graph 5.3 Total false values for 20 more random samples.....	43
Graph 5.4 Total false values for last 20 random samples	46
Graph 5.5 Used fractal methods with their weights for the best random sample result	48
Graph 5.6 Total false values for last 20 Random Samples	50

CHAPTER 1

INTRODUCTION

1.1 Problem Statement and Research Objective

Dermatology is the branch of medicine that focuses on the diagnosis and treatment of diseases of the skin, hair and nails. It also focuses on maintaining the health of the skin. Dermatologists are medical doctors who train in this area for many years, making them the experts in all things related to skin, hair and nails [1]. Skin diseases increase due to changing environment conditions and modern life styles day by day. With this increment, research on the diagnosis and treatments of skin diseases also increases because the most common form of cancer types is skin cancer [2]. The most important step in treatment of skin cancers is early diagnosis. Dermatologists use a handy tool which is named as dermatoscope for taking high resolution images, analyzing dark colored lesion on skin, or diagnosing melanoma and other pigmented lesions. This process is known as dermoscopy[3]. Images taken via dermatoscope are called dermoscopy images. Dermoscopy is now a well-established diagnostic tool to improve the clinical recognition of a broad spectrum of different skin disorders. Skin cancer detection is the most important indication of dermoscopy. There is evidence that the use of dermoscopy reduced false-negative diagnosis rate in recent years. Thus, recent skin cancer guidelines promote the use of dermoscopy in skin cancer screening and diagnosis [4].

The deadliest form of skin cancer is melanoma. Melanoma is a cancer that begins in melanocytes, melanin-producing cells located in the bottom layer of the skin's epidermis. Other name for this cancer is malignant melanoma. Melanoma is much less common than basal cell and squamous cell skin cancers. Like the other cancer types, melanoma is curable in its early stages, but it is much more likely than other cancer types to spread to other parts of the body if not caught early, so it is far more dangerous and deathly. According to statistics, 15% of melanoma cases are fatal [5,6]. In addition to this, women 25-29 years of age are the most commonly affected group from melanoma [7]. Although melanoma accounts for only 4% of all skin cancers [8], it is the cause of 75% of skin cancer related deaths [9]. Accurate diagnosis of melanocytic lesions is amongst the most difficult problems for dermatologists since definitive diagnosis requires biopsy, sampling of skin tissues for examination, which has high risk of metastasis, the spread of the cancer from skin to other organs. Moreover, misdiagnosis of these lesions results in one of the causes of medical malpractice for dermatologists. Even with the help of dermoscopy, 70% of melanoma claims are still a *false-negative* diagnosis [10] mainly because dermatologists' risk averse attitude prevent them from making impetuous positive-diagnosis, and possible surgical biopsy. The focus of this research is to aid dermatologists in the stage of melanoma diagnosis by presenting them some analytic methodologies that are capable to diagnose melanoma more accurately with less risk.

Lesion border irregularity is one of the criteria that could serve to distinguish malignant lesions from benign lesions. However, in current practice, physicians draw lesion borders manually and assess irregularity through their observations. This is subjective, erroneous, and not reproducible due to inter- and intra-observer variability.

One of the high priorities of health care professionals is the ability to diagnose pigmented skin lesions, especially melanocytic neoplasms with high accuracy. Pathologists, dermatologists, and dermatopathologists are in continuous study for objective, reproducible criteria that could serve to distinguish benign from malignant melanocytic proliferations in order to accurately predict biologic outcomes for patients undergoing skin biopsy.

In recent years, dermoscopy has revolutionized disease diagnosis, grading, and prognosis for skin cancer. Dermatoscopes take high resolution images for region of interest (ROI) by analyzing dark colored lesion on skin or diagnosing of melanoma and other pigmented lesions (See Figure 1.1) [3].

These images provide wealth of useful information for computer-assisted diagnosis. Although using dermoscopy images opens the way for more objective analysis of cases, fast and accurate diagnosis in melanocytic lesions is still amongst the most challenging problems for pathologists. Misdiagnosis of these lesions is dangerous and results in over or under treatment of patients. The primary causes of errors include tenuous dermatologic changes that remain undetected to the naked-eye and lack of experience of the observers.



Figure 1.1: Polarized light dermatoscope, Immersion oil dermatoscope, Image taken, respectively

In current practice, dermatologists take dermoscopy images, visually examine them and draw the lesion border manually for a future reference. Even with dermatoscopes, it might be difficult to recognize compartments and tissue structures such as glands based on color, intensity, size, and borders. In addition to this, this process is tedious and prone to intra- and inter-observer variability, which results in limited statistical confidence and low reproducibility [11]. It is also known that manual recognition of color and intensity by dermatoscopes is biased depending on color distributions of a given image [12]. Moreover, delineated lesion border drawn by different dermatologists may not be the same. Sometimes this unlikelihood reaches up to 24% [13].

This situation rises the motivation for computer assisted diagnosis (CAD) techniques to help dermatologists to reduce possible unlikelihood, to standardize the results by alleviating inter- and intra-observer variations, and also to speed up the process [14]. Figure 1.2 illustrates an exemplary dermoscopy image of a region of interest (ROI), manually dermatologist drawn border image, and CAD based algorithmic border detected image, respectively. The third image (CAD image) can capture the lesion border in more detail compared to the second image (dermatologist drawn border image) and it depicts the first image (dermoscopy image) more accurately. With second image, it is much harder to follow the deviation from a smooth shape.

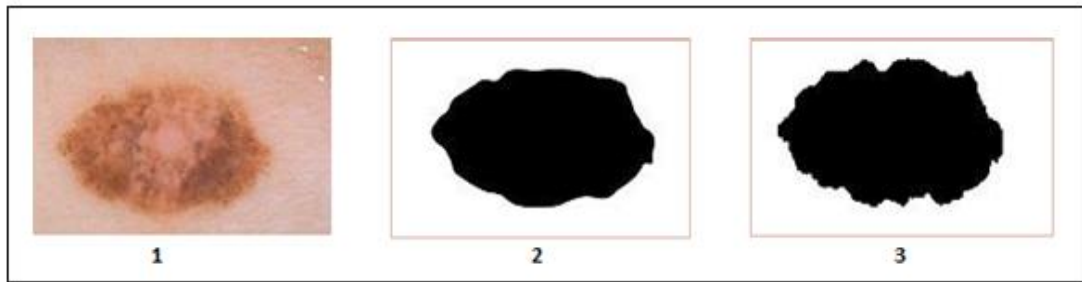


Figure 1.2 Dermoscopy image of ROI, Dermatologist drawn border, Computerized border detection, respectively

In dermatology, as American Cancer Society stated [6], cancer cells grow irregularly so that borders of melanocytic lesions have some shape irregularities. Even if CAD images are used for the identification of border irregularities, some analytical methods are still needed to measure the shape irregularities on CAD drawn lesion border areas and determine possible malignancy of ROI.

1.2 Research Objective

A biopsy is a medical procedure that involves taking a small sample of tissue so that it can be examined under a microscope easily [6]. Biopsies can be used to identify abnormal cells and to help identify a specific type of condition. Where a condition has already been diagnosed, a biopsy can be used to measure how severe it is or which stage it is at. Biopsy is seen as the most accurate way in detection of melanoma, but it is dangerous since it fastens the spread of the malignant lesions through the body which is known as metastasis. Doctors and researchers have noted that biopsy of a tumor can cause seeding, or spread of cancer cells along the path of the needle track at the biopsy site. Additionally, biopsy is not only dangerous, but also an expensive procedure. Therefore, biopsy is usually used as the last resort to diagnose the suspected melanoma cases.

Our ultimate target is to minimize false-negative, false-positive and total-false diagnosis of melanoma as much as possible by using means other than medical procedures (biopsy), such as computer assisted diagnosis (CAD) techniques and mathematical programming methodologies. Despite there are so many studies related to the diagnosis of melanoma, to the extent of our knowledge, none of them has utilized quantitative methods such as mathematical modeling and computer assisted decision systems to diagnose melanoma cases.

The framework of our study is founded upon Kockara et al. [15]. In this article, authors first employed CAD based method for automated lesion border delineation. Then, they implemented different quantitative methods, which are referred as fractal calculation methods, on computerized border detections to quantitatively measure ROI's irregularity. Our main objective is to decide which blend of fractal calculation methods described in Kockara et al. [15] is more accurate for detection of malignancy in dermoscopy images.

1.3 Limitations of the Research

Our study is based upon the computerized border detection and fractal calculation methods discussed in Ercan's study [16]. Although there exists several other fractal calculation methods in the literature, we restrict our attention only on the ones covered in this study. The parametric values of the fractal methods measure the deviation from a smooth shape, i.e. the amount of roughness on the shapes. When these values increase for a lesion border, it means that roughness and correspondingly the possibility to be malignant melanoma also increases. The models and methods we have developed in the thesis assume that parametric values of fractal calculation methods accurately reflect the magnitude of irregularity in lesion borders and are known a priori with certainty.

1.4 Outline of the Thesis

The remainder of this thesis is organized as follows. A review of the existing literature relevant to our work is presented in Chapter 2. Chapter 3 discusses what the Fractal Methods are and examines the methods separately. In Chapter 4, we present our mathematical models that are able to distinguish patients who have malignant lesions from the patients who have benign lesions. Chapter 5 presents the data which we used on our models and results of these models. The best performing model and the best blend of fractal methods are discussed in this chapter. Chapter 6 closes the thesis with conclusions, contributions and future research directions.

CHAPTER 2

BACKGROUND

Since 18th century, there has been abundant work on the dermatology. Willett [17] emphasizes that first dermatology schools were built in 18th-century France for the treatments especially for ancient diseases about skin such as syphilis. Rigel et al. [18] state that dermatologic diseases and especially melanoma has increased in recent years substantially. This trend constitutes the reason why effective educational programs to improve public awareness for early detection become more important than ever [19]. As mentioned earlier, melanoma which begins in melanocytes constitutes the largest part of skin cancers that results in death [1]. Jayaraman et al. [2] state that early detection is important on melanoma to reduce the diseases' fatality rate. In this study, Jayamaran et al. [2]'s target is to distinguish malignant melanoma from benign pigmentations. First, they apply image pre-processing to remove insignificant details such as hairs or effects of air. Then, textural and statistical features obtained are used as neural network classifier input [2]. According to Troxel [10], from 1995 through 2011, only 46 of 362 claims were reported as positive diagnosis of melanoma. However, the article states that 70% of the negatively-diagnosed claims were actually false-negative diagnoses. Mulrane et al. [11] and Conway et al. [12] also states that the diagnosis process is tedious, has low reproducibility and biased on color distributions of a given image. Due to these difficulties in early diagnosis of melanoma, growing motivation for computer assisted diagnosis (CAD) techniques has emerged to help dermatologists to reduce possible unlikelihood, to standardize the results by alleviating inter- and intra-observer variations, and also to speed up the diagnosis process [14].

Among all the studies about dermatology in the literature for a better diagnosis, none of them have utilized mathematical modeling or computerized decision support systems. For instance, Jayaraman et al. [2] study melanoma detection by using dermoscopy images. They have extracted various textural and statistical features using Gray Level Co-occurrence Matrix (GLCM). After that, these attributes are used as stimulants to an Artificial Neural Network (ANN) classifier to separate cancerous and non-cancerous samples. We try to solve the diagnosis problem of skin cancer with computerized systems like Jayaraman et al. [2]'s study. However, different from their work, we prefer to use fractal methods for the classification of melanoma and benign lesions. Lesne [20] states fractals are geometric shapes that have a structure. Berube et al. [21] point out that particles are characterized by their boundary fractal, and then define three fractal methods: box counting, dilation and euclidean distance mapping. Similar to Berube et al. [21], Allen et al. [22] also study fractal methods. In their work, they first divide fractal methods into two groups; vector and matrix based methods. Vector-based methods are then divided into four subgroups which are exact, fast, hybrid and faena. Similarly, matrix-based methods are divided into five subgroups; mosaic amalgamation, lattice interception, dilation method, blanket algorithm, displacement method and distance transform method. In our study, we also use *fast* and *hybrid* methods of vector-based methods, and *dilation* method of matrix-based methods. Allen et al. [22] also claims that the dilation method is the most commonly used matrix-based method. Other than these fractal methods, Cornforth et al. [23] conduct studies on *mass radius* technique, which is now a well-known fractal method.

We use the same data set, which is also used in the analysis conducted in three articles: Kockara et al. [25], Kockara et al. [27], and Kockara et al. [28]. Kockara et al. [25] compare two approaches for automatic border detection in dermoscopy images: density based clustering (DBSCAN) and Fuzzy C-Means (FCM) clustering algorithms. Each approach is examined on a set of 100 dermoscopy images whose manually drawn borders by a dermatologis. As conclusion of this study, visual outcome showed that the DBSCAN delineated targeted lesion more effectively than FCM algorithm. In our study, we have also preferred to use the data generated by DBSCAN algoritm of Kockara et al. [25]. The DBSCAN clustering algorithm, introduced in 1996 [26], is generally used for discovering clusters in large spatial

databases with noise. In Kockara et al. [27], the authors add a new parametric measure to the existing algorithm to enrich the CAD images with color information and to generate more accurate results than existing methods. In Kockara et al.'s another study [28], a novel approach -graph spanner- is examined on the same set of 100 dermoscopic images for automatic border detection. Error rates, false positives and false negatives are quantified by digitally comparing results with manually determined borders from a dermatologist.

In Kockara et al. [15], a recent approach for lesion border detection in dermoscopy images, fast density based lesion detection, has obtained one of the most accurate results. In this study, for the first time, they use a modified version of prominent density based clustering algorithm, DBSCAN with the pre-processing step.

As mentioned earlier, in our study, we use fractal methods for classification. There are various research areas where classification is applied. Classification, differently from clustering, determines which class a new object belongs among a set of predefined classes. As we will explain later in Chapter 6, our algorithms try to find the category (melanoma or benign lesion) that a dermoscopy image belongs to.

Clustering and classification techniques in the literature can be used in various different research areas ranging from classification of living organisms [29] to classification of raw materials [30]. As Hunter [31] expresses, classification can and should be used in sorting anything from documents to students, from books in the library to food in the fridge. It is important to everyday life as we use it in everything we do. McLachlan et al. [32] told that there are many types of clustering methods; hierarchical agglomerative clustering, k-means clustering, the self-organizing map and model-based methods. Differently from McLachlan [32], Rokach and Maimon [33] present clustering techniques as hierarchical, partitioning, density-based, model-base, grid-based and soft-computing methods. So, it is easy to say that clustering techniques have an important place particularly in computer literature. For example, one of the most important machine learning algorithms set for classification of data mining is *Support Vector Machines (SVM)* [34]. A support vector machine is a computer

algorithm which learns by example to assign labels to objects. For example, an SVM can recognize handwritten digits by examining a large collection of scanned images of handwritten numbers. Additionally, like our study, SVM can be applied for biological applications. Theoretically, an SVM can examine the gene expression profile for a diagnosis that comes from a tumor sample. SVM can also classify objects as diverse as protein and DNA sequences [34]. As Shigeo [35] explains, in training a classifier, we aim to maximize classification performance for the training sample. For a two-class problem, a support vector machine is trained so that the direct decision function maximizes the generalization ability.

In Operations Research, cluster and classification analysis is an interesting topic, and there are various means of conducting cluster analysis. One of them is mathematical programming technique that we utilize in our study. Subjects can be grouped into different clusters by mathematical programming as performed by Rao [36]. In his analysis is confined to distance based cluster analysis in which a distance measure between the various entities is available. Since these distances are known, he tries grouping subjects by minimizing either the within groups sums of squares, sum of average within group squared distances, the total within group distances or the maximum within group distance.

Hansen and Jaumard [37] indicate steps of a clustering study, types of clustering methods and criteria. In their work, algorithms for hierarchical, partitioning, sequential and additive clustering are also studied. Differently from other clustering studies, they focus on solution methods of operations research such as dynamic programming, graph theoretical algorithms, branch-and-bound, cutting planes, column generation and heuristics. Kusiak [38], on the other hand, approaches to clustering problems from a different perspective. In his study, five different integer programming formulations of the clustering problems are studied. The relevance of integer programming and combinatorial theory to cluster analysis is discussed. The contribution of this article is that despite there are so many articles on clustering analysis and their applications, the integer programming aspect had not been emphasized before. Since Kusiak [38], practical implications of integer programming and combinatorial theory in cluster analysis has been increasing rapidly from day to day.

CHAPTER 3

FRACTAL METHODS

3.1 Fractal

Fractal comes from the Latin adjective *fractus*, which has the same root as fraction and fragment and means “irregular or fragmented”. Fractals are geometric shapes that have a structure which is self-similar on every scale in a sense that the characteristics on the scale x_1 can be deduced from those on the scale x_2 by a dilation [20]. Fractals provide visual examples of collective critical behavior. It has a self-similar hierarchical organization relying on very different scales. Thus, we can say that the length and the area of a fractal shape depend on the observed resolution. You can see some self-similar objects in Figure 3.1. Let’s first focus on the shapes of the right-most column. Since they are complicated shapes, it is not straight forward to visualize and delineate these imagine. If we did not know these shapes came from a simple pattern, it would be hard to replicate them. As an example, let’s now try to reconstruct the bottom image on the second column. We can start by drawing a triangle with equal sides on a paper. Then, we make a dot in the center of each line of the triangle. When we connect these dots with straight line segments, we obtain a new triangle with equal sides. So, we have four triangles in total now. When we continue to repeat this procedure a certain number of times, we can reach the bottommost image of the last column.

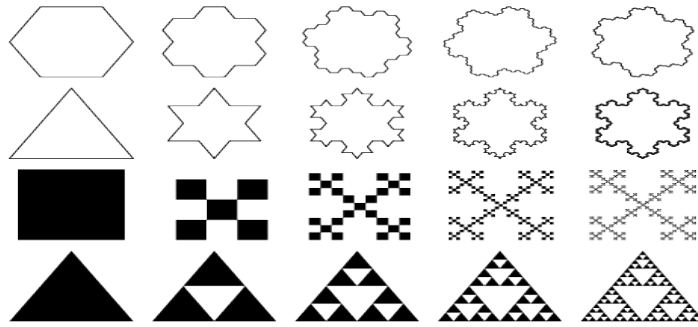


Figure 3.1: Basic fractal shapes

Fractal dimension is a parameter that measures self-similarity in the number of disjoint regions which the dataset can be divided into. Therefore, it can be deduced that that fractal dimension is a quantity which measures irregularity. It produces statistically sound merged information about the spatial distribution of skin lesion, volumetric content, and bulk density. In our thesis, eleven different “fractal dimension” calculation methods have been considered.

3.2 Fractal Methods

As already stated in Section 3.1, fractal dimension (abbreviated as FD or D) is used to describe the ruggedness, complexity, roughness or irregularity of a particle based on the set to which the particle belongs. Different approximation methods are used for the calculation of fractal dimensions. These methods show small methodological differences among each other in terms of the way how fractal dimension values are computed. The methods that are used in this study for FD calculation are *Dilation*, *Euclidean Distance Map*, *Box Counting*, *Fast*, *Fast (Hybrid)*, *Parallel Lines*, *Mass Radius (Long)*, *Mass Radius (Short)*, *Corner (Count)*, *Corner (Perimeter)* and *Cumulative Intersection*. For FD implementations, a toolbox called BIP which is created by Florida International University Biofilm Image Processing research group [39] was used as a reference point.

3.2.1 Dilation

Dilation is an image processing method. It involves convolution of null spaces which have different sizes (diameters) and then convolves them with the image border [21]. After that, each occupied pixel is confined with a square or a circle of size n , the surface of which is accepted as completely occupied. The size n is then dilated and total covered surface area $A(n)$ is calculated for each value of n . Then, the total area of the images is increased by $n \times n$. Final step is to estimate the image's border length. It is done by dividing the total area of the image by the effective diameter of the dilation element for each increment [22]. (See Figure 3.2)

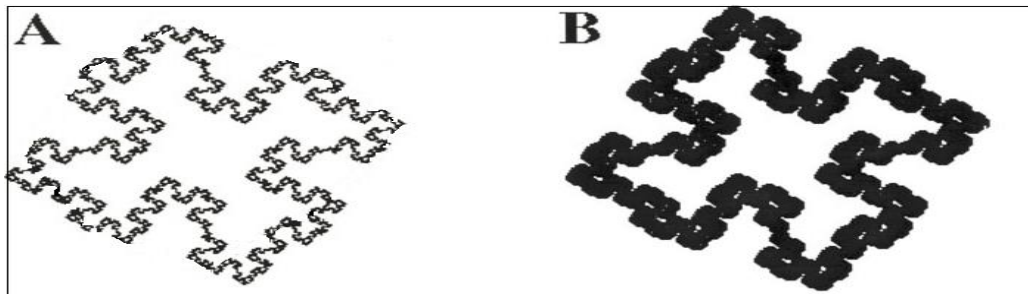


Figure 3.2: Dilation method application. A: Original image, B: Dilation method applied

3.2.2 Euclidean Distance Map

Euclidean distance map method works with monochrome images as black and white. Then gray-scale images occur as a result. The brightness value is found for each pixel of the result according to its distance to the nearest pixel of the outline of the image. The image is divided into several ribbons with a width variable w . Then FD is estimated with the proportion of the logarithm of the area for the each ribbon by the number of selected gray levels. Some FD values can be seen in Figure 3.3 [21].

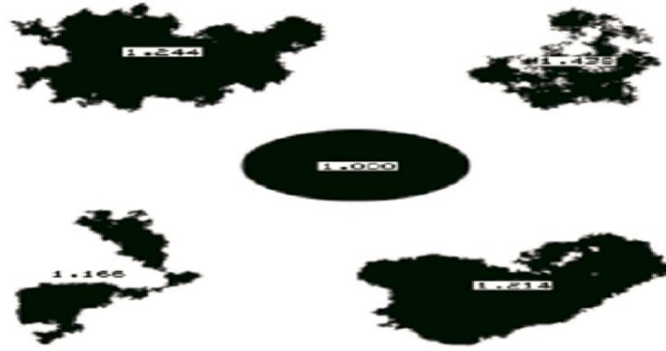


Figure 3.3: Euclidian Distance Map fractal values of different shapes

3.2.3 Box Counting

Box counting fractal method is also known as Minkowski–Bouligand [16]. In this method, we calculate FD by simply counting the number of boxes of a grid lying on the image. The number of the boxes $N(\epsilon)$ to cover the border of the image changes when the size of the unit box of the grid ϵ decreases, because they are proportional to each other. (See Figure 3.4).

The logarithmic proportion at the limit is given by the following equation.

$$FD = \lim_{\epsilon \rightarrow 0} - \frac{\ln(N(\epsilon))}{\ln(\epsilon)}$$

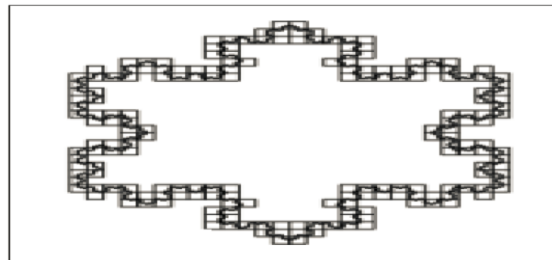


Figure 3.4: Koch Snowflake Box Counting

3.2.4 Fast

The principal target in this method is to approximate the image to a polygon. It is started by choosing a starting point on the image. Then n different points are selected and each side of the polygon is determined with two points respectively. After these, we obtain n sides. All these n points and n sides are kept. The sum of the sides of the polygon approximately gives the perimeter of the original image [16].

3.2.5 Fast (Hybrid)

This method is also known as Richardson's method. In this method, the main goal is to approximate the segmented skin lesion into a polygon by using the same length lines. This approximation entails selecting a starting point on the shape and n different points proceeding with the determination of each side of the polygon with two points respectively. The perimeter of the original image can be approximated as the sum of the edges of the polygon. Shorter the edges, it is closer to the sum of the edges to the actual perimeter of the shape. If the sum of the edges of the polygon is p (an approximation for the perimeter) and scaling factor to the length of edges is s then FD is calculated as $\log(p)/\log(1/s)$. An exemplary illustration is given in Figure 3.5. The figure shows the difference between shorter edges and longer edges. As seen, the perimeter of the first image is far away from the actual perimeter compared to the second image.

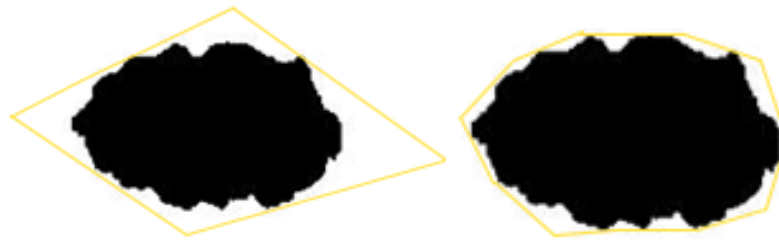


Figure 3.5: Fast FD method

3.2.6 Parallel Lines

Parallel lines algorithm separates the segmented skin lesion with equally-distanced vertical and horizontal parallel lines. After that, these lines are numbered accordingly. The number of points that intersect the lesion are counted along the line. For instance, vertical line 1 in Figure 3.6 intersects the lesion 8 times along horizontal lines. The total numbers of intersecting points P are calculated for different interval distances d between the lines. Then FD value is calculated as $\log(P)/\log(d)$.

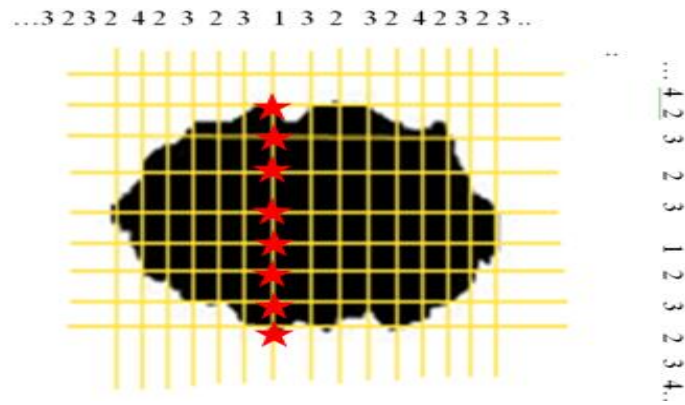


Figure 3.6. Parallel Lines FD method

3.2.7 Corner (Count)

In corner count method, corners of object are counted for different scales of connectivity values. For instance, if connectivity value is set to 4, border pixels of the object assumed to be connected as 4 pixels together. The object's border is covered with, for instance, 4 connected pixels. Then, each of these connections' corners are counted. This counting is repeated recursively with different scales of connectivity such as 2, 4, 8 etc. Then the slope is obtained from the logarithmic plot of the number of corners used to cover border against the scales of the connectivity.

3.2.8 Corner (Perimeter)

It is similar to corner count method. Instead of counting number of corners against the scales of connectivity, corner peripheral (perimeter) method sums the length of each connectivity of different scales. Then the slope is obtained from the logarithmic plot of the total length of perimeter (i.e. periphery of an object with different scale connectivity is different) against the scales of the connectivity such as 2, 4, 8 etc.

3.2.9 Cumulative Intersection

In this method the Hausdorff circles are placed at random places on the image. An iterative process is executed for each circle center. In each iteration, the radius of the circle is increased by constant k , where k is the smallest length unit.

Random but uniquely centered circles are created until whole region is encapsulated by the highest level circles. The number of separate branches (see Figure 3.7 for branches) is counted for each radius and FD is defined as $\log(N_r)/\log(r)$, where r is the circle index number from the center and N_r is the number of cumulative intersections at each exact center circle with changing sizes.

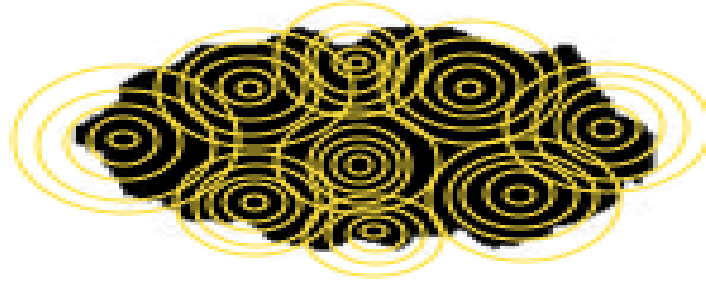


Figure 3.7 Cumulative Intersection FD method

3.2.10 Mass Radius (Long)

This method is very similar to cumulative intersection method. For this method, first circles are drawn which cover the image in changing sizes. FD is determined with the relationship between the areas of the original image found within these circles. (See Figure 3.8) [24].

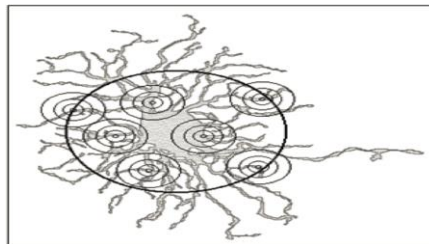


Figure 3.8. Changing centers of Mass Radius (Long) method

3.2.11 Mass Radius (Short)

The mass radius method-short method is also similar to the cumulative intersection method. It is also called as sandbox method. Such as in mass radius (long) method, circles are drawn which cover the image in changing sizes, and FD is determined with

the ratio between the areas of the original image and these circles. The difference from the mass radius (long) is different parameters [16]. Additionally, as differently from cumulative intersection, center of circles are not chosen randomly. First, center of mass of the segmented region is found. (The largest circle in Figure 3.9). Then it is used to find centers of the other circles. For each radius r , average cluster masses $M(r)$ is evaluated. FD value is calculated as the ratio of average mass changes with respect to radius r using the following equation.

$$FD = \frac{\log(M(r))}{\log(1/r)}$$



Figure 3.9. Mass Radius (Short) FD method

3.3 Relevance and Impact to Diagnosis & Treatment

As mentioned in 1.1, when the results for the diagnosis of melanocytic lesions are inaccurate, it might give over-treatment and under-treatment of patients. Misdiagnosis of melanoma is one of the most important issues for malpractice suits. There are different reasons for these misdiagnosis, but the most important reason is the subtleties of the dermatologic changes that distinguish benign from malignant proliferations of melanocytes. These distinctions are slightly different from each other, so it is quite difficult to detect by the naked eye. Because of this reason, a computer-assisted dermoscopy image guided intervention has great potential for objectivity by applying clearly defined subtle dermatologic observations [14].

With this intervention; lesion shaping, compactness, uniformity, dispersion and innumerable other dermatologic criteria can be determined in a more objective manner than the naked eye can do. Such a system has the potential to change the way dermatology and dermatopathology is practiced in the future.

CHAPTER 4

LESION BORDER IRREGULARITY CLASSIFICATION PROBLEM (LBICP)

As we have discussed in Section 1.2, our problem is to classify the patients as (i) patients who have malignant lesions and (ii) patients who have benign lesions. We used irregularities on the lesion borders to distinguish patients into these two groups. Each lesion border image was generated by a CAD based method, and the irregularities of these images were calculated with eleven different fractal methods (See Chapter 3). The parametric values obtained by these FD methods measure the deviations from a smooth shape. We can alternatively define them as the amount of roughness on the shapes. With this definition, it is seen that high-level irregularities in lesion borders are recognized as a strong evidence for higher risk of melanoma [11]. Hence, it can be claimed that when these parametric values increase, roughness in lesion border and consequently any possibility of melanoma also increase.

We have named our problem as “Lesion Border Irregularity Classification Problem” (LBICP). In a nutshell, our primary target in LBICP was to be able to differentiate patients having melanoma as accurately as possible by using the parametric values of FD methods that can quantitatively measure the irregularities on the lesion borders.

4.1 Methodology

All mathematical models in our study that we created for LBICP are mixed integer programming (MIP) models and try to differentiate patients as ones who have cancer and ones who do not have cancer. We developed four different MIP models for this purpose. Each model has one of the following four objectives: (i) maximizing the minimum between-group distance, (ii) minimizing the total within-group distances, (iii) maximizing the total between-group distances, and (iv) minimizing the maximum within-group distance.

Even though each MIP model's objective is different, we followed the same methodology as depicted in Figure 4.1 while implementing either of these models. We can explain this methodology in 5 steps. In Step-1, we choose half of patients randomly as training sample, and the rest inherently constitutes the testing sample. Training sample is a set of data records that is used to learn and discover data patterns. On the other side, testing sample is a set of data records that tests how well the learned data patterns generalize to a data population with a wide range of data records. Next, in Step-2, the MIP model we would like to implement is run for the training sample to determine the usage rate of fractal methods. The usage rates (i.e. weights) that are found as a result of Step-2 show the importance of fractal methods. To find these weights, we use the information of "training sample" patients of whether being melanoma or not as parameter since this information is already known priori for the training sample. Later, in Step-3, we run the MIP model for the testing sample with weighted FD methods (found in step-2) to determine whether the patient has melanoma or not. Hence, the output of Step-2 becomes our input for Step-3. We can summarize the procedure so far as follows: In Step-2, the information of whether a patient has melanoma or not is a parameter, and the weights of fractal methods are decided. However, in Step-3, these weights decided in Step-2 becomes parameter, and we try to decide if the patient has melanoma or not. Next, in Step-4, we compare our results found in Step-3 with the real (true) diagnosis of "testing sample" patients. Finally, in Step-5 we evaluate ratios of false diagnosis (false-negative, false-positive and total-false). The methodology is always the same for each one of the four MIP classification models. These classification models are run for several times with randomly selected training and test samples to monitor the reliability and accuracy of the models.

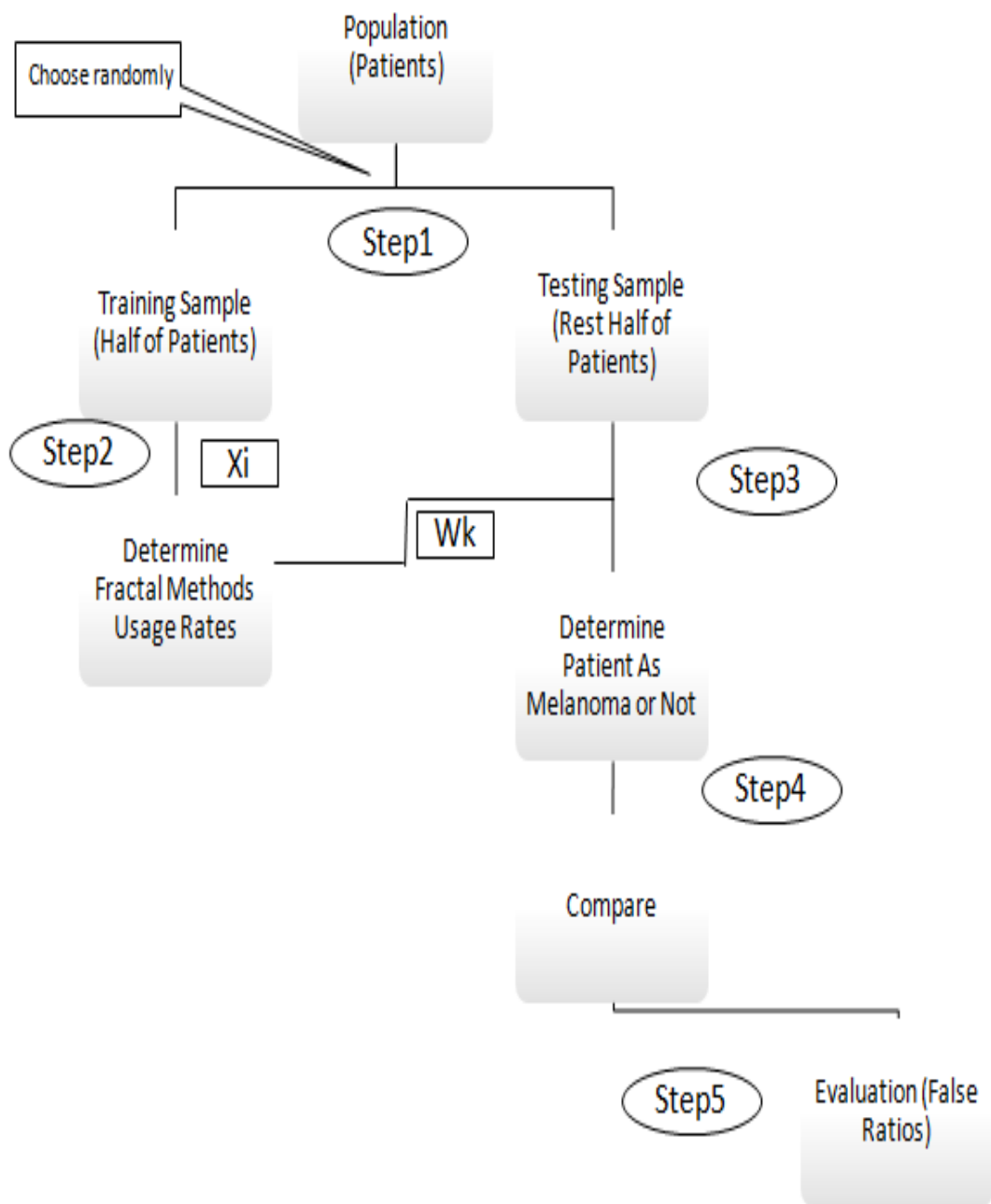


Figure 4.1. The methodology which is used in models

4.2 MIP Classification Models

As discussed in previous section, each of four MIP classification models uses one of the following objectives: (i) maximizing the minimum between-group distance, (ii) minimizing the total within-group distances, (iii) maximizing the total between-group distances, and (iv) minimizing the maximum within-group distance. All models use the same solution procedure described in Figure 4.1.

In Step 2 of the solution procedure, for each fractal method $k \in K$, we find a decision variable w_k which represents the weights of the FD method to be used whose value is between 0 and 1. Additionally, for each patient $i \in I$, B_i is defined as decision variable representing the weighted average of the parameters of the fractal methods. We also define for each patient, $j \in J$, three decision variables $d_{i,j}$, $d_{i,j}^{minus}$, $d_{i,j}^{plus}$. $d_{i,j}$ represents distance between weighted averages (B_i and B_j), but we use $d_{i,j}^{minus}$, $d_{i,j}^{plus}$ to convert non-linear classification models to linear MIP models.

All four models use $X_{i,g}$ for each patient $i \in I$ and each group $g \in G$ to indicate whether patient i is in group 1 or in group 2. This means, for example if group 1 represents the patients who have melanoma and patient i is diagnosed as having melanoma, $X_{i,1}$ and $X_{i,2}$ variables for patient i will have the values of 1 and 0, respectively. In step 2, it is determined already by dermatologists and known.

Finally, to show the values of metric, the models use $A_{i,k}$ for each fractal method $k \in K$ and for each patient $i \in I$. As we mentioned before, these values measure the deviation from a smooth shape. When these values increase, roughness and the possibility to be melanoma increase also.

Despite we use similar sets, decision variables and parameters for both training sample and test sample, there are some important changes also. For example in step 3, w_k is parameter now. Additionally, for each patient $i \in I$, B_i that is defined as the weighted average of the parameters of the fractal methods is parameter also. Because of these values are known now, $d_{i,j}$ which is the distances between weighted averages for each

patient $i \in I$ and patient $j \in J$ can be calculated easily and is given to the model. To show the values of metric, the model uses again $A_{i,k}$ as parameter for each fractal method $k \in K$ and for each patient $i \in I$. Now, $X_{i,g}$ for each patient $i \in I$ and each group $g \in G$ to indicate whether patient i is in group 1 or in group 2 is binary decision variable. One of these groups represents the patients who have melanoma, and other one represents the patients who do not have melanoma.

4.2.1 Maximizing the Minimum Between-group Distance

The purpose of this model is to maximize the minimum distance between any two patients that are in different groups. This model searches maximizing the minimum distance among all patient pairs. We can explain this also as maximizing the distance between weighted average of group-1 (patients who does not have cancer) and maximum weighted average of group-2 (patients who have cancer).

4.2.1.1 Determining Usage Rate of Fractal Methods Using Training Sample

All notations are the same as explained in Section 4.2. We also add a new binary decision variable $y_{i,j}$ that equals 1 if patient i and patient j are in different groups and $y_{i,j} = 0$; otherwise. Note that, however, w_k values are now decision variables, $X_{i,g}$ values are parameters in this step.

The mixed integer programming (MIP) formulation for the training sample can be written as follows:

$$\text{Max } z \tag{1.a.1}$$

s.t.

$$\sum_{k=1}^{11} w_k A_{i,k} = B_i \quad \forall i \in I \tag{1.a.2}$$

$$\sum_{k=1}^{11} w_k = 1 \tag{1.a.3}$$

$$B_i - B_j = d_{plus_{i,j}} - d_{minus_{i,j}} \quad \forall i \in I, j \in J; i < j \tag{1.a.4}$$

$$d_{i,j} = d_{plus_{i,j}} + d_{minus_{i,j}} \quad \forall i \in I, j \in J; i < j \tag{1.a.5}$$

$$d_{plus_{i,j}} \leq M y_{i,j} \quad \forall i \in I, j \in J; i < j \tag{1.a.6}$$

$$d_{minus_{i,j}} \leq M(1 - y_{i,j}) \quad \forall i \in I, j \in J; i < j \tag{1.a.7}$$

$$z \leq d_{i,j} \quad \exists \forall i \in I, j \in J: X_{i,g} + X_{j,g} = 1; i < j \tag{1.a.8}$$

$$w_k \geq 0 \quad \forall k \in K \tag{1.a.9}$$

$$dminus_{i,j} \geq 0 \quad \forall i \in I, j \in J \quad (1.a.10)$$

$$dplus_{i,j} \geq 0 \quad \forall i \in I, j \in J \quad (1.a.11)$$

$$d_{i,j} \geq 0 \quad \forall i \in I, j \in J \quad (1.a.12)$$

$$z \geq 0 \quad (1.a.13)$$

$$y_{i,j} \quad \text{binary variable} \quad \forall i \in I, j \in J \quad (1.a.14)$$

As we described before, objective function (1.a.1) maximizes the distance between minimum weighted average of group-1 and maximum weighted average of group-2 with constraint (1.a.8). Constraint (1.a.8) evaluates the minimum distance between B-values if patient i and j are in different groups, so it is the minimum distance between groups. It also means that this formulation maximizes the minimum distance between groups. Constraint (1.a.2) calculates B-values which are the weighted average of the parameters of the fractal method for each patient. Constraint (1.a.3) ensures that the sum of the weights of each method must be equal to 1. This constraint and constraint (1.a.9) determine the interval of the weights of each method as from 0 to 1. Constraint (1.a.4) and constraint (1.a.5) calculate the distance between B-values for each patients and ensure that this difference must be non-negative with constraint (1.a.10), constraint (1.a.11), constraint (1.a.12) and constraint (1.a.13). We define $y_{i,j}$ as binary decision variables in constraint (1.a.14) that equals 1 if patient i and j are in different groups and equals 0 if they are in same group. This definition comes from the constraint (1.a.8). Constraint (1.a.6) and constraint (1.a.7) calculate distances between patient i and patient j according to their groups. If they are in the different groups, constraint (1.a.6) frees $dplus_{i,j}$ and constraint (1.a.7) forces $dminus_{i,j}$ to be 0 with constraint (1.a.10). So, constraint (1.a.4) evaluates the distance between B-values for the patients which are in the different groups. When they are in the same group, the opposite is true, but because of the constraint (1.a.12), $dminus_{i,j}$ cannot be negative, so it is forced to be 0 again, so the distance between B-values for the patients which are in the same group is forced to be equal to 0. Thus, this constraint set only evaluates the distance between B-values for the patients which are in different groups.

4.2.1.2 Determining Melanoma Using Testing Sample

In addition to the notations described in Section 4.2, we add two new binary variables $y1_{i,j,g}$ and $y2_{i,j,g}$. Note that, however, $X_{i,g}$ values are now decision variables, w_k values are parameters in this step. In the first step, we decided the weights w_k , so we evaluate B_i (the weighted average of the parameters of the fractional (FD) methods for patient i) values with these w_k values on equation (1.b.1). With the B_i values also, we calculate d_{ij} (distances between weighted averages) values for each patient i and patient j on equation (1.b.2.), and we used d_{ij} values as parameter on the following mathematical model.

$$B_i = \sum_{k \in K} w_k A_{i,k} \quad \forall i \in I \quad (1.b.1)$$

$$d_{ij} = |B_i - B_j| \quad \forall i \in I, j \in J \quad (1.b.2)$$

We can show the mathematical model for testing sample as follows:

$$\text{Max } z \quad (1.b.3)$$

s.t.

$$z \leq d_{i,j} + M(1 - y_{i,j,g}) + M(1 - y_{2,i,j,g}) \quad \forall i \in I, j \in J, g \in G \quad (1.b.4)$$

$$\sum_{g \in G} X_{i,g} = 1 \quad \forall i \in I \quad (1.b.5)$$

$$\sum_{i \in I} X_{i,g} \leq 49 \quad \forall g \in G \quad (1.b.6)$$

$$\sum_{i \in I} X_{i,g} \geq 1 \quad \forall g \in G \quad (1.b.7)$$

$$2 - X_{i,g} + X_{j,g} \leq My1_{i,j,g} \quad \forall i \in I, j \in J, g \in G \quad (1.b.8)$$

$$X_{i,g} + X_{j,g} \leq My2_{i,j,g} \quad \forall i \in I, j \in J, g \in G \quad (1.b.9)$$

$$X_{i,g}, y1_{i,j,g}, y2_{i,j,g} \text{ binary variable} \quad \forall i \in I, j \in J, g \in G \quad (1.b.10)$$

Objective function (1.b.3) still tries to maximize the distance between minimum weighted average of group-1 and maximum weighted average of group-2 with constraint (1.b.4), (1.b.8) and (1.b.9). Constraint (1.b.5) ensures that each patient can only appear in one group. Constraint (1.b.6) and constraint (1.b.7) ensure that there will be at least one patient which is melanoma. We define $y1_{i,j,g}$ and $y2_{i,j,g}$ as binary decision variables in constraint (1.b.10) that equals 1 if patient i and j are in different groups and equals 0 if they are in same group. This definition comes from the constraint (1.b.4), (1.b.8) and (1.b.9). If patients are in different groups, constraint

(1.b.8) and constraint (1.b.9) push $y1_{i,j,g}$ and $y2_{i,j,g}$ to be 1. So, when they are equal to 1, constraint (1.b.4) calculates z as minimum distance between weighted averages of patient i and patient j . Because the problem is a maximization problem, it maximizes the minimum distance between weighted averages. When the patients are in the same group, $y1_{i,j,g}$ and $y2_{i,j,g}$ can become 0 or 1, so they free z -value.

4.2.2 Minimizing the Total Within-group Distances

Now, we try to minimize the total within-group distances. On this model, we evaluate all distances among patients who are in same group, and we try to minimize this total distance for both of groups.

4.2.2.1 Determining Usage Rate of Fractal Methods Using Training Sample

We do not add any decision variable or parameter in this model and all the notations are the same as explained in Section 4.2. Note that, however, w_k values are now decision variables, $X_{i,g}$ values are parameters in this step.

The mathematical model is given by:

$$\text{Min} \quad z = \sum_{i \in I} \sum_{j \in J} z_{ij} \quad (2.a.1)$$

s.t.

$$\sum_{k=1}^{11} w_k A_{i,k} = B_i \quad \forall i \in I \quad (2.a.2)$$

$$\sum_{k=1}^{11} w_k = 1 \quad (2.a.3)$$

$$B_i - B_j = d_{plus_{i,j}} - d_{minus_{i,j}} \quad \forall i \in I, j \in J; i < j \quad (2.a.4)$$

$$d_{i,j} = d_{plus_{i,j}} + d_{minus_{i,j}} \quad \forall i \in I, j \in J; i < j \quad (2.a.5)$$

$$d_{i,j} X_{i,g} + d_{i,j} X_{j,g} - z_{i,j} \leq d_{i,j} \quad \forall i \in I, j \in J, g \in G; i < j \quad (2.a.6)$$

$$w_k \geq 0 \quad \forall k \in K \quad (2.a.7)$$

$$d_{minus_{i,j}} \geq 0 \quad \forall i \in I, j \in J \quad (2.a.8)$$

$$d_{plus_{i,j}} \geq 0 \quad \forall i \in I, j \in J \quad (2.a.9)$$

$$d_{i,j} \geq 0 \quad \forall i \in I, j \in J \quad (2.a.10)$$

$$z \geq 0 \quad (2.a.11)$$

Objective function (2.a.1) minimizes the total distance within group distances. Constraint (2.a.2) calculates B-values which are the weighted average of the parameters of the fractal (FD) method for each patient. Constraint (2.a.3) ensures that the sum of the weights of each method must be equal to 1. This constraint and constraint (2.a.7) determine the interval of the weights of each method as from 0 to 1. Constraint (2.a.4) and constraint (2.a.5) calculate the distance between B-values for each patient and ensure that this difference must be non-negative with constraint (2.a.8), constraint (2.a.9) and constraint (2.a.10). Constraint (2.a.6) calculates distances between patient i and patient j according to their groups. If they are in different groups, constraint (2.a.6) frees $z_{i,j}$, but because of the reason that the problem is a minimization problem, the model pushes $z_{i,j}$ to be 0 with constraint (2.a.11). But on the other side, when patient i and patient j are in the same group, $z_{i,j}$ is equal to $d_{i,j}$.

4.2.2.2 Determining Melanoma Using Testing Sample

Our purpose here is again to minimize the total within-group distances. All the notations are the same as explained in Section 4.2. Note that, however, $X_{i,g}$ values are now decision variables, w_k values are parameters in this step. In the first step, we decided the weights w_k , so we evaluate B_i (the weighted average of the parameters of the fractional (FD) methods for patient i) values with these w_k values on equation (2.b.1). With the B_i values also, we calculate d_{ij} (distances between weighted averages) values for each patient i and patient j on equation (2.b.2.), and we used d_{ij} values as parameter on the following mathematical model.

$$B_i = \sum_{k \in K} w_k A_{i,k} \quad \forall i \in I \quad (2.b.1)$$

$$d_{ij} = |B_i - B_j| \quad \forall i \in I, j \in J \quad (2.b.2)$$

The new mathematical formulation is as follows:

$$\text{Min } z = \sum_{i \in I} \sum_{j \in J} z_{ij} \quad (2.b.3)$$

s.t.

$$d_{i,j} X_{i,g} + d_{i,j} X_{j,g} - z_{i,j} \leq d_{i,j} \quad \forall i \in I, j \in J, g \in G ; i < j \quad (2.b.4)$$

$$\sum_{g \in G} X_{i,g} = 1 \quad \forall i \in I \quad (2.b.5)$$

$$z \geq 0 \quad (2.b.6)$$

$$X_{i,g} \text{ binary variable} \quad \forall i \in I, j \in J, g \in G \quad (2.b.7)$$

Objective function (2.b.3) still tries to minimize the total distance within group distances with constraint (2.b.4). Constraint (2.b.5) ensures that each patient can only appear in one group. Constraint (2.b.4) frees $z_{i,j}$ if the patients are in different groups, but because of the reason that the problem is a minimization problem, the model pushes $z_{i,j}$ to be 0 with constraint (2.b.6). But on the other side, when patient i and patient j are in the same group $z_{i,j}$ is equal to $d_{i,j}$. Constraint (2.b.7) defines $x_{i,g}$ as 1 if patient i in group g has melanoma.

4.2.3 Maximizing the Total Between-group Distances

Now, we try to maximize the total distance between group distances. Like in Section 4.2.1., we evaluate the distances among all patient pairs, and we search for distances between groups. Unlike searching for minimum distance, this time we try to find the total distance between two groups. As maximizing this total distance, we separate the patients into two groups as much as possible. Higher value means higher separation for patients.

4.2.3.1 Determining Usage Rate of Fractal Methods Using Training Sample

We only add a constant binary decision variable $y_{i,j}$ that equals 1 if patient i and patient j are in different groups and $y_{i,j} = 0$; otherwise. All other notations are the same as explained in Section 4.2. Note that, however, w_k values are now decision variables, $X_{i,g}$ values are parameters in this step.

The mathematical model is as follows:

$$\text{Max } z = \sum_{i \in I} \sum_{j \in J} z_{ij} \quad i < j \quad (3.a.1)$$

s.t.

$$\sum_{k=1}^{11} w_k A_{i,k} = B_i \quad \forall i \in I \quad (3.a.2)$$

$$\sum_{k=1}^{11} w_k = 1 \quad (3.a.3)$$

$$B_i - B_j = d_{plus_{i,j}} - d_{minus_{i,j}} \quad \forall i \in I, j \in J; i < j \quad (3.a.4)$$

$$d_{i,j} = d_{plus_{i,j}} + d_{minus_{i,j}} \quad \forall i \in I, j \in J; i < j \quad (3.a.5)$$

$$z_{i,j} \leq d_{i,j}(X_{i,g} + X_{j,g}) \quad \forall i \in I, j \in J, g \in G; i < j \quad (3.a.6)$$

$$z_{i,j} \leq d_{i,j}(2 - X_{i,g} - X_{j,g}) \quad \forall i \in I, j \in J, g \in G; i < j \quad (3.a.7)$$

$$dplus_{i,j} \leq My_{i,j} \quad \forall i \in I, j \in J; i < j \quad (3.a.8)$$

$$dminus_{i,j} \leq M(1 - y_{i,j}) \quad \forall i \in I, j \in J; i < j \quad (3.a.9)$$

$$w_k \geq 0 \quad \forall k \in K \quad (3.a.10)$$

$$dminus_{i,j} \geq 0 \quad \forall i \in I, j \in J \quad (3.a.11)$$

$$dplus_{i,j} \geq 0 \quad \forall i \in I, j \in J \quad (3.a.12)$$

$$d_{i,j} \geq 0 \quad \forall i \in I, j \in J \quad (3.a.13)$$

$$z \geq 0 \quad (3.a.14)$$

$$y_{i,j} \text{ binary variable} \quad \forall i \in I, j \in J \quad (3.a.15)$$

Objective function (3.a.1) maximizes the total distance between group distances with constraint (3.a.6) and constraint (3.a.7). Constraint (3.a.2) calculates B-values which are the weighted average of the parameters of the fractal (FD) method for each patient. Constraint (3.a.3) ensures that the sum of the weights of each method must be equal to 1. This constraint and constraint (3.a.10) determine the interval of the weights of each method as from 0 to 1. Constraint (3.a.4) and constraint (3.a.5) calculate the distance between B-values for each patient and ensure that this difference must be non-negative with constraint (3.a.11), constraint (3.a.12) and constraint (3.a.13). We define $y_{i,j}$ as binary decision variables in constraint (3.a.15) that equals 1 if patient i and j are in different groups and equals 0 if they are in same group. This definition comes from the constraint (3.a.8) and (3.a.9). Constraint (3.a.6) and constraint (3.a.7) calculate $z_{i,j}$ for the patients who are in different groups. When they are in same groups, these constraints force $z_{i,j}$ to be 0 together.

4.2.3.2 Determining Melanoma Using Testing Sample

Here, we do not need to add any new decision variables or parameters. All notations are the same as explained in Section 4.2. Note that, however, $X_{i,g}$ values are now decision variables, w_k values are parameters in this step. In the first step, we decided the weights w_k , so we evaluate B_i (the weighted average of the parameters of the fractional (FD) methods for patient i) values with these w_k values on equation (3.b.1). With the B_i values also, we calculate d_{ij} (distances between weighted averages) values for each patient i and patient j on equation (3.b.2.), and we used d_{ij} values as parameter on the following mathematical model.

$$B_i = \sum_{k \in K} w_k A_{i,k} \quad \forall i \in I \quad (3.b.1)$$

$$d_{ij} = |B_i - B_j| \quad \forall i \in I, j \in J \quad (3.b.2)$$

The mathematical formulation is as follows:

$$\text{Max } z = \sum_{i \in I} \sum_{j \in J} z_{ij} \quad (3.b.3)$$

s.t.

$$\sum_{g \in G} X_{i,g} = 1 \quad \forall i \in I \quad (3.b.4)$$

$$z_{i,j} \leq d_{i,j}(X_{i,g} + X_{j,g}) \quad \forall i \in I, j \in J, g \in G ; i < j \quad (3.b.5)$$

$$z_{i,j} \leq d_{i,j}(2 - X_{i,g} - X_{j,g}) \quad \forall i \in I, j \in J, g \in G ; i < j \quad (3.b.6)$$

$$X_{i,g} \text{ binary variable} \quad \forall i \in I, j \in J, g \in G \quad (3.b.7)$$

Objective function (3.b.3) still tries to maximize the total distance between group distances with constraint (3.b.5) and (3.b.6). Constraint (3.b.4) ensures that each patient can only appear in one group. Constraint (3.b.5) and constraint (3.b.6) calculate $z_{i,j}$ for the patients who are in different groups. When they are in same groups, these constraints force $z_{i,j}$ to be 0 together. Constraint (3.b.7) defines $x_{i,g}$ as 1 if patient i in group g has melanoma.

4.2.4 Minimizing the Maximum Within-group Distance

Now, we try to minimize the maximum distance within group distances. Here again, we target grouping by getting patients who have minimum distance between each other closer.

4.2.4.1 Determining Usage Rate of Fractal Methods Using Training Sample

Note that, however, w_k values are now decision variables, $X_{i,g}$ values are parameters in this step.

The mathematical model can be written as follows:

$$\text{Min} \quad z \quad (4.a.1)$$

s.t.

$$\sum_{k=1}^{11} w_k A_{i,k} = B_i \quad \forall i \in I \quad (4.a.2)$$

$$\sum_{k=1}^{11} w_k = 1 \quad (4.a.3)$$

$$B_i - B_j = d_{plus_{i,j}} - d_{minus_{i,j}} \quad \forall i \in I, j \in J; i < j \quad (4.a.4)$$

$$d_{i,j} = d_{plus_{i,j}} + d_{minus_{i,j}} \quad \forall i \in I, j \in J; i < j \quad (4.a.5)$$

$$d_{i,j} X_{i,g} + d_{i,j} X_{j,g} - z \leq d_{i,j} \quad \forall i \in I, j \in J, g \in G; i < j \quad (4.a.6)$$

$$w_k \geq 0 \quad \forall k \in K \quad (4.a.7)$$

$$d_{minus_{i,j}} \geq 0 \quad \forall i \in I, j \in J \quad (4.a.8)$$

$$d_{plus_{i,j}} \geq 0 \quad \forall i \in I, j \in J \quad (4.a.9)$$

$$d_{i,j} \geq 0 \quad \forall i \in I, j \in J \quad (4.a.10)$$

$$z \geq 0 \quad (4.a.11)$$

Objective function (4.a.1) minimizes the maximum within group distances with constraint (4.a.6). Constraint (4.a.2) calculates B-values which are the weighted average of the parameters of the fractal (FD) method for each patient. Constraint (4.a.3) ensures that the sum of the weights of each method must be equal to 1. This constraint and constraint (4.a.7) determine the interval of the weights of each method as from 0 to 1. Constraint (4.a.4) and constraint (4.a.5) calculate the distance between B-values for each patient and ensure that this difference must be non-negative with constraint (4.a.8), constraint (4.a.9) and constraint (4.a.10). Constraint (4.a.6) calculates distances between patient i and patient j according to their groups. If they are in the different groups, constraint (4.a.6) pushes z to be 0 with constraint (4.a.11). However, when they are in same groups, z is forced to be $d_{i,j}$ with objective function.

4.2.4.2 Determining Melanoma Using Testing Sample

Note that, however, $X_{i,g}$ values are now decision variables, w_k values are parameters in this step. In the first step, we decided the weights w_k , so we evaluate B_i (the weighted average of the parameters of the fractional (FD) methods for patient i) values with these w_k values on equation (4.b.1). With the B_i values also, we calculate $d_{i,j}$ (distances between weighted averages) values for each patient i and patient j on equation (4.b.2.), and we used $d_{i,j}$ values as parameter on the following model.

$$B_i = \sum_{k \in K} w_k A_{i,k} \quad \forall i \in I \quad (4.b.1)$$

$$d_{ij} = |B_i - B_j| \quad \forall i \in I, j \in J \quad (4.b.2)$$

The mathematical formulation can be written as follows:

$$\text{Min } z = \sum_{i \in I} \sum_{j \in J} z_{ij} \quad (4.b.3)$$

s.t.

$$d_{i,j}X_{i,g} + d_{i,j}X_{j,g} - z \leq d_{i,j} \quad \forall i \in I, j \in J, g \in G ; i < j \quad (4.b.4)$$

$$\sum_{g \in G} X_{i,g} = 1 \quad \forall i \in I \quad (4.b.5)$$

$$X_{i,g} \text{ binary variable} \quad \forall i \in I, j \in J, g \in G \quad (4.b.6)$$

$$z \geq 0 \quad (4.b.7)$$

Objective function (4.b.3) still tries to minimize the maximum within group distances with constraint (4.b.4). Constraint (4.b.4) calculates distances between patient i and patient j according to their groups. If they are in the different groups, constraint (4.b.6) pushes z to be 0 with constraint (4.b.7). However, when they are in same groups, z is forced to be $d_{i,j}$ with objective function. Constraint (4.b.5) ensures that each patient can only appear in one group. Constraint (4.b.6) defines $x_{i,g}$ as 1 if patient i in group g has melanoma.

CHAPTER 5

DATA ANALYSIS AND RESULTS

5.1 Dataset

As mentioned in Chapter 2, density based clustering is an effective clustering method used in data mining for discovering spatial databases. BD-DBSCAN is improved version of prominent density based clustering algorithm (DBSCAN) [15]. In our study, we consider a dermoscopy image as a dataset in which each pixel belongs to one group (cluster) according to its spatial location and color.

In our study first, CAD based delineation of lesion borders was taken with two ways. First the dermatologist took this border manually. Additionally, a CAD based method was used by taking delineation of lesion borders such as in Figure 5.1. As it is seen easily, the border with the CAD based method is much more accurate than manually drawn border and reflects the shape irregularity better.

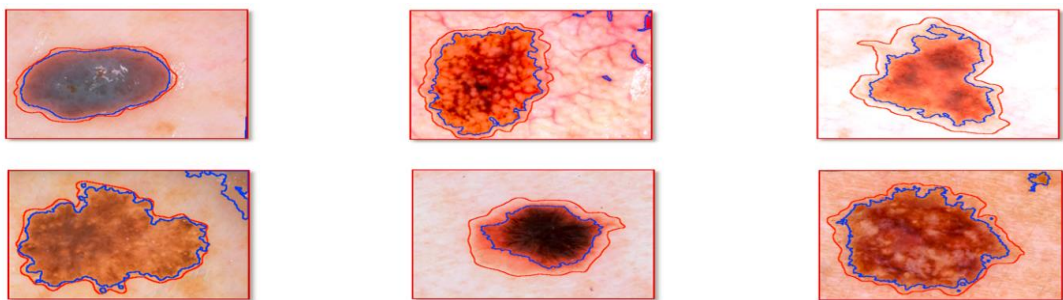


Figure 5.1: Sample skin cancer images showing borders delineated by the dermatologist (red) and a CAD based method, BD-DBSCAN (blue).

Then for measuring irregularities of every lesion border image generated by a CAD based method, eleven different fractal methods were used as we mentioned in Chapter 3. It was already known that which images belong to malignant lesions and which images belong to benign lesions because they were determined before by dermatologists, so the dataset we obtained for this project contains 100 dermoscopy images which are diagnosed by dermatologist. We used the images showing borders delineated by a CAD based method which was boundary driven density based clustering algorithm (BD-DBSCAN). 30 of them were diagnosed as malignant melanoma and 70 of them were diagnosed as benign.

5.2 Computational Experiments

The objective of this section is to evaluate the performances of the models which were defined in Chapter 4 in terms of false ratios. At the first stage, we created 4 different training samples randomly. Each training sample was constituted of 50 patients. The rest 50 patients formed testing sample automatically. We ran each model in Step-2 with training sample and in Step-3 with testing sample respectively four times, and observed the results.

We used false negative, false positive and total false values for comparison. For our study, true negative and true positive are the proper evaluations. Dermatologists diagnose your situation correct. On the other hand, false negative and false values are false diagnosis. False negative means that you have malignant lesion, but dermatologists say that you do not have. On the contrary, false positive means that dermatologists say that you have malignant lesion despite you do not have in real (Figure 5.2). Total false values are the summations of mistaken diagnosis, so we mostly based on total false values during our comparison.

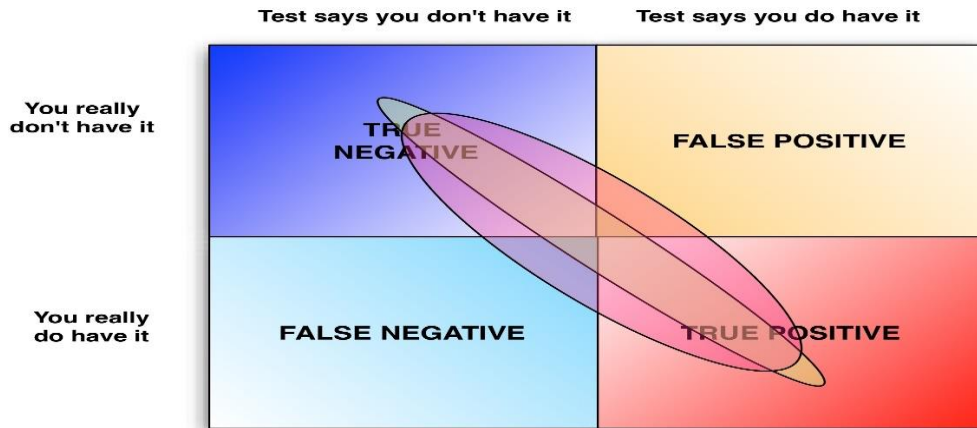


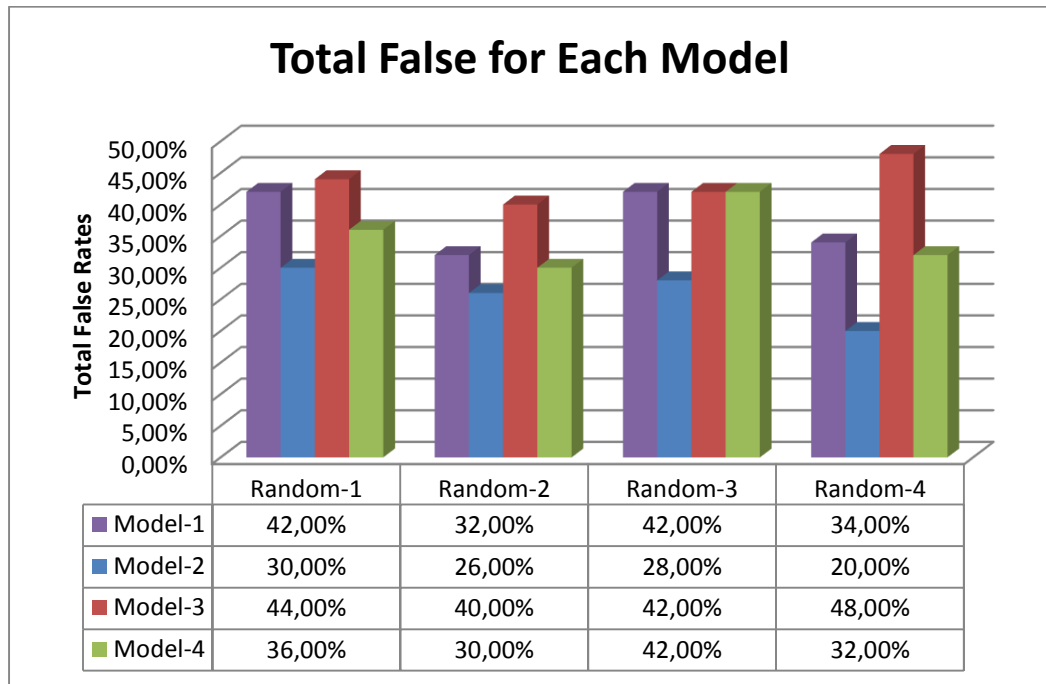
Figure 5.2 False-negative, false-positive, true-negative, true positive

All the experiments were performed on an Intel Core™ i5 (5200U) 2.20GHz machine with 6 GB RAM. They were implemented using GAMS 23.9.5 that runs with CPLEX 12.4 solver.

As it is seen from Table 5.1, the best results were observed in Model-2 on each false-negative, false-positive and total false results. Despite the best total false result is 34% in Model-1, 40% in Model-3 and 30% in Model 4, we observed 20% in Model-2. In addition to this, the worst total-false result is still better than the best results of other models. Not only total false values are important in our evaluations, but also false negative and false positive values are important. Model-1 and Model-4 are really bad in false-negative results. For example let us examine the results for Model-4, Random-1. As it is seen on the figure, false-negative value is 90%. There is written that 18/20. It means that the model found 18 patients as healthy despite they have cancer in the real life. For Random-1, there are 20 patients who have cancer in real life. But, the model could find only 2 of them as cancer patients. As we seen, false-positive value is 0%. It means that the model could find all 30 patients who do not have any cancer cells. Additionally, our model did not claim patients as sick patients despite they are healthy. So, we know that the model diagnosed 18 patients from 50 patients wrongly in total. Because of this, total false value is 18/50 which means 36%.

Table 5.1 False-negative, false-positive and total-false results for 4 models and 4 random groups

	Random-1			Random-2			Random-3			Random-4		
Model-1	False Negative	20\20	100.00%	14\14	100.00%	20\21	95.24%	16\16	100.00%	False Negative	16\16	100.00%
	False Positive	1\30	3.33%	2\36	5.56%	1\29	3.45%	1\34	2.94%	False Positive	1\34	2.94%
	Total False	21\50	42.00%	16\50	32.00%	21\50	42.00%	17\50	34.00%	Total False	17\50	34.00%
Model-2	False Negative	12\20	60.00%	6\14	42.86%	10\21	47.62%	6\16	37.50%	False Negative	6\16	37.50%
	False Positive	3\30	10.00%	7\36	19.44%	4\29	13.79%	4\34	11.76%	False Positive	4\34	11.76%
	Total False	15\50	30.00%	13\50	26.00%	14\50	28.00%	10\50	20.00%	Total False	10\50	20.00%
Model-3	False Negative	12\20	60.00%	9\14	64.29%	14\21	66.67%	12\16	75.00%	False Negative	12\16	75.00%
	False Positive	10\30	33.33%	11\36	30.56%	7\29	24.14%	12\34	35.29%	False Positive	12\34	35.29%
	Total False	22\50	44.00%	20\50	40.00%	21\50	42.00%	24\50	48.00%	Total False	24\50	48.00%
Model-4	False Negative	18\20	90.00%	14\14	100.00%	20\21	95.24%	15\16	93.75%	False Negative	15\16	93.75%
	False Positive	0\30	0.00%	1\36	2.78%	1\29	3.45%	1\34	2.94%	False Positive	1\34	2.94%
	Total False	18\50	36.00%	15\50	30.00%	21\50	42.00%	16\50	32.00%	Total False	16\50	32.00%



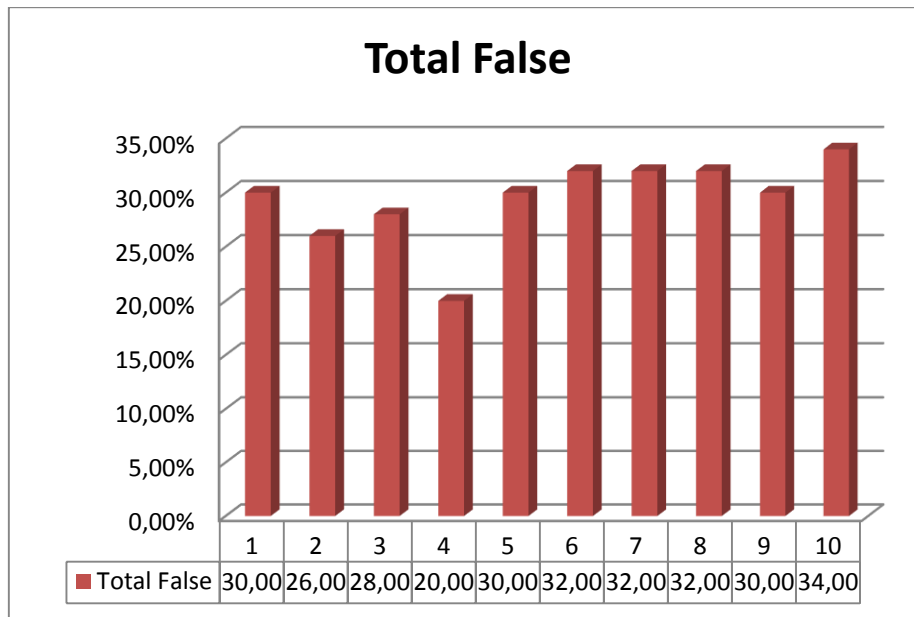
Graph 5.1. Total false values for first 4 random samples

The graph also shows the best results were taken from Model-2 and the worst results were taken from Model-3. After observing this, we decided to continue our analysis with Model-2 and created 6 more random training samples. You can see the ratios for these 10 samples with the first 4 samples in Table 5.2 and Graph 5.2.

As seen from Table 5.2, our best result did not change despite we added 6 more samples. The total false result interval is from 20% to 34%.

Table 5.2 Model 2 results for 10 random groups

Model-2		Random-1		Random-2		Random-3		Random-4		Random-5		Random-6		Random-7		Random-8		Random-9		Random-10	
	False Negative	12\20	60.00%	6\14	42.86%	10\21	47.62%	6\16	37.50%	2\11	18.18%	7\16	43.75%	5\13	38.46%	8\18	44.44%	9\19	47.37%	8\17	47.06%
	False Positive	3\30	10.00%	7\36	19.44%	4\29	13.79%	4\34	11.76%	13\39	33.33%	11\37	29.73%	8\32	25.00%	6\31	19.35%	6\31	19.35%	9\33	27.27%
	Total False	15\50	30.00%	13\50	26.00%	14\50	28.00%	10\50	20.00%	15\50	30.00%	16\50	32.00%	16\50	32.00%	16\50	32.00%	15\50	30.00%	17\50	34.00%



Graph 5.2 Total false values for 10 random samples for model-2

After we examined the results, we wanted to check which fractal methods are used in which ratio. Let us remember fractal methods that were used in our study. These methods are; Dilation, Euclidean Distance Map, Box Counting, Fast, Fast (Hybrid), Parallel Lines, Mass Radius (Long), Mass Radius (Short), Corner (Count), Corner (Perimeter) and Cumulative Intersection. We examined usage of these methods on our 10 random samples. As it is seen in Table 5.3, some methods are always used in all random samples and some of them are not. For example, Fractal Methods-3,4,6,7,8,10 are never used in these random samples. On the other hand, Method-2 is always used in highest fraction. So, we concluded that Euclidean Distance Map method is a good method on determining the patient who has melanoma or not. You can see and observe other methods from Table 5.3.

Table 5.3 Used fractal methods in first 10 random groups

Random-1			
0,7382611	0,2614543	0,00028	
2	5	9	Corner (Count)
Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	

Random-2			
0,26529	0,71741	0,000151	0,01715362
1	2	9	11
Dilation	Euclidean Distance Map	Corner (Count)	Cumulative Intersection

Random-3			
0,20627613	0,728073	0,0653966	0,00025
1	2	5	9
Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)

Random-4			
0,1523	0,739001	0,1010092	0,00021
1	2	5	9
Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)
			Cumulative Intersection

Random-5			
0,1707272	0,7968333	0,03196	0,00047692
1	2	5	9
Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)

Random-6			
0,234774	0,746578	0,00998178	3,9311E-05
1	2	5	9
Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)
			Cumulative Intersection

Random-7			
0,7388022	0,26073	0,00047	
2	5	9	Corner (Count)
Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	

Random-8			
0,238275	0,7481501	0,01331	0,00026329
1	2	5	9
Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)

Random-9			
0,2358222	0,7376876	0,01758	0,0089085
1	2	5	11
Dilation	Euclidean Distance Map	Fast (Hybrid)	Cumulative Intersection

Random-10			
0,259521	0,732434	0,00020042	0,00784387
1	2	10	11
Dilation	Euclidean Distance Map	Corner (Perimeter)	Cumulative Intersection

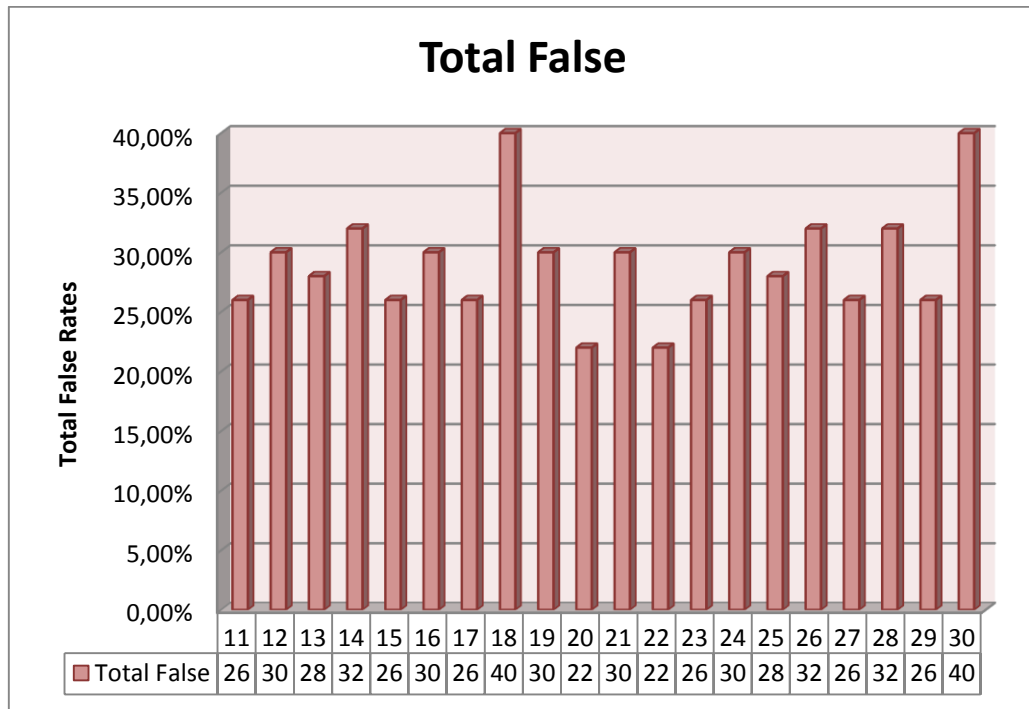
After observing these results, we decided to try a new approach and created our random scenarios according to patients who have melanoma. As we mentioned before in Section 4.1, we were choosing our training and testing sample randomly. Now, we changed this selection method. We chose again 50 random patients, but this time we paid attention by choosing these 50 patients as 15 random patients from 30 patients who have melanoma and 35 random patients from 70 patients who do not have melanoma. Again this selection was randomly but this time we changed the selection clusters. With this scenario, we created 20 more random training samples as 15-35 and ran Model-2. (You can see the results in Table 5.4) In addition to these 20 random samples, we created 20 different more random training samples with the first usual way by choosing 50 patients from 100 patients randomly. (See Table 5.6)

We ran Model-2 for these 40 samples and examined the results. One of our targets was to find a better solution than 20%, other one was not to find really bad results. Interval was important for our study. During our study, we have never encountered with a result higher than 40% total false.

After that, again we observed the usage rates of fractal methods for these 40 random samples (See Table 5.5 & Table 5.7). As before, fractal methods 3,4,6,8 were never used again.

Table 5.4 20 random scenarios for 15-35 grouping

Model-2			
Random-11			
False Negative	5\15	33.33%	
False Positive	8\35	22.86%	
Total False	13\50	26.00%	
Random-12			
False Negative	8\15	53.33%	
False Positive	7\35	20.00%	
Total False	15\50	30.00%	
Random-13			
False Negative	4\15	26.67%	
False Positive	10\35	28.57%	
Total False	14\50	28.00%	
Random-14			
False Negative	6\15	40.00%	
False Positive	10\35	28.57%	
Total False	16\50	32.00%	
Random-15			
False Negative	6\15	40.00%	
False Positive	7\35	20.00%	
Total False	13\50	26.00%	
Random-16			
False Negative	6\15	40.00%	
False Positive	9\35	25.71%	
Total False	15\50	30.00%	
Random-17			
False Negative	7\15	46.67%	
False Positive	6\35	17.14%	
Total False	13\50	26.00%	
Random-18			
False Negative	9\15	60.00%	
False Positive	11\35	31.43%	
Total False	20\50	40.00%	
Random-19			
False Negative	7\15	46.67%	
False Positive	8\35	22.86%	
Total False	15\50	30.00%	
Random-20			
False Negative	5\15	33.33%	
False Positive	6\35	17.14%	
Total False	11\50	22.00%	
Random-21			
False Negative	7\15	46.67%	
False Positive	8\35	22.86%	
Total False	15\50	30.00%	
Random-22			
False Negative	5\15	33.33%	
False Positive	6\35	17.14%	
Total False	11\50	22.00%	
Random-23			
False Negative	5\15	33.33%	
False Positive	8\35	22.86%	
Total False	13\50	26.00%	
Random-24			
False Negative	8\15	53.33%	
False Positive	7\35	20.00%	
Total False	15\50	30.00%	
Random-25			
False Negative	4\15	26.67%	
False Positive	10\35	28.57%	
Total False	14\50	28.00%	
Random-26			
False Negative	6\15	40.00%	
False Positive	10\35	28.57%	
Total False	16\50	32.00%	
Random-27			
False Negative	6\15	40.00%	
False Positive	7\35	20.00%	
Total False	13\50	26.00%	
Random-28			
False Negative	6\15	40.00%	
False Positive	10\35	28.57%	
Total False	16\50	32.00%	
Random-29			
False Negative	7\15	46.67%	
False Positive	6\35	17.14%	
Total False	13\50	26.00%	
Random-30			
False Negative	9\15	60.00%	
False Positive	11\35	31.43%	
Total False	20\50	40.00%	



Graph 5.3 Total false values for 20 more random samples

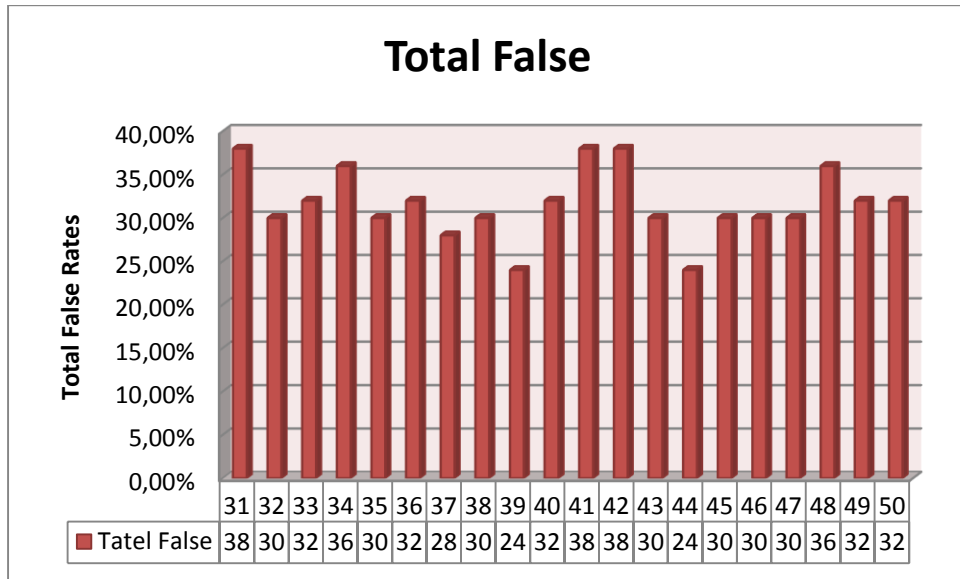
As it is seen from Graph 5.3, the best result is 22%, but still worse than Random-4 which we found in first test. Additionally, our worst result is 40% in these 20 random samples which is worse than our first 10 results.

Table 5.5 Fractal methods for additional 20 random sets

Random-11			Random-12			Random-13			Random-14							
0.247896833	0.734555682	8.42E-05	0.017463314	0.766632117	0.01157426	0.004360271	0.000433294	0.179104	0.785342523	0.0345906	0.000963	0.204278	0.770032134	0.025116296	0.0005731	
1	2	9	11	1	2	5	7	1	2	5	9	1	2	5	9	
Dilation	Euclidean Distance Map	Corner (Count)	Cumulative Intersection	Dilation	Euclidean Distance Map	Fast (Hybrid)	Mass Radius (Long)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	
Random-15			Random-16			Random-17			Random-18							
0.239589264	0.750665795	0.000912	0.008832574	0.22595168	0.755655629	0.01691683	0.000522018	0.2280381	0.737820706	0.0338295	0.000312	0.203426	0.778766117	0.000250954	0.0175565	
1	2	5	11	1	2	5	9	1	2	5	9	1	2	9	11	
Dilation	Euclidean Distance Map	Fast (Hybrid)	Cumulative Intersection	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Corner (Count)	Cumulative Intersection	
Random-19			Random-20			Random-21			Random-22							
0.213213066	0.771018796	0.015424	0.000344265	0.170837551	0.725633811	0.094206707	4.25E-05	0.2132131	0.771018796	0.0154239	0.000344	0.170838	0.725633811	0.094206707	4.25E-05	
1	2	5	9	1	2	5	9	1	2	5	9	1	2	5	9	
Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Mass Radius (Count)	Corner (Count)
Random-23			Random-24			Random-25			Random-26							
0.247896833	0.734555682	8.42E-05	0.017463314	0.216900058	0.766632117	0.01157426	0.004360271	0.179104	0.785342523	0.0345906	0.000963	0.204278	0.770032134	0.025116296	0.0005731	
1	2	9	11	1	2	5	7	1	2	5	9	1	2	5	9	
Dilation	Euclidean Distance Map	Corner (Count)	Cumulative Intersection	Dilation	Euclidean Distance Map	Fast (Hybrid)	Mass Radius (Long)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	
Random-27			Random-28			Random-29			Random-30							
0.239589264	0.750665795	0.000912	0.008832574	0.243894699	0.743369686	0.012146382	0.000589223	0.2280381	0.737820706	0.0338295	0.000312	0.203426	0.778766117	0.000250954	1.76E-02	
1	2	5	11	1	2	7	9	1	2	5	9	1	2	9	11	
Dilation	Euclidean Distance Map	Fast (Hybrid)	Cumulative Intersection	Dilation	Euclidean Distance Map	Mass Radius (Long)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Corner (Count)	Cumulative Intersection	

Table 5.6.20 random scenarios for 50 grouping

Model-2		Random-31			Random-32			Random-33			Random-34		
	False Negative	8\12	66.67%	9\19	47.37%	8\17	47.06%	6\15	40.00%				
	False Positive	11\38	28.95%	6\31	19.35%	8\33	24.24%	12\35	34.29%				
	Total False	19\50	38.00%	15\50	30.00%	16\50	32.00%	18\50	36.00%				
Random-35		Random-36			Random-37			Random-38					
	False Negative	7\15	46.67%	9\17	52.94%	6\14	42.86%	7\13	53.85%				
	False Positive	8\35	22.86%	7\33	21.21%	8\38	21.05%	8\37	21.62%				
	Total False	15\50	30.00%	16\50	32.00%	14\50	28.00%	15\50	30.00%				
Random-39		Random-40			Random-41			Random-42					
	False Negative	2\9	22.22%	10\18	55.56%	7\16	43.75%	16\17	94.12%				
	False Positive	10\41	24.39%	6\32	18.75%	12\34	35.29%	3\33	9.09%				
	Total False	12\50	24.00%	16\50	32.00%	19\50	38.00%	19\50	38.00%				
Random-43		Random-44			Random-45			Random-46					
	False Negative	9\20	45.00%	9\11	81.82%	4\12	33.33%	9\20	45.00%				
	False Positive	6\30	20.00%	3\39	7.69%	11\38	28.95%	6\30	20.00%				
	Total False	15\50	30.00%	12\50	24.00%	15\50	30.00%	15\50	30.00%				
Random-47		Random-48			Random-49			Random-50					
	False Negative	7\13	53.85%	6\9	66.67%	9\15	60.00%	8\17	47.06%				
	False Positive	8\37	21.62%	12\33	36.36%	7\35	20.00%	8\33	24.24%				
	Total False	15\50	30.00%	18\50	36.00%	16\50	32.00%	16\50	32.00%				



Graph 5.4 Total false values for last 20 random samples

The worst total-false result is 38% in these 20 random sample set. When we analyzed Table 5.7, method 2 was not used for random sample-42 and random sample-44 for the first time. Fractal method-11 was the only method that was used in these samples.

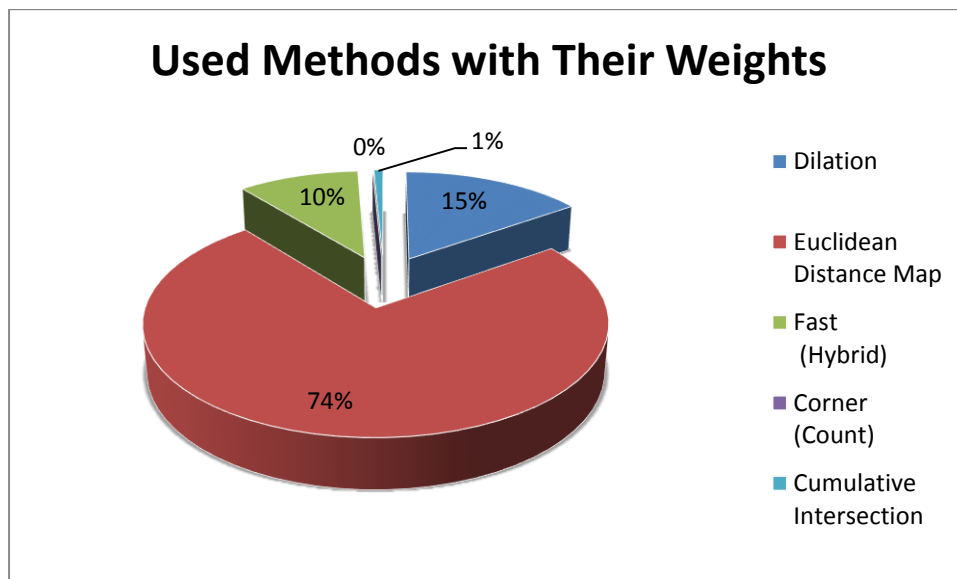
Table 5.7 Fractal methods for last 20 random groups

Random-31			Random-32			Random-33			Random-34					
0.228833943	0.755271015	0.000413	0.015481968	0.797687622	0.017581548	0.008908499	0.7619254	0.231486086	0.0002725	0.006316	0.18549134	0.784054477	0.0288874	0.000566801
1	2	9	11	1	2	5	11	2	5	9	11	2	5	9
Dilation	Euclidean Distance Map	Corner (Count)	Cumulative Intersection	Dilation	Euclidean Distance Map	Fast (Hybrid)	Cumulative Intersection	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)
Random-35			Random-36			Random-37			Random-38					
0.182811471	0.772285252	0.044612	0.000291121	0.238726578	0.748885045	0.011998105	0.000410272	0.0597448	0.751902805	0.1877307	0.000622	0.254438684	0.712361715	0.0003955
1	2	5	9	1	2	7	9	1	2	5	9	1	2	9
Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Miss Radius (Long)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Corner (Count)
Cumulative Intersection														Cumulative Intersection
Random-39			Random-40			Random-41			Random-42					
0.206501516	0.77581839	0.017273	0.00030725	0.21739289	0.744452268	0.000813199	0.037341622	0.2089311	0.748394041	0.0422769	0.000366	0.2089311	0.748394041	0.0422769
1	2	5	9	1	2	9	11	1	2	5	9	1	2	10
Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Corner (Count)	Cumulative Intersection	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Cumulative Intersection
Random-43			Random-44			Random-45			Random-46					
0.7344795	0.264909521	0.000611	0.000511	0.2063458	0.761479347	0.0318337	0.000341	0.2063458	0.761479347	0.0318337	0.000341	0.727765461	0.272079525	0.000155
2	5	9	11	1	2	5	9	1	2	5	9	2	5	9
Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Cumulative Intersection	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)
Random-47			Random-48			Random-49			Random-50					
0.269752352	0.707246986	0.00019	0.022811026	0.194686822	0.767979562	0.001916732	0.01769538	0.2357427	0.730458667	0.0279648	0.000337	0.20061238	0.720297437	0.0710377
1	2	9	11	1	2	5	10	1	2	5	9	1	2	5
Dilation	Euclidean Distance Map	Corner (Count)	Cumulative Intersection	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Perimeter)	Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Dilation	Euclidean Distance Map	Fast (Hybrid)

After examining these 50 results, we decided to choose the best fractal methods. The best fractal method means here, fractal methods which were used in random sample that gave the best total-false results. As we have mentioned above, the best total-false results were observed in Random-4. You can see the fractal methods which are used in this sample with their weights in Table 5.8 and Graph 5.5. These were method-1, method-2, method-5, method-9 and method-11.

Table 5.8 Fractal methods with their fractions for best random sample

0.152337612	0.73900079	0.101009	0.000207	0.007445602
1	2	5	9	11
Dilation	Euclidean Distance Map	Fast (Hybrid)	Corner (Count)	Cumulative Intersection

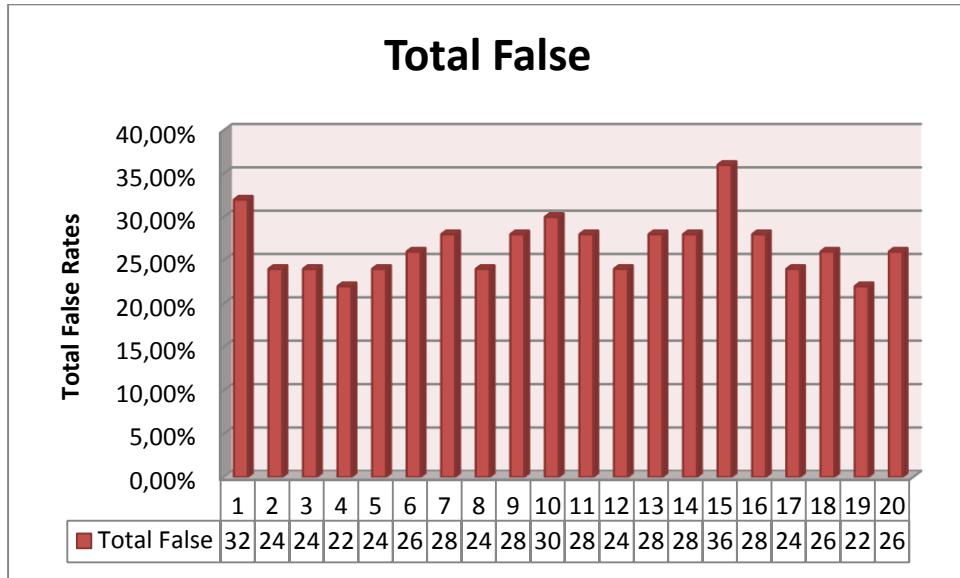


Graph 5.5 Used fractal methods with their weights for the best random sample result

After observing this, we created 20 more random samples by usual way. This time, we did not run the model for Step-2 because we already chosed our fractal methods and their weights. We only ran the model for Step-3 with the weights given above.

Table 5.9 20 last random scenarios to determine melanoma

Model-2			
Random-1			
False Negative	10\21	47.62%	
False Positive	6\29	20.69%	
Total False	16\50	32.00%	
Random-2			
False Negative	4\13	30.77%	
False Positive	8\37	21.62%	
Total False	12\50	24.00%	
Random-3			
False Negative	5\18	27.78%	
False Positive	7\32	21.88%	
Total False	12\50	24.00%	
Random-4			
False Negative	4\11	36.36%	
False Positive	7\39	17.95%	
Total False	11\50	22.00%	
Random-5			
False Negative	6\13	46.15%	
False Positive	6\37	16.22%	
Total False	12\50	24.00%	
Random-6			
False Negative	6\15	40.00%	
False Positive	7\35	20.00%	
Total False	13\50	26.00%	
Random-7			
False Negative	4\13	30.77%	
False Positive	10\37	27.03%	
Total False	14\50	28.00%	
Random-8			
False Negative	2\11	18.18%	
False Positive	10\39	25.64%	
Total False	12\50	24.00%	
Random-9			
False Negative	5\16	31.25%	
False Positive	9\34	26.47%	
Total False	14\50	28.00%	
Random-10			
False Negative	7\17	41.18%	
False Positive	8\33	24.24%	
Total False	15\50	30.00%	
Random-11			
False Negative	5\12	41.67%	
False Positive	9\38	23.68%	
Total False	14\50	28.00%	
Random-12			
False Negative	6\14	42.86%	
False Positive	6\36	16.67%	
Total False	12\50	24.00%	
Random-13			
False Negative	2\11	18.18%	
False Positive	12\39	30.77%	
Total False	14\50	28.00%	
Random-14			
False Negative	4\10	40.00%	
False Positive	10\40	25.00%	
Total False	14\50	28.00%	
Random-15			
False Negative	8\19	42.11%	
False Positive	10\31	32.26%	
Total False	18\50	36.00%	
Random-16			
False Negative	8\18	44.44%	
False Positive	6\32	18.75%	
Total False	14\50	28.00%	
Random-17			
False Negative	3\10	30.00%	
False Positive	9\40	22.50%	
Total False	12\50	24.00%	
Random-18			
False Negative	6\17	35.29%	
False Positive	7\33	21.21%	
Total False	13\50	26.00%	
Random-19			
False Negative	6\21	28.57%	
False Positive	5\29	17.24%	
Total False	11\50	22.00%	
Random-20			
False Negative	4\15	26.67%	
False Positive	9\35	25.71%	
Total False	13\50	26.00%	



Graph 5.6 Total false values for last 20 Random Samples

After 70 runs, we saw that our best result for total-false value is still 20%. It means 80% success in diagnosis of melanoma skin cancer on dermoscopy images. As it is observed in Table 5.10, the average value for total false values of these last 20 random samples is 26.6%, the minimum total false value is 22%, and the maximum total false values is 36%.

Table 5.10 Average, min., max. values for last 20 Random Samples

Average	26,6%
Minimum	22%
Maximum	36%

CHAPTER 6

CONCLUSION

6.1 Summary of the Research

Dermatology is a medical specialty which deals with the diagnosis and treatment of diseases of the skin [25]. As American Cancer Society declared, the most significant step in treatment of a disease is early and effective diagnosis. Early diagnosis saves lives. From cancer types, the most prevalent one is skin cancer [3]. For determination of skin cancer, dermatologists use a tool, dermatoscope. It takes high resolution images and analyzes dark colored lesion or diagnoses melanoma. This process is known as dermoscopy. Melanoma which begins in melanocytes constitutes the largest part of skin cancers that results in death [1][4]. To distinguish malignant from benign lesions is not an easy process. As normal procedure, dermatologists take dermoscopy images, examine these images visually and then draw the lesion borders manually to follow the process since cancer cells grow irregularly so that cancer regions' borders have shape irregularities. Lesion border irregularity is one of the criteria which serves to distinguish malignant from benign lesions. Physicians currently draw lesion borders manually with the naked eye, and this process is subjective and hard to recognize compartments and tissue structures. Unfortunately, rate of false diagnosis is high when lesion borders are manually drawn by physicians. Despite physicians use dermoscopy, 70% of melanoma claims are still false-negative diagnosis due to the lack of knowledge about the diagnosis of melanocytic lesions [7]. Misdiagnosis of these lesions results in one of the causes of medical malpractice for this group of physicians

because it results in overtreatment or under treatment of patients. Even though using dermoscopy images decreases misdiagnosis of lesions, still one of the most important objectives of dermatologists is to gain ability to diagnose pigmented skin lesions with high accuracy as much as possible. This is the main motivation to utilize computer assisted diagnosis (CAD) techniques for the diagnosis of the melanoma. The techniques that are developed by fractal methods can specify irregularities on the lesion borders since fractal properties of skin lesions reflect irregularity of shapes. Fractal dimension (FD) values can be calculated by several different approximation methods that differ from each other according how they measure the irregularity on a shape. The fractal methods used in our research are Dilation, Euclidean Distance Map, Box Counting, Fast, Fast (Hybrid), Parallel Lines, Mass Radius (Long), Mass Radius (Short), Corner (Count), Corner (Perimeter) and Cumulative Intersection. Our study aims to determine which fractal methods are more effective on distinguishing malignant lesions from benign lesions to decrease ratios of false-positive, false-negative and total false diagnosis. To reach our objective, we use the same data that are also used in three articles of Kockara et al. [25, 27, 28]. On the first part of our study, CAD based method for automated lesion border delineation is employed, then 11 fractal methods are implemented to measure lesion's irregularity. After measuring irregularities, we use the FD values found by these 11 fractal methods as input in our study. All mathematical models that we create for Lesion Border Irregularity Classification Problem (LBICP) in our study are mixed integer programming (MIP) models. We developed four different MIP models which have four distinct objectives: (i) maximizing the minimum between-group distance, (ii) minimizing the total within-group distances, (iii) maximizing the total between-group distances, and (iv) minimizing the maximum within-group distance. Despite all models have different objective functions, all of them tries to classify patients as ones who have cancer and ones who do not have cancer. First, we find optimum usage rate of fractal methods for each models on training sample. Next, true diagnosis performances of each models on testing sample are evaluated by using usage rates which were found on the training sample. After several experiments have been conducted, we find that the rate of false diagnosis is 20% at the best case, which means 80% of the time diagnosis of melanoma skin cancer on dermoscopy images is successful. The average rate of false diagnosis, on the other hand, is found to be 26.6%.

6.2 Opportunities for Future Work

As we mentioned in Chapter 4, the parametric values obtained by these FD methods measure the deviations from a smooth shape, so they are the amount of roughness on the shapes. High level irregularities in lesion borders mean high risk of melanoma. The values that measure the deviation from a smooth shape (i.e. irregularity) are the most important parameters of our study. Therefore, in the future, the number of these parametric values can be increased by using a sample larger than 100 patients. Thus, all the experiments can be repeated for more randomly generated sets to achieve more successful true diagnosis rates.

REFERENCES

- [1] *Dermatology*. (2015). Encyclopaedia Britannica.
<http://academic.eb.com/EBchecked/topic/158586/dermatology>
- [2] Jayaraman, V.S., Narasimhan, K. & Sankaranarayanan, V. (2014). Melanoma Detection Using Dermoscopy Images. *International Journal of Applied Engineering Research*, 9, 13.
- [3] *The Dermoscopy Site* (2012). Retrieved February, 2015, from
<http://www.dermoscopy.org/atlas/base.htm>
- [4] Thatte, S.S. & Khopkar, U.S. (2014). The Utility of Dermoscopy in the Diagnosis of Evolving Lesions of Vitiligo. *Indian Journal of Dermatology, Venereology Leprology*, 80, 506-508.
- [5] Naldi, L., Jordan, L. & Durisilla H. (2010). Etiological Factors in Skin Cancers: Environmental and Biological. *Cancer of the Skin*, (2nd ed.) London: Elsevier.
- [6] AmericanCancerSociety. (2010). *Cancer Facts&Figures*. Available. Retrieved March, 2015, from
<http://www.cancer.org/acs/groups/content/@nho/documents/document/acspc-024113.pdf>
- [7] Carucci, J.A. & Darrell S.R. (2000). Malignant Melanoma: Prevention, Early Detection, and Treatment in the 21st Century. *CA Cancer J Clin*, 5, 21.
- [8] AmericanCancerSociety. (2010). *Melanoma Skin Cancer Overview*. Available. Retrieved March, 2015, from
<http://www.cancer.org/acs/groups/cid/documents/webcontent/003063-pdf.pdf>
- [9] NationalCenterforChronicDiseasePreventionandHealthPromotion. (2011). *The Burden of Skin Cancer*. Available. Retrieved April, 2015, from
<http://www.cdc.gov/HealthyYouth/skincancer/pdf/facts.pdf>
- [10] Troxel, D.B. (2003). Pitfalls in the Diagnosis of Malignant Melanoma: Findings of a Risk Management Panel Study. *The American Journal of Surgical Pathology*, 27, 1278-1283.
- [11] Mulrane, R.E., Penney, S., Callanan, J.J. & Gallagher W.M. (2008). Automated Image Analysis in Dermatology: a Valuable Tool in Medical Diagnostics. *Expert Review of Molecular Diagnostics*, 8, 707-725.

- [12] Conway, C., Dobson, L., O'Grady, A., Kay, E., Costello, S. & O'Shea, D. (2008). Virtual Microscopy as an Enabler of Automated/quantitative Assessment of Protein Expression in TMAs. *Histochemistry and Cell Biology*, 130, 447-463.
- [13] Binder M., Schwarz M., Winkler A., Steiner A., Kaider A., Wolff K. & Pehamberger H. (1995). Epiluminescence microscopy. A Useful Tool for the Diagnosis of Pigmented Skin Lesions for Formally Trained Dermatologists. *Archives of Dermatology*, 131(3), 286-291.
- [14] Celebi, M. E., Iyatomi, H., Schaefer G. & Stoecker W.V. (2009). Lesion Border Detection in Dermoscopy Images. *Computerized Medical Imaging and Graphics*, 33, 148-153.
- [15] Kockara, S., Mete, M. & Aydin, K. (2010). Fast Density-Based Lesion Detection in Dermoscopy Images. *Computerized Medical Imaging and Graphics*, 35(2), 128-136.
- [16] Ercan, M. (2012). *Fractal Methods as a Prognostic Factor to Determine Malignancy in Dermoscopy*. (Unpublished Master Thesis Report). University of Central Arkansas, Conway, Arkansas.
- [17] Willett, J.A. (2010). *The American Beauty Industry Encyclopedia*. Santa Barbara, Calif: Greenwood.
- [18] Rigel, D. S., Friedman, R. J. & Kopf, A.W (1996). Lifetime Risk for Development of Skin Cancer in the U.S. Population: Current Estimate is Now 1 in 5. *Journal of the American Academy Dermatology*, 35, 1012-1013.
- [19] Rigel, D. S., Friedman, R. J. & Kopf, A.W (1996). The Incidence of Malignant Melanoma in the United States: Issues as we Approach the 21st Century. *Journal of the American Academy Dermatology*, 34(5), 839-847.
- [20] Lesne, A. (1988). *Renormalization Methods Critical Phenomena, Chaos, Fractal Structures*. West Sussex, UD, England: John Wiley & Sons Ltd.
- [21] Berube, D., Jebrak, M. (1999). High Precision Boundary Fractal Analysis for Shape Characterization. *Computers & Geosciences*, 25, 1059-1071.
- [22] Allen, M., Brown, G.J. & Miles, N.J. (1995). Measurement of Boundary Fractal: Review of Current Techniques. *Powder Technology*, 84, 1-14.
- [23] Cornforth, D., Jelinek H. & Peichl, L. (2002). *Fractop: a Tool for Automated Biological Image Classification*, 6th Australasia-Japan Joint Workshop, ANU, Canberra.
- [24] Jelinek, F. H. & Fernandez, E. (1998). Neurons and Fractals: How Reliable and Useful are Calculations of Fractal . *Journal of Neuroscience Methods*, 81, 9-18.

- [25] Kockara, S., Mete, M., Chen, B. & Aydin, K. (2011). Analysis of Density-based and Fuzzy C-means Clustering Methods on Lesion Border Extraction in Dermoscopy Images. *BMC Bioinformatics*, 11(6), 26.
- [26] Ester, M., Kriegel, H.P., Sander, J. & Xu, X. (1996). A Density-based Algorithm for Discovering Clusters in Large Spatial Databases With Noise. *Conference on Knowledge Discovery and Data Mining*. 226-231.
doi: 10.1.1.71.1980
- [27] Kockara, S., Suer, S. & Mete, M. (2011). An Improved Border Detection in Dermoscopy Images for Density Based Clustering. *BMC Bioinformatics*, 12.8.
- [28] Kockara, S., Mete, M., Yip, V., Lee, B. & Aydin, K. (2010). A Soft Kinetic Data Structure for Lesion Border Detection. *Oxford Bioinformatics*, 26, 21-28.
- [29] Kadhila, N. *Characteristics and Classification of Living Organisms*. Cambridge: Cambridge University Press – NSSC Biology Module 1.
- [30] Gu, J., Liu, C. (2012). Discriminative Illumination: Per-Pixel Classification of Raw Materials based on Optimal Projections of Spectral BRDF. *Computer Vision and Pattern Recognition*, 797-804.
- [31] Hunter, E. (2000) Do We Still Need Classification? *The Future of Classification*, Gower, 1-17.
- [32] McLachlan, G.J., Bean, R.W. & Nh, S.K. (2008). Clustering. *Methods in Molecular Biology*, 453, 423.
- [33] Rokach, L. & Maimon, O. (2005). Clustering Methods. *Data Mining and Knowledge Discovery Handbook*, 321-352.
- [34] Noble, W.S. (2006). What is a Support Vector Machine? *Nature Biotechnology*, 24(12), 1565-1567.
- [35] Shiego, A. (2010). Support Vector Machines for Pattern Classification. (2nd ed.), *Two Class Support Vector Machines* (pp. 21-112). London: Springer.
- [36] Rao, M. R. (1971). Cluster Analysis and Mathematical Programming. *Journal of the American Statistical Association*, 66(335), 622-626.
- [37] Hansen, P. & Jaumard, B. (1997). Cluster Analysis and Mathematical Programming. *Mathematical Programming*, 79, 191-215.
- [38] Kusiak, A. (1984). Analysis of Integer Programming Formulations of Clustering Problems. *Image and Vision Computing*, 2(1), 35-40.
- [39] Florida International University Biofilm Image Processing Research Group (2000). Available. Retrieved January, 2015, from <http://users.cs.fiu.edu/~giri/BIP/>

Curriculum Vitae

Zühal Özcan was born on 12 July 1990, in Aydın. She received her BS degree in Industrial Engineering from Işık University in 2014 with 1st ranking out of all graduates. Additionally, she took her double major from International Trade in 2014. She completed her M.S. degree in 2015 in Operations Research from Işık University. She worked as a teaching assistant at the department of Industrial Engineering of Işık University from 2014 to 2015 during her Master Degree. She earned Fulbright Scholarship in 2015 for doctorate program in United States of America. Additionally, her admission to Arizona State University to study in the Industrial Engineering (PhD) program was accepted, and she has been awarded Teaching Assistantship and CIDSE Doctoral Fellowship. Her research interests include operations research, scheduling and mathematical modelling.