2018

SKIN CANCER DETECTION USING SVM-BASED CLASSIFICATION AND PSO FOR SEGMENTATION

osamah A. Almasiri

SKIN CANCER DETECTION USING SVM-BASED CLASSIFICATION AND PSO FOR SEGMENTATION

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

BY

OSAMAH AYAD ALMASIRI
Bachelor of Biomedical Engineering, Baghdad University, Iraq 2008

Director: DING-YU FEI, PH.D
ASSOCIATE PROFESSOR, BIOMEDICAL ENGINEERING

Virginia Commonwealth University
Richmond, Virginia
May 2018

Acknowledgment

I would first like to express my gratitude towards my family for everything that they have done for me, but especially for how they have believed in me and encouraged me to complete this thesis. My dear dad has been faithfully by my side, and without his sacrifices, I would not be the man I am today. To my beloved mom, I believe God has heard all of her prayers on my behalf because of your sincere heart; thanks to her from the very bottom of my heart.

Also, I would like to express my special gratitude and thanks to uncle Read. I have turned to him many times when I have faced difficult situations; he has consistently been beside me and I could always count on him.

Thanks to my lovely Richmond family; there are no words to describe my appreciation and gratefulness to them. I consider myself so fortunate to have three moms, one at home and two abroad. I am deeply indebted to my 2nd Mama, Sandra, for her support, encouragement, love and a host of other things she has done for me; this journey would have been more difficult without her.

I would also like to convey my sincere thanks to Louis and Linda (3rd Mama); they have been by my side. Louis's wisdom has guided me and he has helped me in countless ways. I have felt Linda's continuing prayers on my behalf; I know they have come from her big and beautiful heart.

I would also like to thank my thesis advisor, Dr. Ding-Yu Fei, Virginia Commonwealth University School of Engineering, Department of Biomedical Engineering, Richmond Virginia. His office door was always open whenever I had an unresolved issue. He consistently allowed this paper to be my own work, but with his wise counsel.

My thanks and appreciations also go to Dr. Azhar Rafiq, another committee member, who has willingly advised me with his able guidance and all his efforts in helping me complete this thesis.

Dr. Paul Wetzel, the 3rd member of my advisory committee, has guided me with his insightful comments and questions. I say “thank you” to him

Finally, I must express my very profound gratitude to the Higher Committee of Education Development in Iraq (HCED). Thanks to the HCED for giving me this opportunity to study in the USA and receive my Master’s Degree in Biomedical Engineering. This degree would not have been possible without their confidence in me.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

List of Abbreviations

SI	Swarm Intelligence
PSO	Particle Swarm Optimization
SVM	Support Vector Machine
GLCM	Gray Level Co-occurrence Matrices
GLRLM	Gray Level Run Length Matrix
ROI	Region of Interest
SD	Standard deviation
Var	Variance of column data.
MI	Mutual Information
HD	Hammoude Distance
TDR	True Detection Rate
FPR	False Position Rate
GE	Gross Error
SE	Sensitivity
SP	Specificity

Abstract

SKIN CANCER DETECTION USING SVM-BASED

CALSSIFICATION AND (PSO) FOR THE SEGMENTATION

By OSAMAH A. ALMASIRI, B.E

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2018

Director: DING-YU FEI, PH.D
ASSOCIATE PROFESSOR, BIOMEDICAL ENGINEERING

Various techniques are developed for detecting skin cancer. However, the type of maligned skin cancer is still an open problem. The objective of this study is to diagnose melanoma through design and implementation of a computerized image analysis system. The dataset which is used with the proposed system is Hospital Pedro Hispano (PH^2).

The proposed system begins with preprocessing of images of skin cancer. Then, particle swarm optimization (PSO) is used for detecting the region of interest (ROI).

After that, features extraction (geometric, color, and texture) is taken from (ROI). Lastly, features selection and classification are done using a support vector machine (SVM).

Results showed that with a data set of 200 images, the sensitivity (SE) and the specificity (SP) reached 100% with a maximum processing time of 0.03 sec.

CHAPTER 1 INTRODUCTION

1.1 Background

Skin cancer can be defined as an abnormal propagation of cells in the skin. The malicious propagation is created due to a rapid increase of skin cells [1]. Skin cancer is considered as one of the most dangerous types of cancers. Skin cancer has been identified in different forms such as melanoma and nonmelanoma skin cancer. Examples of non-melanoma skin cancer include basal cell and squamous cell carcinoma. Melanoma is a very serious cancer because it has the ability to spread to other parts of the human body [2]. Rare types of skin cancer also occur, including Merkel cell carcinoma, skin lymphoma, and Kaposi sarcoma.

The Center for Disease Control tracks melanoma type of skin cancer with cancer registries, and the data show that in 2014 76,665 people were diagnosed with melanomas in the United States [3]. More specifically, of those diagnosed, 9,234 patients died from melanoma in 2014. Malignant melanoma is considered one of the most common malignant tumors, with an incidence rate of ten times greater than other cancers. Since the early 1970s, malignant melanoma incidence has risen. For example, in the USA, it increased by about 4 percent every year[4]. The major reason for the rapid growth of melanoma situations is the redundant exposure to ultraviolet rays (UV) from the sun. Experts indicate that about 90% of melanomas are connected with severe UV exposure and sunburns over a lifetime [5]. Therefore, the immediate diagnosis at early stages is essential since delayed diagnosis of melanoma can have a negative influence in terms of personal health, extensive medical treatment procedures and treatment cost [6].

Diagnosis of melanoma can be considered a challenge even when using tools such as the standard dermatome scanner. Overall, the precision of melanoma diagnosis by experts is within 75%–84%. This emphasizes the need for innovative tools in diagnostics methods such as computer-based analysis of disease-related images [7].

In spite of the current efforts to develop image data analysis with Artificial Intelligence (AI) capabilities, no analytic software has been produced that can fully replicate human intelligence [8]. Advanced computational analysis methodologies are being used to develop computer algorithms, which can be utilized to identify clinical characteristics in the image, which might be missed by human observation. The medical literature discloses that many algorithms have been evaluated for image classification [8]. Therefore, diagnostic image processing and analysis by software systems can play an essential role in the clinical domain.

Humans can classify objects easily such as medical images from various patients, but this task is a complex analytical problem for computers. Classification of an image includes identification labels being assigned to the sections in the image such as object classification and feature extraction which are considered basic techniques in different applications. Usually, classification systems are developed with a defined database of images that includes predefined patterns in order to classify a detected object in the proper class [9].

Chapter two gives more details on melanoma disease, however this study is attempting to develop melanoma detection using a computer-aided diagnosis system with an evaluation of two image segmentation techniques and classification of the images using a Support Vector Machine (SVM) method.

1.2 Related Work

Various researchers have investigated the extraction and classification of considerable clinical image features to provide detection of melanoma, such as:

- A. Bono et. al. (2004) [10], described a clinical study of 22 cases of melanoma with the diameter of less than 3 mm. They extracted geometric and dermoscopic features from the images. A lesion was considered suspicious for melanoma when at least one diagnostic parameter was identified as being positive. In their work, they noticed that the ABCD diagnostic algorithm becomes redundant while evaluating micro-melanomas and that traditional dermoscopic diagnostic methods often fail.

- N. K. Al abbadi et. al. (2008) [11], proposed a method for skin texture recognition using both color and texture features of lesions and image classification using three layer neural networks.

- J. Blackledge and D. A. Dubovitskiy (2009) [12], focused on the development and implementation of a skin cancer image screening software system that could be used in a general practice by non-experts to 'filter' normal from abnormal cases. Thus, the positive lesion patients would be referred to a specialist. A fuzzy logic analytic engine was used to classify an image object based on both its Euclidean and fractal geometric properties.

- Capdehourat et al. (2011) [13], utilized statistical, color, and geometric features for the classification of images. The study was performed on 655 images of melanocytic lesions with 544 benign and 111 malignant melanoma regions. The result obtained indicated their computational analysis to have an 89% specificity and 95% sensitivity using the AdaBoost/C5.4 approach. The same disease classifications were applied to the images by an expert dermatologist who segmented them as normal versus abnormal manually; the result was 85% specificity and 90% sensitivity.

- Safi et al. (2012) [14], proposed a machine learning approach to classify melanocytic lesions into malignant and benign from dermoscopic images. For segmentation, they used multiphase soft segmentation method with total variation and H^1 regularization. Thus, each lesion image was characterized by a feature vector that contained shape, color and texture information. Additionally, the images were analysed with local and global parameters that try to reflect on lesion structures used in medical diagnosis. The learning and classification stage was performed using Support Vector Machine (SVM) with polynomial kernels. The classification method delivered accuracy of 98.57% with a true positive rate of 0.991% and a false positive rate of 0.019%.

- Huang et al. (2012) [15], focused their research on capillary detection in the skin cancer images. They used a SVM classifier to identify capillary structures in the image pixels. They estimated the likelihood of capillaries based on the distance to the red color in the "CIE Lab" color space. The accuracy with this method was 89.8% (44/49), the sensitivity was 90.5% (19/21) and the specificity was 89.3% (25/28).

- Korotkov et al. (2012) [16], presented a study on an analysis of pigmented skin lesions. Their method included a comparison of various analytic techniques, such as fuzzy C-means, center split, multi-resolution, split and merge, PCT/median cut and adaptive thresholding. From the results of their comparative study, it was concluded that the adaptive snake was the best among gradient vector flow followed by level set, adaptive thresholding, expectation-maximization (EM) level set and fuzzy based split-and-merge algorithm.

- Ma et al. (2013) [17], focused on a recognition of melanoma cases specifically. They used statistical measures such as entropy wavelet energy and geometric based irregularity measures of the images being analyzed. The geometric measures included radial deviation, contour roughness, and

irregularity measures. They used Discrete Wavelet Transform (DWT) to confine signal components to dynamically increase width frequency bands in different resolutions.

- Arroyo et al. (2014) [18], analyzed dermoscopy images by detecting pigment network using supervised machine learning and structural analysis. They used high-level design which is composed of two main analytic blocks, a machine learning process and searching of the structures for specific patterns. For feature extraction from images, gray level co-occurrence matrix (GLCM) was implemented on the segmented regions. Using C4.5 algorithm for 220 images (120 without a reticular pattern and 100 with such structure), the method achieves a sensitivity of 86% and the specificity of 81.67%.

- Poornima M S and Shailaja K (2017) [1], presented a method for the detection of melanoma by checking various melanoma parameters such as color, area perimeter, diameter by texture, size and shape. The extracted feature parameters were applied to images in order to classify them as non-melanoma or melanoma cancer.

- Uzmabano Ansari and Tanujasarode (2017) [19], proposed a skin cancer detection system using SVM for early detection of skin cancer disease. The features of each image were extracted using GLCM methodology. These features are given as the input to the classifier software. Support Vector Machine (SVM) is used for classification purpose. It classifies the given image into cancerous or non-cancerous.

1.3 Objective of the Study

Skin cancer diagnosis for humans remains an open problem according to the skin cancer image type complexity. Different research studies have been done in this field, but maligned skin

cancer type is still a challenge in detection and localization since the intensity of the lesion area can be similar to the skin image.

The objective of this thesis is to design and implement a robust, efficient, and computerized analytic software system to analyze skin images in order to detect and diagnose melanoma. Overall, this thesis is divided into different stages: the first stage is the image preprocessing, the second stage is the segmentation of the preprocessed image, using particle swarm optimization (PSO) as an intelligent tool for detecting the region of interest (ROI). The third stage is about analyzing features such as geometric, color, and texture which are extracted from ROI detected image. The fourth stage features normalization and selection, and the last stage involves features classification using a support vector machine (SVM).

CHAPTER 2 Background

2.1 Skin Cancer

One of the most common cancers in humans is skin cancer. There are two main types of skin cancer: non-melanoma skin cancer and malignant melanoma [20].

2.1.1 Non-Melanoma (Benign)

Benign cancers are rarely life-threatening, and they do not propagate to other parts of the body. They are often removed by surgical techniques and usually do not recur. It is one of the types of skin cancer that greater than 99,500 people in the US are diagnosed with every year. There are different types of non-melanoma skin cancers but, the most common types of benign skin cancer are basal cell carcinoma and squamous cell carcinoma with approximately 5.4 million cases per year in the US [21].

A- Basal Cell Carcinoma (BCC)

BCC is the most common type of skin cancer and arises in cells of the deepest layer of the skin. Basal cell carcinoma usually occurs on the face and neck, where the skin is most exposed to sunlight [22].

Basal cell carcinoma rarely kills. However, because it can cause large degrees of damage and disfigurement, it is still considered life-threatening by invading surrounding tissues [23].

B- Squamous Cell Carcinoma (SCC)

Squamous cell carcinoma is the second most common skin cancer. It is a more aggressive cancer than basal cell carcinoma because it has the ability to spread to other parts of the body e.g. nearby lymph nodes. It is dangerous, but not as dangerous as a melanoma [21]. Squamous cell carcinoma begins in squamous cells which are thin, flat cells in appearance. It's normally found in areas that are not exposed as much to the sun, such as the legs or feet [24].

2.1.2 Malignant Melanoma

Melanoma is a malignant tumor of melanocytes. Melanocytes are cells that produce the melanin, dark pigments that are responsible for the color of the skin. Melanoma is less common than other skin cancers and behaves differently as compared to basal cell and squamous cell cancers. However, it is much more dangerous and causes a large majority of skin cancer deaths since it can spread widely in the body [24]. It can grow fast and requires immediate treatment. Melanoma can appear on any skin surface across the body. In men, it's often found on the head, on the neck, or between the shoulders and the hips. In women, it's often found on the skin of the lower legs, between the shoulders or between the hips. Melanoma is scarce in people with dark skin. When it does get expressed in people with dark skin, it's normally found below the fingernails, on the palms of the hands, or on the soles of the feet [22].

2.1.3 Skin Cancer Diagnosis

There are different ways to evaluate skin lesions and diagnose skin cancers. Most dermatologists depend on biopsy of the lesion for definitive diagnosis rather than shave biopsy. Pathologists examine histological sections derived from such biopsies to make a definitive

diagnosis. The histological exam depends on evaluating cellular morphology and architectural distribution of the cancer cells within the lesion. In some instances, the definite histopathological diagnosis of malignancy is difficult; this is particularly evident when there is overlap in morphological features between some malignant and benign lesions [25]. A computer-based automatic diagnosis system can be an important tool for such difficult cases. The computer-based analytic tool is not necessarily more intelligent than the human brain, but it may be useful in extracting some information, such as texture features of the skin lesion that may not be easily seen by human eyes [26]. Therefore, it is important to develop efficient computer schemes for the physicians and pathologists through supporting their decision-making process based on digital images of skin lesions. It is anticipated that the computer-based image analysis could help in identifying early skin cancers and aid in early diagnosis in order to help reduce the death rate caused by these serious diseases [27].

2.2 Typical Medical Image

Medical images acquired for skin cancer diagnoses are pictures of lesion distributions are typically captured by an image acquisition system. Most of today's images are digital and may be archived and processed for analysis by a computer-assisted method. Examples of medical imaging include X-ray imaging, computed tomography (CT), magnetic resonance imaging (MRI), nuclear imaging, ultrasound imaging, photography, and microscopic images. The main imaging techniques are [28]:

- X-ray imaging technique: measures the absorption of shortwave electromagnetic waves which can vary between different tissues.

- Magnetic Resonance Imaging (MRI) technique: measures the density and molecular binding of selected atoms, most notably hydrogen which may vary with tissue type, molecular composition, and functional status.
- Ultrasound imaging technique: captures reflections at the boundaries between and within tissues with different acoustic impedance.
- Nuclear imaging technique: measures the distribution of radioactive tracer material administered to the subject through the blood flow.

Following capturing and storing of clinical images the computer-aided diagnostic (CAD) system for skin cancer examination is one of the medical image analytic processing methods that has received considerable attention [27]. This procedure requires the image to be managed in a series of sequential steps before actual analysis.

2.3 Image Preprocessing

The aim of preprocessing is to improve the image data such that there is suppression of undesired distortions in the image and enhances some image features relevant for further processing and analysis task [29]. The main purpose of this phase is to improve the quality of skin lesion image by removing unrelated and surplus parts in the background of the image before further processing. Effective preprocessing techniques can greatly improve the accuracy of the system [30].

2.3.1 Image Enhancement and Noise Removing

The basic goal of image enhancement is to process the image so that visual information can be viewed and assessed by software systems with greater clarity. The primary condition for image enhancement is that the desired information to be extracted, emphasized or restored must exist in the image [31]. Perhaps the most critical goal of image enhancement is that the processed

image should be more suitable than the original one for the required task of additional image analysis. This makes the process of image enhancement, by its nature, rather subjective and, hence it is difficult to quantify its performance [32].

Image filtering is one of the most important operations used in image enhancement. Image filtering is the first step within medical image preprocessing. The goal of the process is a smoothing of the image, canceling noise in the image while preserving the integrity of edges and information details. Medical images usually contain one or more types of noise or artifact. Because of diagnostic and therapeutic applications being dependent on interpretation of the image, this issue is critical. Therefore, it is necessary to remove the noise in the preprocessing procedure before any image analyses can be performed [33]. The two important image filters are:

- Median Filter

A Median filter is a nonlinear spatial filter based on the order statistics theory. In order to perform median filtering of a pixel in an image, color and brightness values of the pixels in its neighborhood are first sorted. The median is determined, and the value of this analysis is assigned to that pixel. A median filter is conducted by moving a pixel selection frame or window over the image and computing the output pixel as the median value of the brightness within the input window. If the pixel selection window is (j×k) in size, the (j*k) pixels can be ordered in brightness value from smallest to largest. If (j*k) is odd, then the median will be the (j*k+1)/2 entry in the list of ordered brightness [33].

- Gaussian Filter

The Gaussian filter is an important filter both for theoretical and practical reasons. The image is filtered using a discrete kernel derived from a radially symmetric form at the center of the image. The continuous 2-D Gaussian function defined in Eq. (2.1) [31]:

$$f(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \ldots\ldots (2.1)$$

Where σ is the standard deviation of the Gaussian function

2.3.2 Image Segmentation

Image segmentation is the next phase of image handling with its aim of partitioning images into areas which are meaningful for the next specific analytic task. Segmentation is the main step before the descriptions, recognitions or classifications of images is possible. There are two main segmentation methods: a) regions-based technique, in which likeness of pixels is discovered, b) boundaries-based technique, in which edges of the lesion are discovered and linked to shape boundaries. Segmentation is one of the difficult tasks in digital image processing. It is a technique that is employed to separate the lesion pigment from the adjacent normal skin. The dermoscopy image is a challenge for segmentation algorithms because of the different variances between skin and lesion. Some images have low disparity or smooth transmission between the lesion and the surrounding skin, and they commonly contain some skin features like black frame, air bubbles, hairs, blood vessels and skin lines. A number of segmentation algorithms have been proposed to address these challenges[34]. Different methods are used for segmentation such as: k-means clustering, thresholding or fuzzy c-mean clustering. But, in the recent days, swarm intelligence

algorithms are used as intelligent tools for image segmentation such as Particle Swarm Optimization (PSO).

2.4 Swarm Intelligence

Swarm Intelligence (SI) is an artificial intelligence technique based on the study of collective behavior in decentralized and self-organized object systems. The idea of SI comes from systems found in nature, including schools of fish, bird flocks and animal herding behaviors that can be applied to a computationally intelligent system. SI systems are typically made up of a population of agents interacting locally with one another and with their environment. Local interactions between such nodes often lead to the emergence of a global behavior[35]. The overall intent in image analysis is to look for the best solution as each iteration interacts with the previous one and collectively conclude on the ideal solution.

Mark Millonas in 1994, who developed swarm models for applications in artificial life, has articulated five basic principles of swarm intelligence [35]:

-The proximity principle: The swarm should be able to perform simple space and time computations.

-The quality principle: The swarm should be able to respond to quality factors in the environment.

-The principle of diverse response: The swarm should not commit its activity along excessively narrow channels.

-The principle of stability: The swarm should not change its behavior every time the environment changes.

-The principle of adaptability: The swarm must be able to change its behavior as long as the computational cost is not prohibitive.

Particle Swarm Optimization (PSO)

Optimization is the act of obtaining the best solution possible under specific situations. In PSO, each member of the population is called a particle and the population is called a swarm. Thus in image analysis each pixel can be the particle being analyzed. The basic PSO model consists of a swarm of particles, which are initialized with a population of random candidate solutions. They move iteratively through the *d*-dimension problem space to search for the new solutions, where the fitness f can be calculated as the certain qualities measure. Each particle has a position represented by a position-vector $\boldsymbol{xi}$ (i is the index of the particle) and a velocity represented by a velocity-vector $\boldsymbol{vi}$. Each particle remembers its own best position in a vector *i*-th, and its *d*-dimensional value is pi (t) [36].

The best position-vector among the swarm after each analysis iteration is then stored in the vector i-th, and its d-th dimensional value is g(t). During the iteration time *t*, the update of the velocity (vi (t+1) from the previous velocity to the new velocity is determined by Eq. (2.2). The new position (xi (t+1) is then determined by the sum of the previous position and the new velocity by Eq. (2.3).

$$vi\ (t+1) = w * vi + c1 * r1 * (pi\ (t) - xi\ (t)) + c2 * r2 * (g(t) - xi\ (t)) \ldots (2.2)$$

$$xi\ (t+1) = xi\ (t) + vi\ (t+1) \quad .... (2.3)$$

where i =1,2,…..,N; w is the inertia weight, r1 and r2 are the random numbers, which are used to maintain the diversity of the population, and are uniformly distributed in the interval [0,1] for the d-th dimension of the i-th particle. c1 is a positive constant, called coefficient of the self-recognition component; c2 is a positive constant, called coefficient of the social component. From Eq. (2.2), a particle decides where to move next, considering its own experience, which is the memory of its best past position, and the experience of its most successful particle in the swarm. In order to guide the particles effectively in the search space, the maximum moving distance during one iteration must be clamped in between the maximum velocity [−vmax,vmax].

The general concept behind the algorithm for the Particle Swarm Optimization (PSO) can be described in the algorithm (2.1)[37].

Algorithm (2.1) PSO Algorithm

Input: Initialize the algorithm parameters (c1, c2, w, vmax, Swarm_Size, Max_Iter).

Output: The optimization having the highest fitness as found by PSO.

Step 1: Randomly generate the initial particles and velocity to form a swarm.

Step 2: Calculate the fitness function of each of the particles.

Step 3: If the current position of the particle is better than the previous one based on step 2, then update the particles to indicate this fact.

Step 4: Find the best particle of the swarm. Update the positions of the particles by using equations (2.2) and (2.3).

Step 5: If the maximum number of iterations has exceeded or if the value with very high fitness is found, then go to step 6 or else go to step 2.

Step 6: Copy the best value and exit.

The stop criteria of PSO algorithm is usually one of the following [38]:

• Maximum number of iterations: the optimization process is terminated after a fixed number of iterations.

• Number of iterations without improvement: the optimization process is terminated after some fixed number of iterations without any improvement.

• Minimum objective function error: the error between the obtained objective function value and the best fitness value is less than a pre-fixed anticipated threshold.

2.5 Feature Extraction

A feature can be defined as an "interesting" part of an image. Feature extraction is used to reduce the original data by measuring certain properties, or features that distinguish one input image pattern from another [39]. Features extraction can be divided into two approaches, global and local, used to distinguish the images. The global features deal with the whole image without being attentive to pixels position, while local features characteristics care about objects in images. The features could be classified into three categories: Structural features, Physical features, and Mathematical features. Physical and structural features are commonly used by human beings for recognizing patterns amongst images because these features are easily detected by the eye. Machines can be designed to extract mathematical features or patterns within each image, which are somewhat difficult to determine by human beings. There are various methods of extracting features from images and some of these methods are structural, spatial, and statistical [39].

- Structural Methods

The structural approach to extract features in an image provides a good symbolic description of the image. These features are useful for synthesis objects, such as a geometric shape. In these features, a binary image is used to identify geometrical features such as [40]:

- Area

The actual number of pixels of the nucleus (white pixels) as in Eq.(2.4).

$$\text{Area} = \sum_{i=1}^{m} \sum_{j=1}^{n} I_{white}(i,j) \quad \ldots. (2.4)$$

Where m and n are height and width of the binary image (object mask)

And I_{white} is a white pixel in the binary image (object)

- Perimeter

The perimeter of an object is obtained from the boundary chain code by using the formula (2.5).

$$P = Ne + \sqrt{2}No \quad \ldots\ldots (2.5)$$

Where Ne is the number of even values and No is the number of odd values in the boundary chain code.

- Distance to Centroid

It is the distance between the points which are located on the boundaries of an object to the center of the nucleus of that object (object centroid). The distance to the centroid point is measured using Euclidian distance Eq. (2.6).

$$D(B,C) = \sqrt{(B_x - C_x)^2 - (B_y - C_y)^2} \quad \ldots\ldots (2.6)$$

Where: B_x,B_y are coordinates of boundary points and ($C_x$and C_y) are coordinates of the centroid point of the object as in formula (2.7 and 2.8).

$$C_x = \frac{1}{n}\sum_{i=1}^{n} x_i \ \ldots\ldots (2.7)$$

$$C_y = \frac{1}{m}\sum_{j=1}^{m} y_j \ \ldots\ldots (2.8)$$

Where: x represents the (x) coordinates of region pixels.

y represents the (y) coordinates of region pixels.

- Distance to Nearest Center

It is the distance between the boundary points of an object to the geometric center (image center). Distance to the nearest point is measured using Euclidian distance Eq. (2.9).

$$D(B,C) = \sqrt{(B_x - C_{row})^2 - (B_y - C_{col})^2} \ \ldots\ldots (2.9)$$

Where: B_x ,B_y are coordinate of boundary points.

$C_{row}$and C_{col} are coordinates of the centroid point for the geometric center (image center).

- Spatial Methods

Spatial features of an image are categorized by its color divided into major color bands, amplitude and spatial distribution [41].

- Intensity Color Features

Intensity features from an image, represent an image with their dominant color and average intensity. The color intensity feature is a three-dimensional vector composed of Red Green Blue (RGB) colors. Mean, standard deviation and variance can be extracted for each color as shown in Eq.s (2.10), (2.11) and (2.12).

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \ \ldots\ldots (2.10)$$

$$St = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \quad \ldots\ldots. (2.11)$$

$$Var = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2 \quad \ldots\ldots (2.12)$$

Where:

(μ) is the mean value of column data, (n) is the size of the column data, and (x_i) is the column data, (St) is the standard deviation value and (Var) is the variance of column data.

- Statistical Methods

Statistically based feature analysis of images involve first-order statistics, second-order statistics and higher-order statistics. These statistics properties give the distributions and relationships between the gray levels of an image. Features are computed from the statistical distribution pixels. The first-order deals with a unique pixel, the second-order deals with a pair of pixels, while the third-order deals with more than two pixels [42].

A- First-Order Statistics

Features in the first-order statistic are extracted exclusively from the information provided by the gray scale intensity histograms; ignoring the information about the location of pixels such as mean, standard deviation, skewness, energy and entropy [42].

B- Second-Order Statistics

Second-order statistics take the specific position of pixel relative to another pixel into account. Gray Level Co-occurrence Matrix (GLCM) is one of the statistical tools for extracting second order texture information from images. It is a representation of the spatial distribution and the interdependence of the levels within a local area[35]. The co-occurrence matrix represented by the probability of an N×N of image levels. The image must be quantized to a

finite number of levels before the co-occurrence matrix can be calculated. The direction of GLCM should be predefined, generally restricted to the values 0, 45, 90, and 135 [39].

GLCM is used to extract texture features; these co-occurrence based descriptors include[43]:

- Contrast (Inertia)

The inertia and the inverse difference moment are measures for the distribution of gray-scales in the image, and the intensity contrast between a pixel and its neighbor over the whole image is clarified in Eq. (2.13):

$$\text{Cont}= \sum_{i,j=1}^{n} |i-j| p(i,j) \ldots\ldots (2.13)$$

Where: Cont is contrast, n is Co-occurrence matrix dimension, p is the probability of Gray level Co-occurrence matrix (GLCM).

- Dissimilarity

In GLCM, dissimilarity is defined as in Eq. (2.14).

$$Dissimilarity = \sum_{i,j=1}^{n} (i-j)^2 p(i,j) \quad \ldots\ldots (2.14)$$

Where n is Co-occurrence matrix dimension, and p is the probability of GLCM,

- Energy

In GLCM, energy indicates the uniformity of the texture. It is defined as the sum of squared elements as in Eq. (2.15).

Energy $= \sum_{i,j=1}^{n} p(i,j)^2$

Where n is Co-occurrence matrix dimension, and p is the probability of GLCM.

- Entropy

In GLCM, entropy is defined as in Eq. (2.16).

$$Entropy = -\sum_{i,j=0}^{n} p(i,j)\log(p(i,j)) \quad \text{..... (2.16)}$$

Where: n is Co-occurrence matrix dimension, and p is the probability of GLCM.

Entropy measures the randomness of the pixel elements in the matrix. Therefore, a homogeneous image has a lower entropy than a heterogeneous image. Also, when energy gets higher, entropy should get lower.

- Autocorrelation

The autocorrelation function is used as a measure of periodicity of texture, as well as a measure of the scale of the texture primitives. Autocorrelation is related to the size of the texture primitive such as the fitness of the texture. If the texture is rough or unsmooth, then the autocorrelation function will go down slowly, if not it will go down very quickly. For normal textures, the autocorrelation function will show peaks and valleys. It has a relationship with the power spectrum of the Fourier transform as in Eq. (2.17).

$$Auc(i,j) = \frac{\sum_{u=1}^{n}\sum_{v=1}^{m} I(u,v)I(u+i,v+j)}{\sum_{u=1}^{n}\sum_{v=1}^{m} I^2(u,v)} \quad \text{.... (2.17)}$$

Where n and m are size of the image, I is the image Intensity u and v are related to the image in the frequency domain.

2.6 Feature Normalization

Feature scaling and data normalization are other names for feature normalization. It can be calculated as in Eq. (2.18).

$$Fnorm = \frac{feature - \mu_{feature}}{\sigma_{feature}} \quad \ldots. (2.18)$$

Where: Feature vector is the original feature vector, $\mu_{feature}$ is the mean of the feature vector and $\sigma_{feature}$ is the standard deviations of the feature vector.

2.7 Features Selection

Features selection technique is used for two main purposes: it increases the classifier efficiency by decreasing the size of features and increasing the classification accuracy by eliminating noise features[44]. The outcome is that a selection of a relevant subset of vectors with significant features are used in the training image set preparing for eventual image classification. A common features selection method is the mutual information (MI). MI measures the dependence between two random variables, in the context of feature selection between one feature and the class labels. If the MI is zero, both variables are independent and contain no information about each other; thus the feature is irrelevant for the target. Higher MI values indicate more information about the target and hence a higher relevance. Simple feature ranking chooses a certain number of the highest ranked features or all features above a threshold as relevant features for the learning machine as shown in Eq. (2.19).

$$MI(X,Y) = \sum_y \sum_x P(x,y)\ Log\ (\frac{P(x,y)}{P(x)p(y)}) \qquad \dots (2.19)$$

Where MI is mutual information of feature Vector X, Y is random variables, P(x) is the distribution probability of variable X and P(y) is distribution probability of variable.

2.8 Image Classification

Classification techniques predict discrete responses when classifying input data into categories. Typical applications for object classification include medical imaging, speech recognition, and credit scoring. Classification of images from skin lesions is the final step in computerized analysis of the images and to estimate whether the lesion is malignant or benign. To follow out the classification process, the current analytic systems utilize different classification methods to apply the feature descriptors that have been extracted from the images in the prior stages. The efficiency of these methods depend on both extracted features and the selected classifier[45].

- Support Vector Machine

Support Vector Machine (SVM) is a state of the art of classification method that was introduced in 1992 by Boser, Guyon, and Vapnik[44]. SVM classifier is widely used in bioinformatics because of its high accuracy, its ability to process high-dimensional data, and flexibility in modeling different kinds of data. SVMs have also been applied in various applications of biomedical and automatic recognition systems of different diseases, such as micro-classification detection in mammograms, polyps in CT colonography, cancer in optical images, and many melanoma detections in dermoscopic images[41].

SVM is a statistical machine learning method. When separating the different categories of objects, it applies a linear separating hyperplane for creating a classifier to maximize the margin that will include all the objects as shown in figure (2.1). The width of the margin between the different classes is considered as the optimization criterion. The margin is defined as the distance of the optimal hyperplane and the nearest training data points of a class. In cases of non-linearly separation of original input space, Guyon, Boser, and Vapnik, in 1992, presented an approach to generate nonlinear classifiers using kernel functions[46]. SVM transforms the original feature to a higher-dimensional feature space. The transformation may be obtained using different nonlinear mappings.

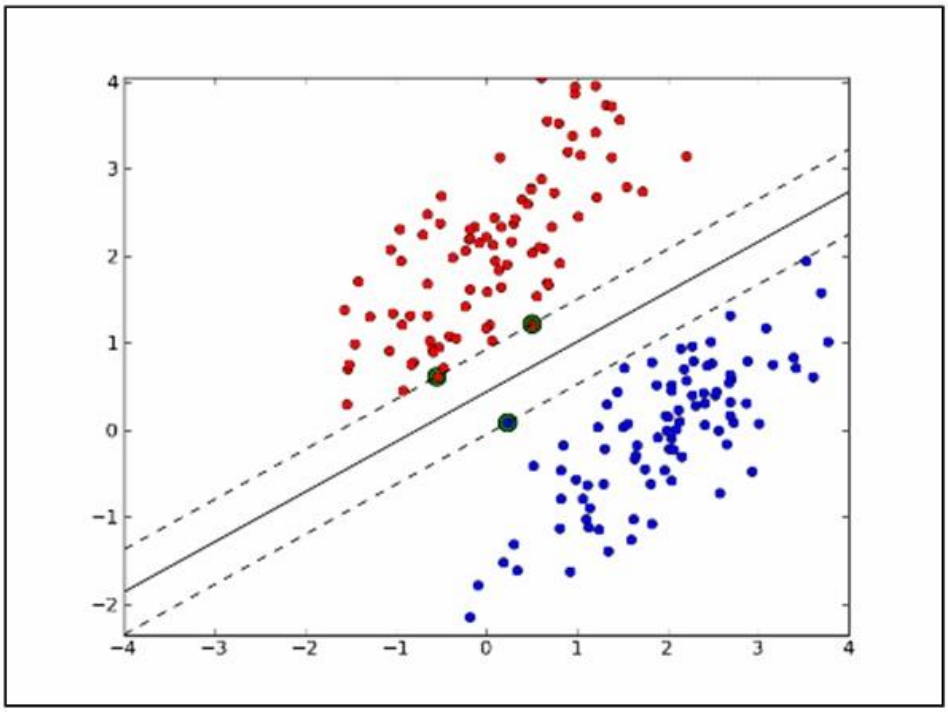


Figure (2.1) Support Vector Machine

However, computational analysis of features in images along vectors can be very hard. Instead, kernels offer an alternative. Thus, a single kernel function can be used to compute similarity

between images since a kernel is a similarity function. It is a function that is provided to the machine learning algorithm and it analyzes how similar two points are.

- Kernel Function in SVM

The kernel function K(x; y) may be chosen to suit the problem[40].

Linear Case:

In a case of two classes' problem with N number of training samples as shown in figure (2.2), each group of samples is described by a Support Vector (SV) X_i. It is composed of different data (features) with a number of dimensions from each side of the hyperplane. For a case of two classes, the label of a sample is assigned (-1) for the first class's label and (+1) for the other. The SVM classifier consists in defining the function of Eq. (2.20) [47].

$$f(x) = sign(\langle w, x \rangle + b) \quad \ldots. (2.20)$$

Where w is the distance between the hyperplane and the margin, w needs to be maximized. w has to be perpendicular to the hyperplane, x is the data (feature), and b is the bias.

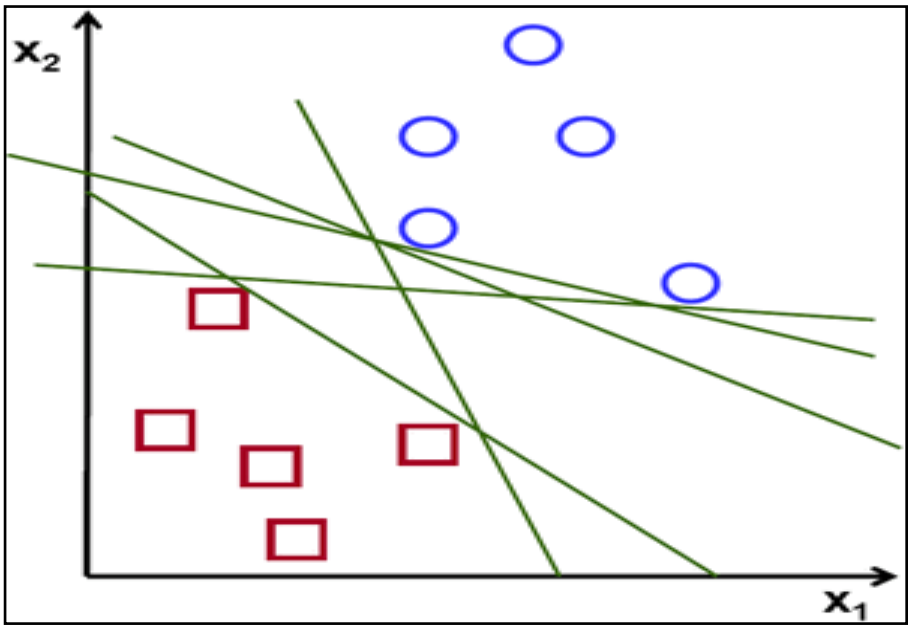


Figure (2.2) Linear Case in SVM

The sign of f(x) gives the label of the sample. The goal of the SVM is to maximize the margins between the support vectors and the optimal hyperplane. Therefore, it is easier to use the Lagrange multiplier without worrying about the constraints during the maximizing solution as in Eq. (2.21)[47].

$$f(x) = sign(\sum_{i=1}^{n} y_i \,.\, \alpha i \,\langle x.\, x_i\rangle + b) \;\ldots.\; (2.21)$$

Where (αi) is the Lagrange multiplier.

Non-Linear Case:

A non-linear case is difficult to solve by a conventional linear SVM, therefore soft margins are used by adding a noise data to solve a non-linear problem. The more efficient solution for SVM is to use a kernel. The kernel is a function that aims to separate the data

more easily by simulating the projection of the initial data in a feature space with higher dimension.

This case is illustrated in figure (2.3)[48].

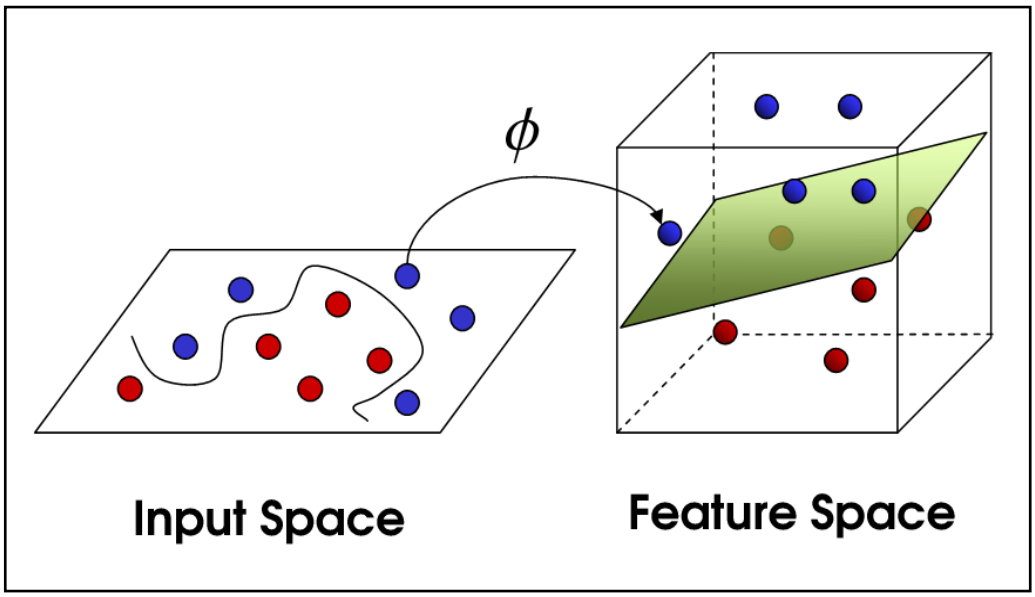


Figure (2.3) Non-Linear Case in SVM

In this new space, the data could become linearly separable or better structured. To apply this, the dot product ⟨xi, xj⟩ is replaced by the function in Eq. (2.22)[48].

$$K(x, xi) = \langle \emptyset(x)\emptyset(xi) \rangle \quad \ldots. (2.22)$$

Then the new function to classify the data are clarified in Eq. (2.23)[48].

$$(x) = sign(\sum_{i=1}^{n} y_i \,.\propto_i K\langle x.x_i\rangle + b) \quad \ldots\ldots (2.23)$$

Three methods of kernels are commonly used in SVM: Polynomial kernel, sigmoid kernel, and Gaussian kernel. These functions are shown in Eq.s (2.24, 2.25 and 2.26) sequentially[48].

$$K\langle x.x_i\rangle = \langle x.x_i\rangle^p \quad \ldots.. (2.24)$$

$$K\langle x.x_i\rangle = \tanh(\langle x.x_i\rangle + 1) \text{ ….. (2.25)}$$

$$\mathrm{K(x,y)} = \exp\left(\frac{-\|x-y\|^2}{2\sigma^2}\right), \sigma \neq 0 \text{ …… (2.26)}$$

CHAPTER 3 Methods and Algorithms Applied

The proposed methodology of skin cancer detection from skin lesion images using the previously defined methods and algorithms are clarified in this chapter. The general proposed detection system consists of sequential stages as shown in figure (3.1).

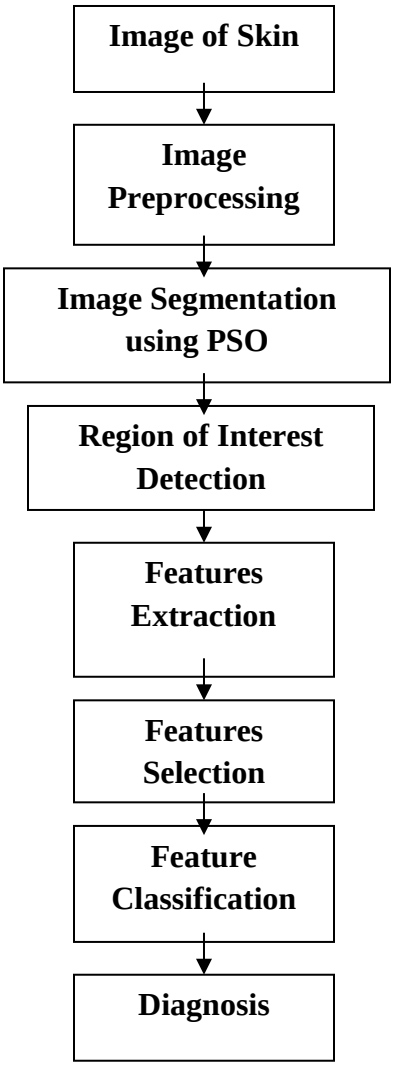


Figure (3.1) Block Diagram of The Proposed Detection System

The proposed tumor detection system starts by preprocessing the skin lesion image from two groups of skin images, benign and melanoma. These images are captured by a specialized camera used for this purpose. Removing hair from the skin image is considered as a part of the preprocessing stage to eliminate obscuring objects. However, PSO is used as a segmentation tool. After the segmentation, a region of interest (ROI) relevant to the skin lesion is identified. These phases of image preparation are followed by the features extraction to find the characteristics of the skin image serving as input for computational analysis. Some of these features are identified as local features and some as global, and each group will be stored in a separate dataset. The local datasets are used for an evaluation and enhancement of the predefined image in order to get the best ROI. Features extraction is followed by features selection that is used to choose only the discriminating features for image classification. The stages in figure (3.1) are explained in sections 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6.

3.1 Image Preprocessing

The preprocessing of a skin cancer image includes: evaluating the color qualities of the image by splitting the image into its three main color bands, Red, Green, and Blue, and applying (3×3) median filter to each band, and applying Eq. (3.1) by three morphological masks, zero, 45, and 90 degrees as shown in figure (3.2).

$$g(x,y) = \begin{cases} 1 \; iff(x,y) > T \\ 0 \; iff(x,y) \leq T \end{cases} \quad \ldots. (3.1)$$

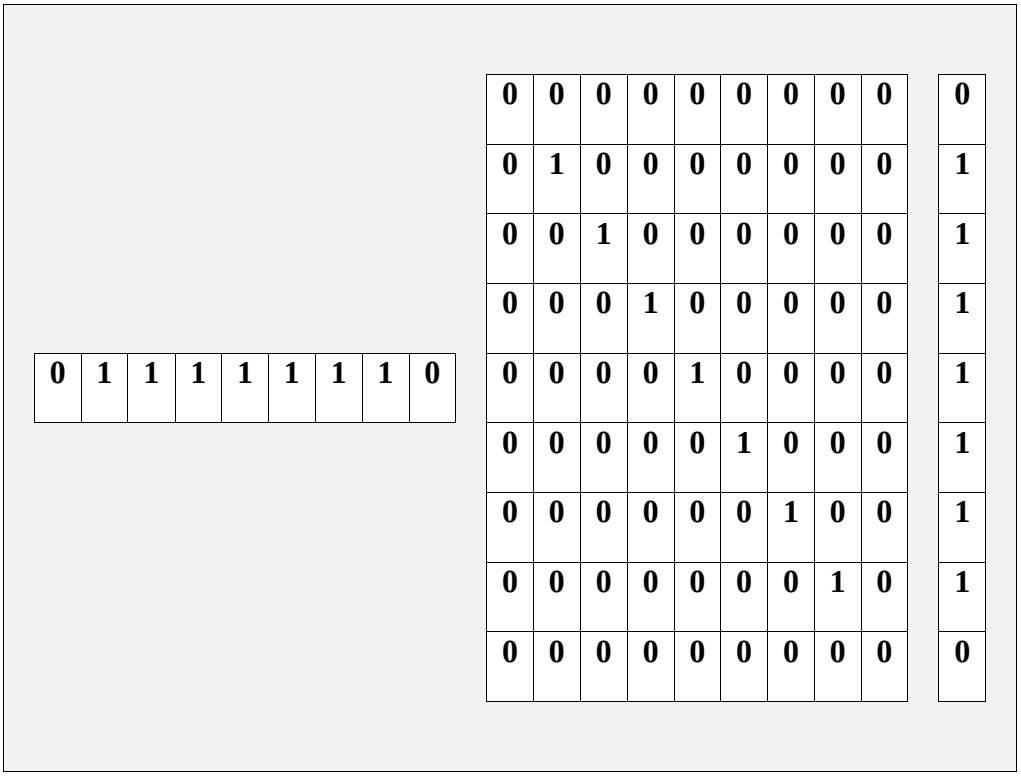


Figure (3.2) Morphological Mask

The three results from the three masks are gathered as one output to replace the input image, then overlaid images of hairs are removed and a Gaussian filter is applied using Eq. (2.1), where the filter size is (3×3) pixels and the coefficient sigma is equal to 1.1. The additional details of the image preprocessing stage are illustrated in figure (3.3).

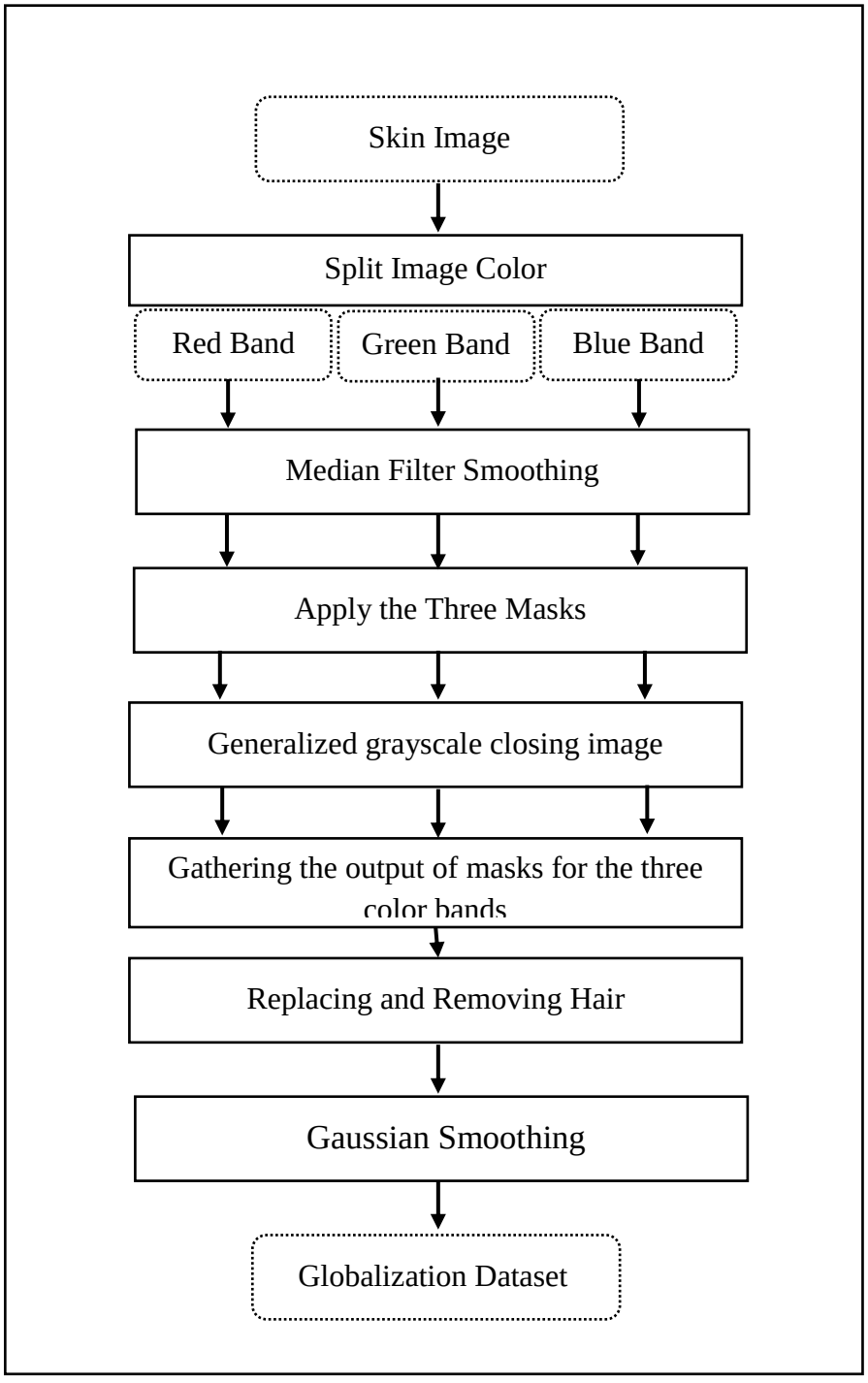


Figure (3.3) **S**kin Image Preprocessing

Also, the image of skin preprocessing can be illustrated in algorithm (3.1).

Algorithm (3.1) Image Skin Preprocessing

Input: Colored Image (Im)

Output: Preprocessed Image (PrIm)

1: Begin

2: Specify Mask0, Mask45, and Mask90

3: Im=Read Color Image RGB

4: Imr=Im.red; Img=Im.green; Imb=Im.blue

5: Apply median filter on three bands (Imr, Img and Imb)

6: Apply image dilation followed by erosion on three bands (Imr, Img and Imb) with mask0

7: Apply image dilation followed by erosion on three bands (Imr, Img and Imb) with mask 45

8: Apply image dilation followed by erosion on three bands (Imr, Img and Imb) with mask 9 0

9: Gather (union) three color bands by replacing max of three color bands

10: Construct binary mask equivalent to hair filling

11: From the processed image get the filled region to replace in the original image

12: Apply Gaussian smoothing filter on the resulting image to find PrIm

13: Return PrIm

14: End Algorithm

In the algorithm (3.1), the dilation of a binary image can be performed by successively placing the center pixel of the structuring element on each background pixel. If any of the neighboring pixels are foreground pixels (value 1), then the background pixel is switched to the foreground. Formally, the dilation of image A by structuring element B, is denoted as (A$\oplus$ B).

Opening is defined as an erosion followed by a dilation. The image erosion process removes some foreground (bright) pixels from the edges of regions of foreground pixels, while the dilation part adds foreground pixels. The foreground features remain about the same size, but their contours are smoother. Therefore, the opening function can be stated as in Eq. (3.2).

$$AoB = (A \ominus B) \oplus B \ldots (3.2)$$

Closing is defined as a dilation followed by an erosion. Closing smooths the contours of foreground objects in the image but, in contradistinction to opening, it merges narrow breaks or gaps and eliminates small holes as stated in Eq. (3.3).

$$A \bullet B = (A \oplus B) \ominus B \ldots (3.3)$$

3.2 Image Segmentation

The lesion area is separated from healthy skin using image segmentation. The segmentation method is very effective in giving better diagnosis results. Therefore, PSO is used as an intelligent tool to perform an image segmentation process depending on intensity distributions of the skin image. Hence, PSO makes image segmentation as an optimization of the image. The proposed approach is an intelligent search of the best class for each pixel in the skin image using a population of artificial particles. Classes of pixels are assigned by each particle iteratively depending on the fitness function. Algorithm (2.1) is used to implement PSO method for segmenting, depending on

Eq.s (2.2 and 2.3). Practically, the maximum moving distance during one iteration is clamped between [-5, 5]. Random numbers (r1 and r2) are between [0, 1].

The first step in the proposed PSO method is determining the number of particles which is equal to 150 for the first iteration and one example image from the dataset. In the beginning, the positions of these particles are set randomly, however, the velocities of them are set to zero. Note that the number of iterations is (4 × length of skin image × width of skin image).

The fitness function is the variance of the intensity distributions of skin image between classes as clarified in Eq. (3.4).

$$Var = \sum_{j=1}^{n} w_j (\mu_j - \mu_t)^2 \quad \ldots. (3.4)$$

Where: j is a specific class.

μ_j is the average occurrence of class j.

w_j is the probability of occurrence of class j, which is shown in Eq.(3.5).

$$w_j = \sum_{i=1}^{t_j} p_i \quad \ldots. (3.5)$$

Where: p_i is the probability of occurrence of each pixel in the skin image, which can be obtained by Eq. (3.6).

$$p_i = \frac{h_i}{N} \quad \ldots. (3.6)$$

Where: i is an intensity level.

N is the total number of the pixels in the skin image.

h_i is the image histogram.

μ_j can be obtained by Eq. (3.7).

$$\mu_j = \sum_{i=1}^{t_j} \frac{ip_i}{w_j} \quad \text{.... (3.7)}$$

μ_t can be obtained by Eq. (3.8).

$$\mu_t = \sum_{i=1}^{L} ip_i \quad \text{.... (3.8)}$$

Where: L is the maximum intensity.

Maximum variance can be used as a dividing tool for dividing a skin image into two regions based on intensities.

3.3 Region of Interest Detection

The region of interest detection process of a given skin lesion image is based on the proposed method and depends on the PSO segmentation method which is explained in section (3.2). The region of interest (ROI) is later used for localization of image feature extraction.

The ROI can be calculated from the boundary region that is specified by PSO segmentation. This is done by taking the maximum and minimum values of x and y in the boundary region. The color of ROI can be converted for convenience.

The region of interest detection relies on four factors, listed below, that are used to determine the differences between the segmented image after using PSO and the ground truth lesion GT (binary image) that is extracted manually by a medical expert and saved in the dataset:

- Hammoude distance (HM): HM compares a pixel by pixel of the pixels enclosed by the two borders, can be calculated by Eq. (3.9).

$$HM(SR, GT) = \frac{\#(SR \cup GT) - \#(SR \cap GT)}{\#(SR \cup GT)} \quad \text{.... (3.9)}$$

Where (SR) is the lesion segmented image by using PSO (binary image), (GT) is the ground truth lesion (binary image), explaining as follows:

($SR \cap GT$) The number of white pixels in the common areas only,

($SR \cup GT$) The number of white pixels in the white areas of both images.

- True detection rate (TDR): TDR computes the average of pixels that are categorized as lesions by both the medical expert and the automatic segmentations; can be calculated by Eq. (3.10).

$$TDR(SR, GT) = \frac{\#(SR \cap GT)}{\#GT} \quad \ldots. \ (3.10)$$

- False position rate (FPR): FPR computes the average of pixels that are categorized as lesions by the automatic segmentations that were not categorized as lesions by the medical specialist; these can be calculated by Eq. (3.11).

$$FPR(SR, GT) = \frac{\#(SR \cap \overline{GT})}{\#GT} \quad \ldots (3.11)$$

- Gross error: Gross-error counts if HM ≥ 30 %

The results of these four parameters are very important for evaluating the effectiveness of the image segmentation method.

ROI will be used for the image's feature extraction which consists of three categories: color feature, geometric feature and texture feature. The geometric feature depends on the binary image while color feature depends on the color within the interested region that will be specified by multiplying the binary mask of ROI by the three bands of color skin image. Texture feature

depends on a grayscale image that will be specified from the ROI in the color image. The circularity of ROI is found for each skin image that will be useful for diagnosing and specifying the geometric shape of lesion.

3.4 Image Feature Extraction

Accurate skin lesion image classification depends on features that are extracted from within the ROI. In the method utilized in this study, three types of features are considered in subsections (3.4.A), (3.4.B) and (3.4.C).

A. Geometric Features

These features depend on a binary mask dependent on the shape generated in the mask. The binary mask is extracted from a skin image using PSO segmentation method as shown in figure (3.4). Geometric features will be extracted from this binary mask.

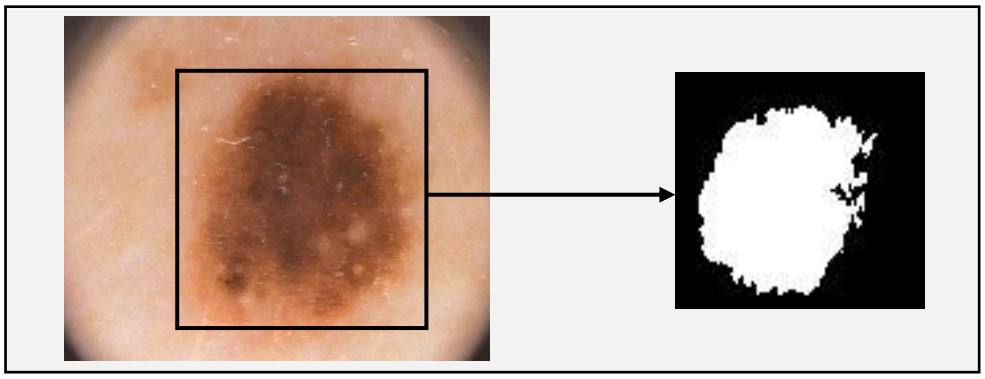


Figure (3.4) Binary Mask of ROI

- **Area**

The mean area of ROI (first feature) is computed by the actual number of pixels of the nucleus depending on equation (2.4) as shown in figure (3.5).

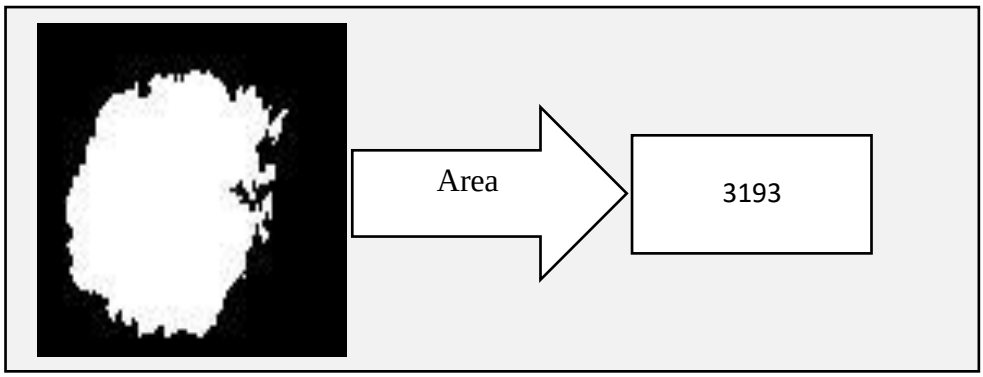


Figure (3.5) Area of ROI

- **Perimeter**

The perimeter feature is computed by summing up the distances between each adjoining pair of pixels around the border of the nucleus, depending on equation (2.5). From this result, the mean and standard deviation (second and third) can be calculated as shown in figure (3.6).

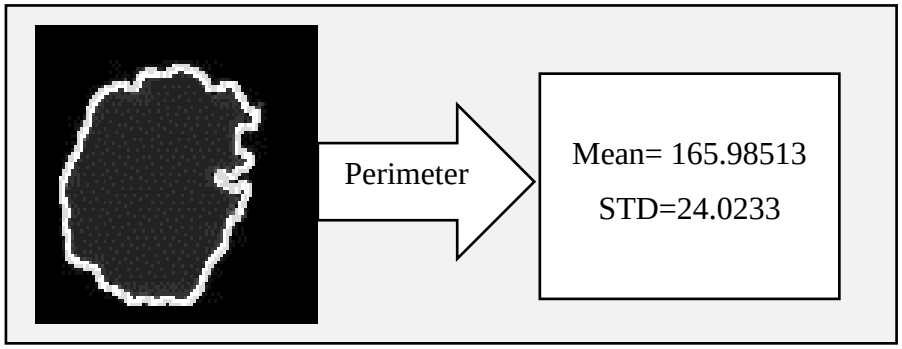


Figure (3.6) Mean and STD of ROI Perimeter

- **Distance to Centroid**

It is the distance between the boundary points of the object to the center of the nucleus of the object (object centroid). It depends on equations (2.6, 2.7 and 2.8). Also, the mean and standard deviation can be obtained as fourth and fifth features depicted in figure (3.7).

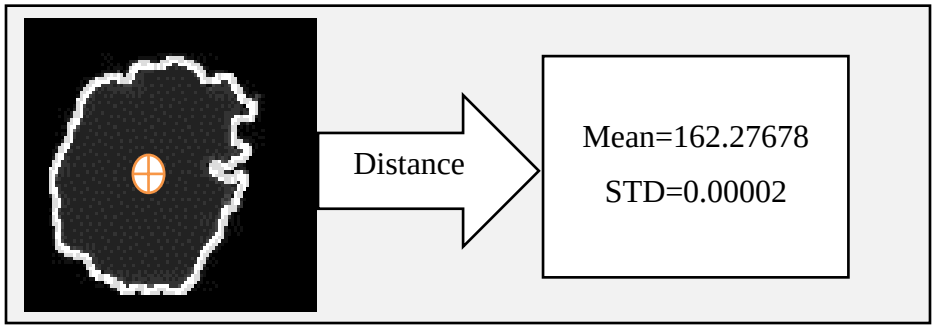


Figure (3.7) Mean and STD of ROI Distance

- **Distance to Center of Nearest**

The distance between the boundary points of the object to the geometric center (image center). It can be calculated using equation (2.9). Also, mean and standard deviation can be computed as the sixth and seventh features as shown in figure (3.8).

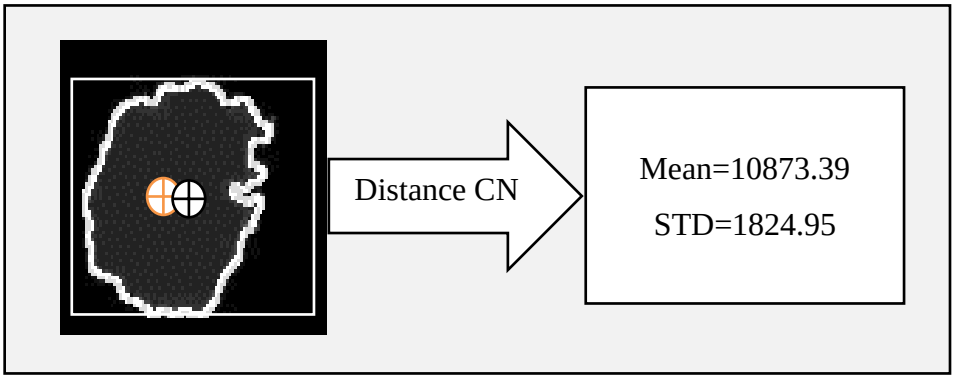


Figure (3.8) Mean and STD of ROI Distance to Nearest

B. Color Features

The color features are used in this study as global and local features as shown in figures (3.9) and (3.10).

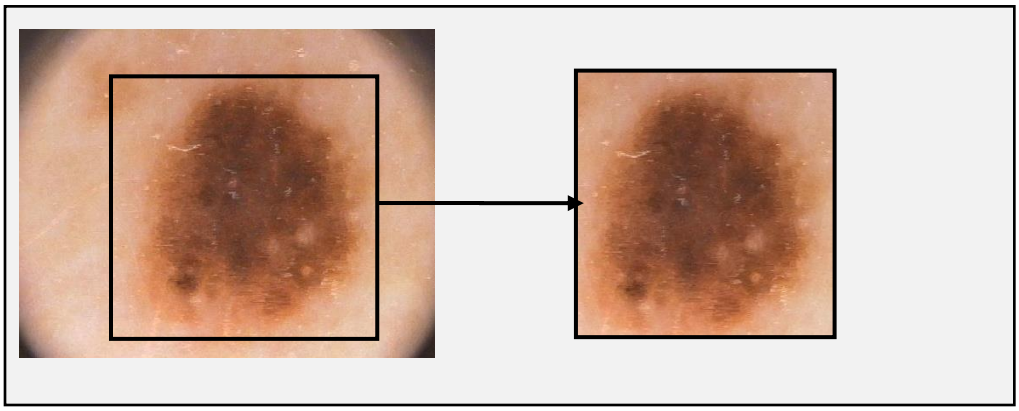


Figure (3.9) Global and Local Region of Color Skin Image

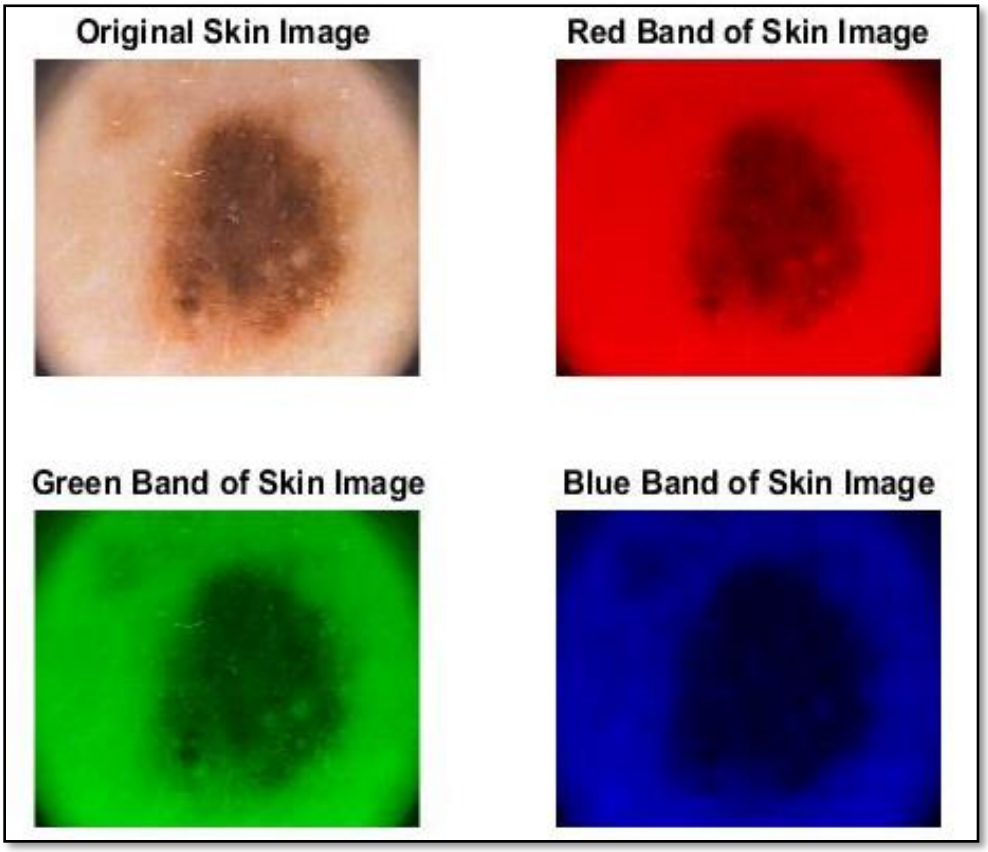
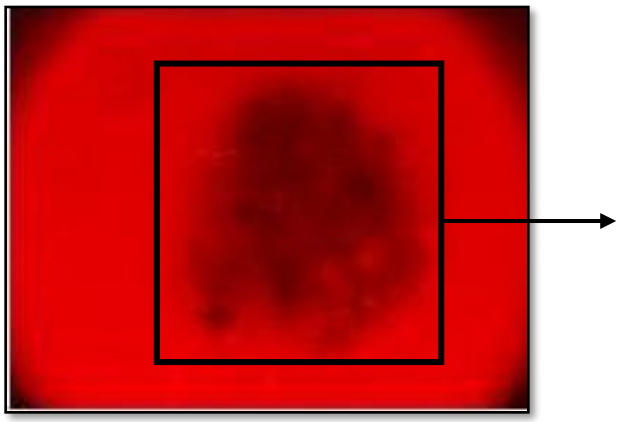


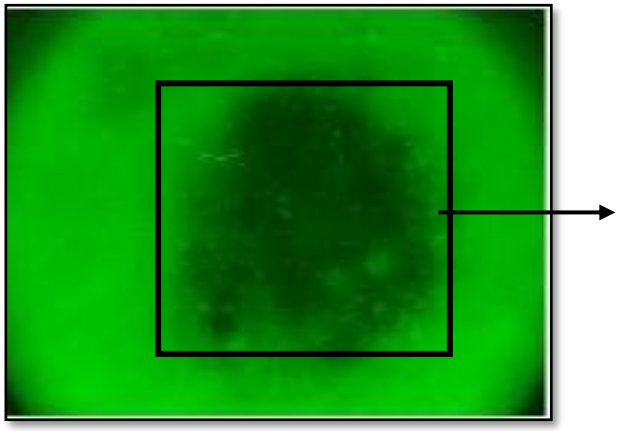
Figure (3.10) RGB Splitting Color of Skin Image

Intensity values are represented as matrix shown in figures (3.11, 3.12, and 3.13) for red, green, and blue respectively.



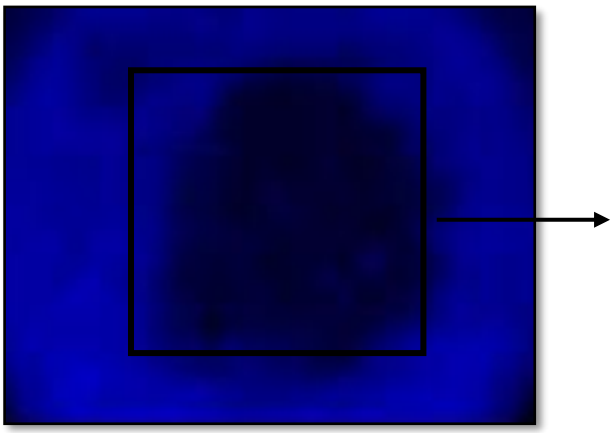
214	206	205	208	206	206	209
210	203	205	211	209	206	206
205	200	204	212	211	206	204
202	199	204	209	207	204	204
202	202	206	206	201	202	204
198	203	207	204	199	202	204
194	200	206	205	200	203	206

Figure (3.11) Red Band Color Samples



165	157	156	159	157	157	158
161	154	156	162	160	157	155
154	149	155	163	160	155	153
151	148	153	160	156	153	153
149	151	155	155	150	151	154
145	150	156	153	148	151	154
141	147	153	152	148	151	156

Figure (3.12) Green Band Color Sample



171	163	162	165	163	163	163
167	160	162	168	166	163	160
161	156	161	169	165	160	158
158	155	160	166	161	158	158
155	156	160	160	155	156	157
151	156	161	158	153	156	157
147	153	159	158	152	155	159

Figure (3.13) Blue Band Color Sample

In order to calculate the mean, standard deviation and variance for each color band, equations (2.10, 2.11 and 2.12) are used. The mean of red, green, and blue bands represents the eighth, ninth and tenth features respectively. The standard deviation of red, green, and blue bands represents the eleventh, twelfth and thirteenth features respectively. The variance of red, green, and blue bands represents the fourteenth, fifteenth and sixteenth features respectively. Mean and STD ROI color features are shown in figure (3.14). Mean and STD ROI Color Features (Variance) are shown in figure (3.15).

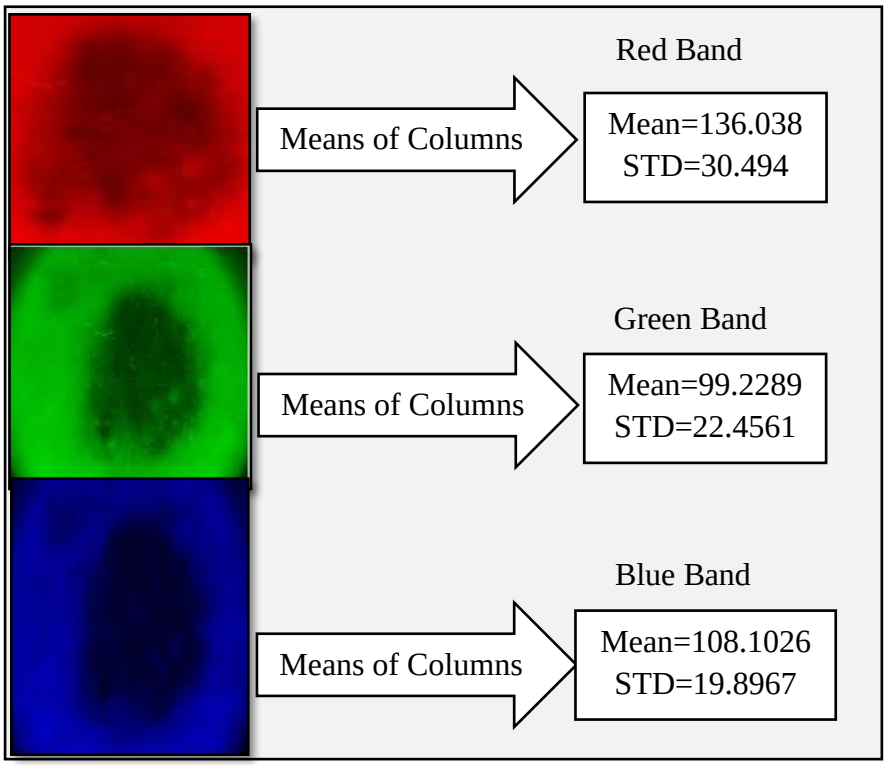


Figure (3.14) Mean and STD ROI Color Features (Mean)

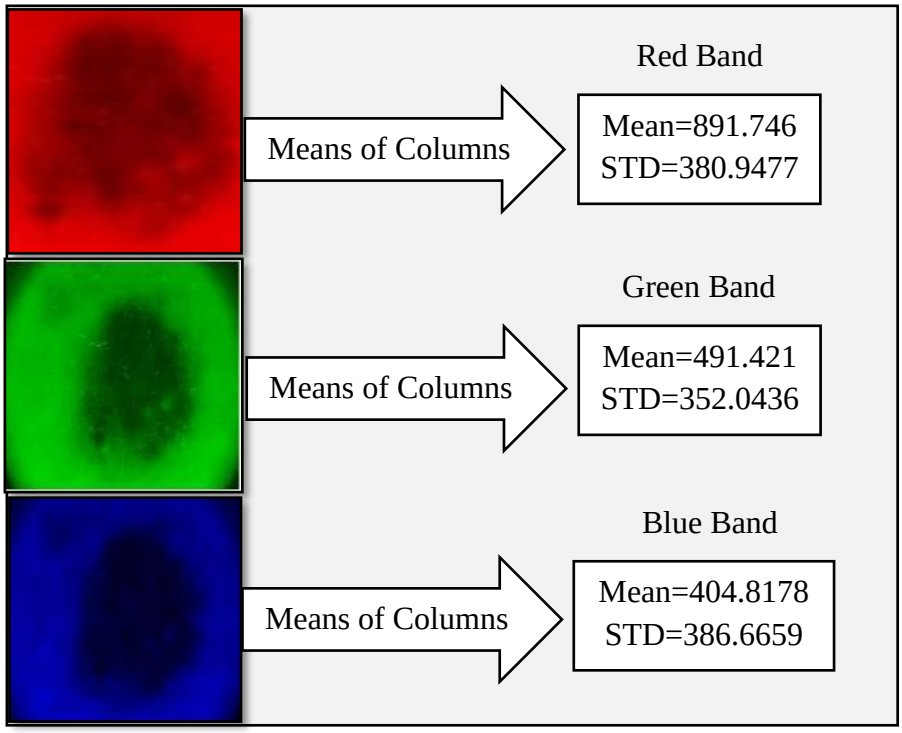


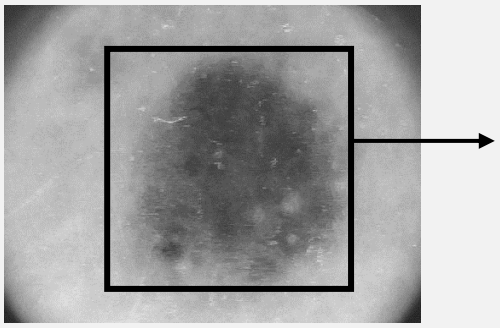
Figure (3.15) Mean and STD ROI Color Features (Variance)

C. Texture Features

Texture features depend on the gray level co-occurrence matrix (GLCM) and the gray level run length matrix (GLRLM).

- **Gray level co-occurrence matrix (GLCM) Features**

The content of the GLCM matrix depends on the scan direction and the distance relationship between pixels. GLCM features are local features which depend on the relation of two adjacent pixels that give some information about location. The samples of grayscale image matrix are shown in figure (3.16). GLCM finds all probabilities of occurrence pixels' values in a gray level. These samples are represented as sparse matrix for counting values and the probabilities of these samples that are divided by total number of occurrences.



180	172	171	174	172	172	174
176	169	171	177	175	172	171
170	165	170	178	176	171	169
167	164	169	175	172	169	169
166	167	171	171	166	167	169
162	167	172	169	164	167	169
158	164	170	169	164	167	171

Figure (3.16) Gray Image Matrix Sample

The measures require that each co-occurrence matrix cell does not contain a count, but rather a probability. The GLCM features that differentiate pixels within images are shown in table (3.1).

Table (3.1) GLCM Features

Feature Number	Feature Name
17	contrast
18	dissimilarity
19	energy
20	entropy
21	homogeneity
22	variance (sum of square)
23	inverse difference normalize
24	inverse difference moment
25	maximum probability
26	sum of entropy
27	sum of variance
28	difference of entropy
29	difference of variance
30	information measures of correlation first
31	information measures of correlation second
32	cluster shade
33	cluster prominence
34	auto correlation

35	correlation

The total extracted GLCM features are clarified in algorithm (3.2).

Algorithm (3.2) Extracted GLCM Features

Input: Image Color(Imcolor), angle, Distance

Output: Pcm

Steps:

1: Begin

2: Ang=Read Angle // Direction angle

3: Distance=2 // Distance between neighbor pixels

4: ImC=ReadColorImage(ImHr or ROIm) //

5: ImGray=ConvetToGray(ImC) //

6: Set h= ImGT.height

7: Set w= ImGT.width

8: Set CM(1 to h, 1 to w)=0, count=0

9: For i=1 to h

10: For j=1 to w-1

11: Temp1=ImGray(i,j)

12: Temp2=ImGray(i,j+1)

13: CM(Temp1,Temp2)= CM(Temp1,Temp2)+1

14: CM(Temp2,Temp1)= CM(Temp2,Temp1)+1

15: count=count+1

16: End for j

17: End for i

18: For each element in CM

19: Pcm(i,j)=CM(i,j)/count

20: End for

21: Return Pcm

22: End Algorithm

- **Run Length Matrix**

A run is a string of pixels which has the same intensity along a specific linear direction. The run length matrix is calculated for horizontal and vertical directions. After the calculation of total runs of pixels' intensities, then the probability is found by dividing each value by total occurrence issues; this will give a new matrix, gray level run length matrix (GLRLM). This in turn, is used to calculate the run length features values. Run length features of an image being analyzed are shown in table (3.2).

Table (3.2) Run Length Features

Feature Number	Feature Name
36	Short Run Emphasis
37	Long Run Emphasis
38	Gray Level Distribution
39	Run Length Distribution
40	Run Percentage
41	Low Gray-level Run Emphasis
42	High Gray-level Run Emphasis
43	Short Run Low Gray-level Emphasis
44	Short Run High Gray level Emphasis
45	Long Run Gray-level Emphasis
46	Long Run High Gray level Emphasis

Algorithm (3.3) illustrates the steps to calculate the run length matrix features.

Algorithm (3.3) Extracted Run Length Features

Input: ImHr, MaxRun, ImGT

Output: Prlm

Steps:

1: Begin

2: Read MaxRun

3: ImC=ReadColorImage(ImHr or ROIm) //

4: ImGray=ConvertToGray(ImC) //

5: Set h= ImGT.height

6: Set w= ImGT.width

7: Set RLM(0 to 255, 1 to MaxRun)=0

8: For k=1 to MaxRun

9: For i=1 to h

10: For j=1 to w-k

11: If I sEqual(ImGray(i,j to j+k-1)) then

12: F=1

13: j=j+k

14: RLM(ImGray(i,j),k)=RLM(ImGray(i,j),k)+1

15: count=count+1

16: End if

17: End for j

18: End for i

19: End for k

18: For each element in RLM

19: Prlm(i,j)=RLM(i,j)/count

20: End for

21: Return Prlm

22: End Algorithm

3.5 Feature Selection

The features selection step is used to find feature subsets that are relevant to the result by using the mutual information method. This method gives weight to each feature; these weights will then be sorted in order to rank the features from high to low. This process is used as a filter to reduce features dimensions in order to decrease the complexity. Mutual information is modeled according to equation (2.19). The elimination of significant features might decrease the accuracy rate of the classification. However, some features of the dataset may have no weight at all or have a high level of noise. The elimination of such features can increase the search speed and the accuracy rate of the image analysis.

3.6 Classification

Classification is the last step in the methodology used in this image anlaysis system. This step uses the SVM method for the eventual diagnosis based on the overall image analysis. SVM considers two approaches, global and local, depending on the global features and local features from within the ROI. Two kinds of kernel function are used for classifying data: linear, and a polynomial kernel function. The detailed steps of the classifier process are shown in figure (3.17).

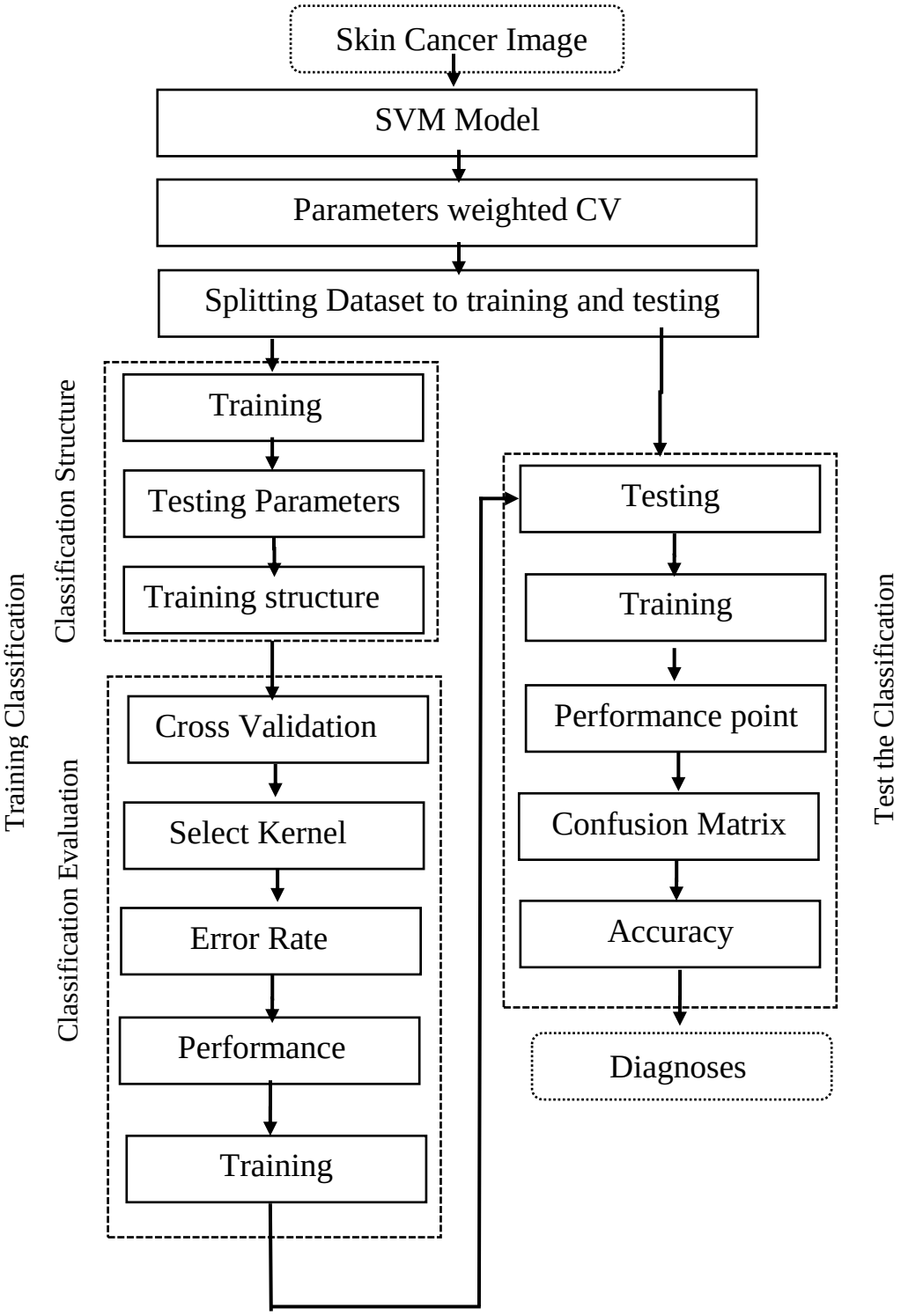


Figure (3.17) Skin Image Classifier

A. SVM Model

The logic in implementing the SVM is to create a hyperplane for splitting the data of the pixels analyzed in each image. The maximization of margins between negative and positive examples depend on SVM. The kernel function is related to the performance of classification. There are several types of kernel functions used in SVM, linear kernel and polynomial kernel.

- **Linear Kernel**

The linear kernel is used when the number of features in the image analyzed is larger than the number of observations. It is implemented in the proposed system using equation (2.20).

- **Polynomial Kernel**

The polynomial kernel affects the features of the skin image relative to the similarity and combination (collaboration) perspectives. It is implemented in the current study using equation (2.24).

B. Training

SVM is a discriminative model which uses a subset of the training image data in order to construct a protocol that separates skin image that is identified as benign from melanoma. Two kernels, linear and polynomial, are used for finding a hyperplane that separates feature vectors into two regions. Only feature vectors that are close to the decision surface are used by SVM.

- **Training Structure**

The structured SVM is a machine learning algorithm that generalizes the image classifier. After training, the structured SVM model is capable of predicting the classification of any new image samples. This stands in contrast to the simpler approaches of classification where input data (instances) are mapped to "atomic" labels without any internal structure, or regression.

- **Cross-Validation**

Classifiers accurately predict training data that are known as class labels. Training data is separated into two parts which are unknown to the training classifier. As a result, the prediction accuracy on this set can more precisely reflect the performance on classifying unknown data. An improved version of this procedure is cross-validation. In v-fold cross-validation, the training set is first divided into v subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining (v − 1) subsets. Thus, each instance of the whole training set is predicted once. Consequently, the cross-validation accuracy is the percentage of data which are correctly classified. The cross-validation procedure can prevent the over fitting problem.

C. Testing

The testing phase classifies each row of the vector of features that is not labeled; while in training the rows are labeled, using the information in a support vector machine classifier structure. Note that the structure is created using the training step. The sample is a matrix of observations where each row corresponds to it or replicates, and each column relates to a feature or variable. Thus, the sample should have the same number of columns as the training data. This is because the number of columns defines the number of features.

- **Performance**

Performance evaluates the prediction against true labels. It depends on the calculation of the confusion matrix that contains True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). There are multiple measurements such as: sensitivity or true positive rate (TPR), specificity (SPC) or true negative rate, precision or positive predictive value (PPV), and accuracy. These parameters can be used for the evaluation of the results.

Finally, the total processing of the images in this study is shown in figure (3.18).

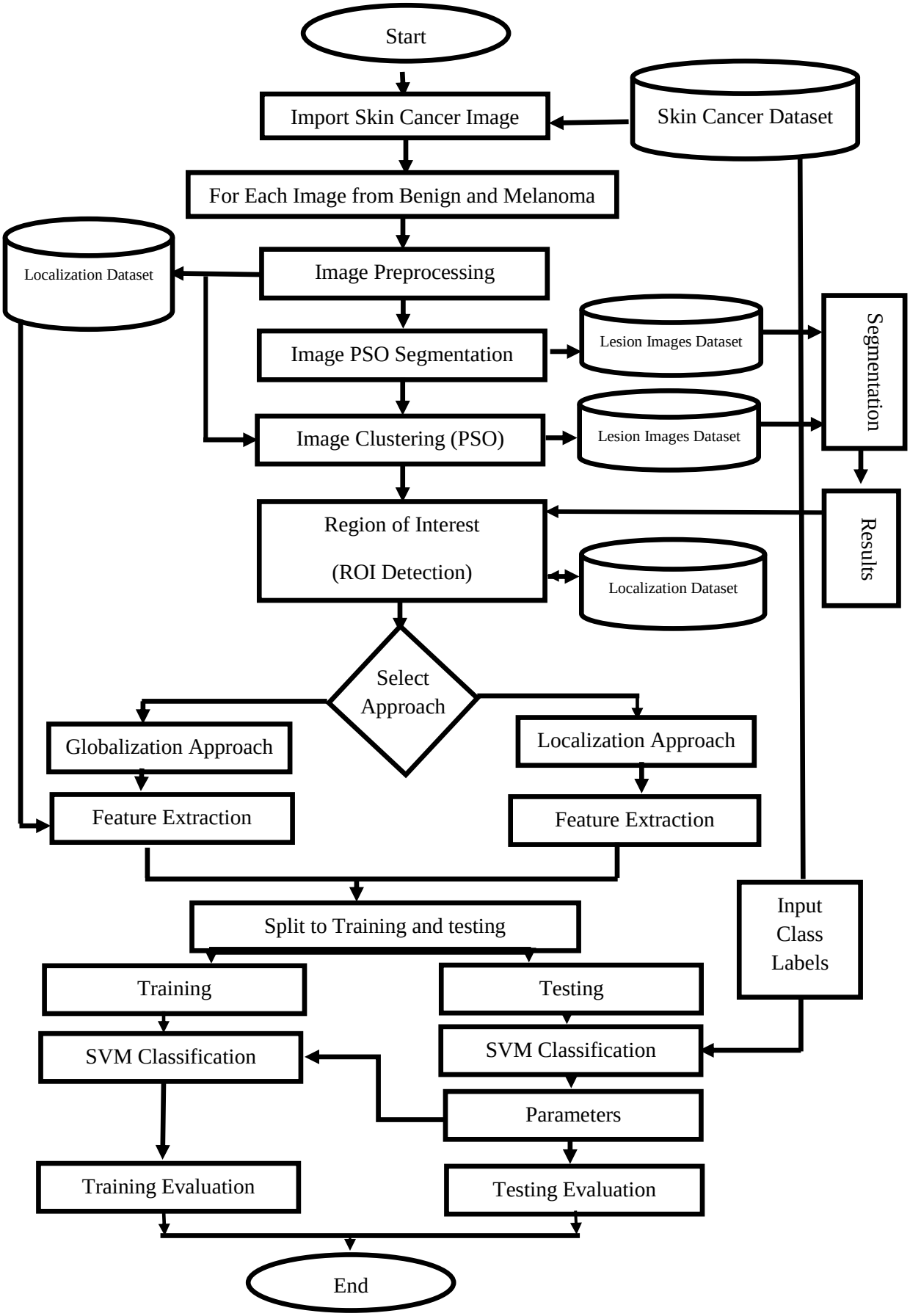

Start
Import Skin Cancer Image
Skin Cancer Dataset
For Each Image from Benign and Melanoma
Localization Dataset
Image Preprocessing
Image PSO Segmentation
Lesion Images Dataset
Segmentation
Image Clustering (PSO)
Lesion Images Dataset
Region of Interest
(ROI Detection)
Localization Dataset
Results
Select
Approach
Globalization Approach
Localization Approach
Feature Extraction
Feature Extraction
Input
Class
Labels
Split to Training and testing
Training
Testing
SVM Classification
SVM Classification
Parameters
Training Evaluation
Testing Evaluation
End

Figure (3.18) Flowchart of the Detailed Proposed System

Chapter 4 Results

This chapter demonstrates the results at different stages in the image analytics and diagnosis detection system. Knowing that, the implementation of the proposed system is done using a PC with a processor: (Intel (R) COR(TM), I5 3210M CPU @ 2.50 GHz, and 4GB of RAM). The operating system is Windows 8 (64 bits) and MATLAB 2015(The Mathworks) is used as a software tool.

PH^2 dataset includes 200 dermoscopic images of melanocytic lesions, is used in this study. This dataset consists of 80 common nevi, 80 atypical nevi, and 40 melanomas. Medical annotation of all the images, namely medical segmentation of the lesion, clinical and histological diagnosis and the assessment of several dermoscopic criteria, are included in the PH^2 dataset. These several dermoscopic criteria are comprised of colors, pigment network, dots/globules, streaks, regression areas and blue-whitish veil.

Some of the benign images that are used in the proposed system are shown in figure (4.1), while some of the melanoma images that are used in the proposed system are shown in figure (4.2).

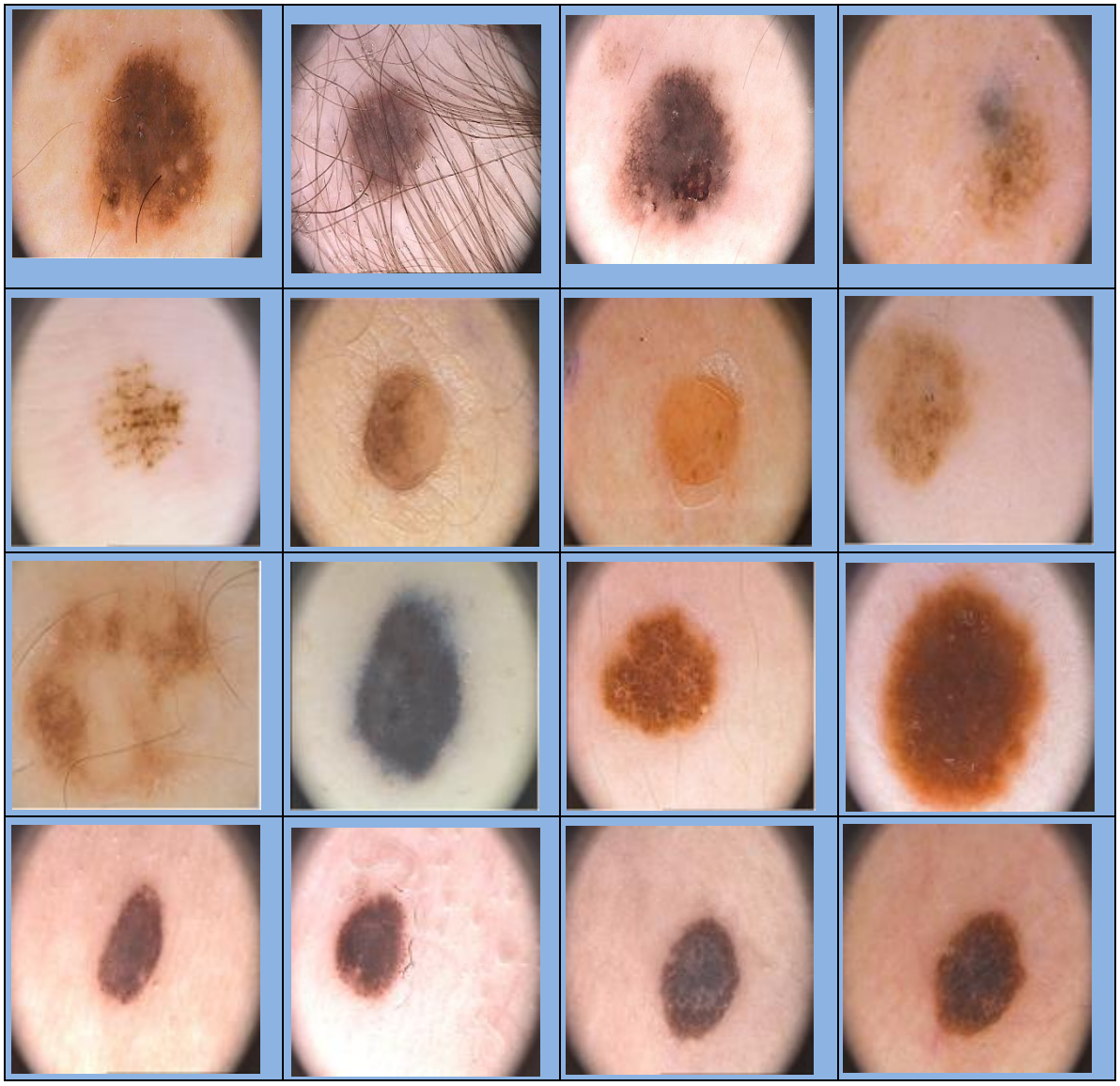


Figure (4.1) Some Skin Images (Benign Samples)

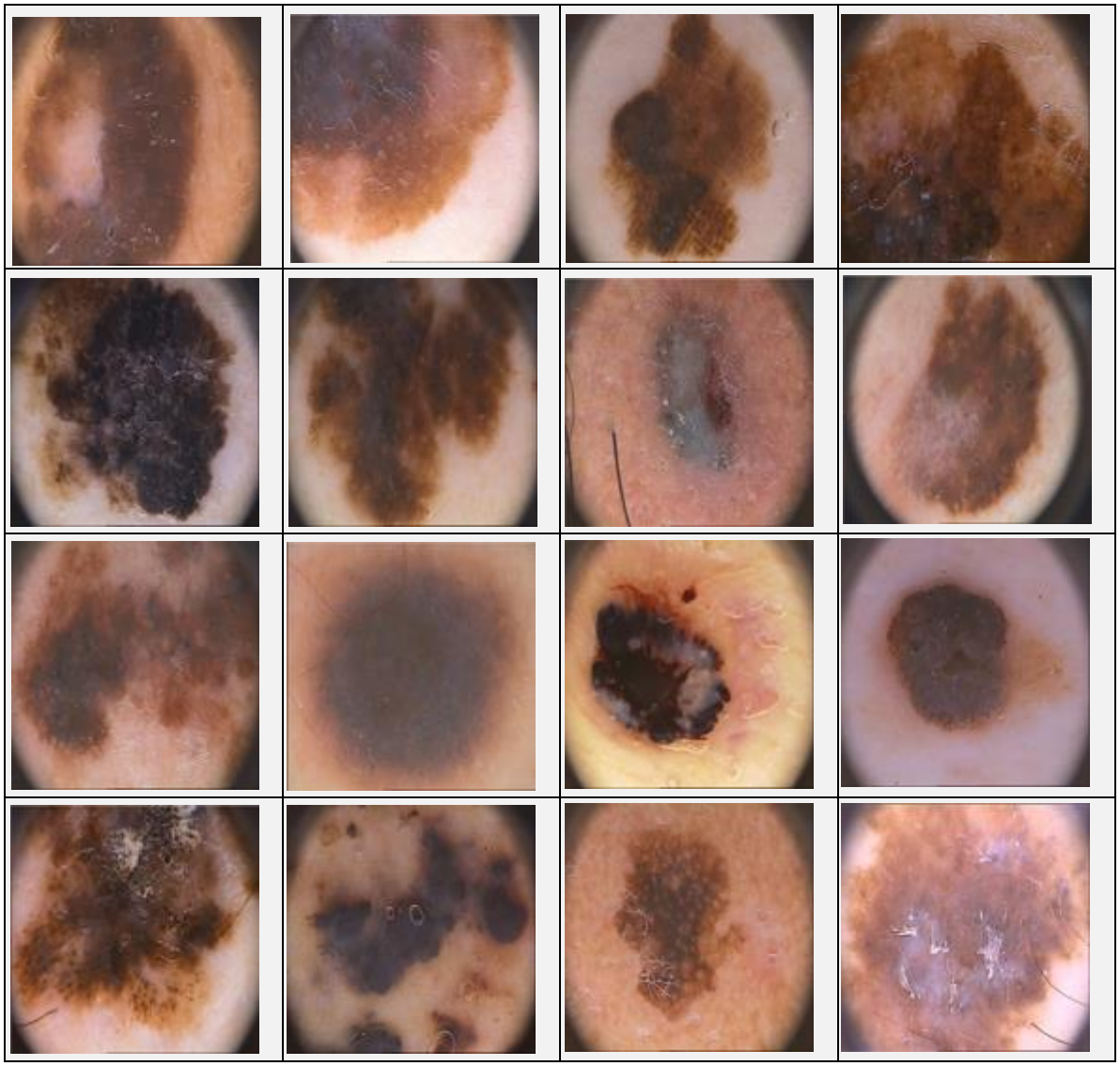


Figure (4.2) Some Skin Images (Melanoma Samples)

4.1 Results of Skin Image Preprocessing Stage

Image preprocessing is applied to both skin image sets, benign and melanoma. The results of this stage depend on section (3.1) and algorithm (3.1). One skin image example is shown in figure (4.3 a). The first step in the image preprocessing stage is done by separating the colors to RGB bands and then the median filter is applied on each band as shown in figure (4.3 b). A

morphological closing operation is applied to the image with line structure element: Horizontal line (zero angle), sloped line (45 angle), and vertical line (90 angle). From these three bands, maximum values are obtained and the hair area, which is close to dark in color, is replaced, as shown in figure (4.3 c). The last step in this stage is applying the Gaussian smoothing filter on the result as shown in figure (4.3 d).

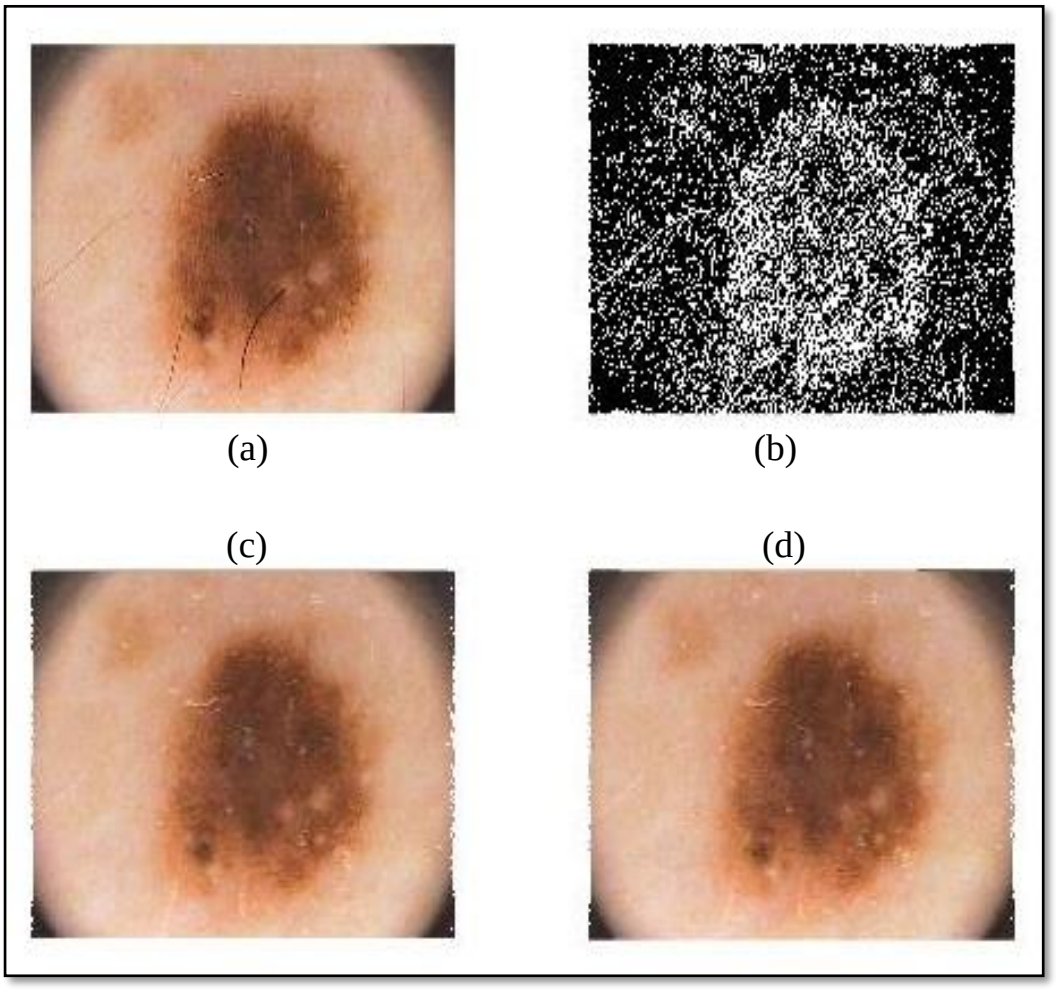


Figure (4.3) (a) Original Skin Image (b) Image after Splitting and Applying Median Filter (c) Image after Replacing Hair Area (d) Image after Applying Gaussian Filter

Image preprocessing of 16 skin images for benign and melanoma is shown in figures (4.4 and 4.5) respectively.

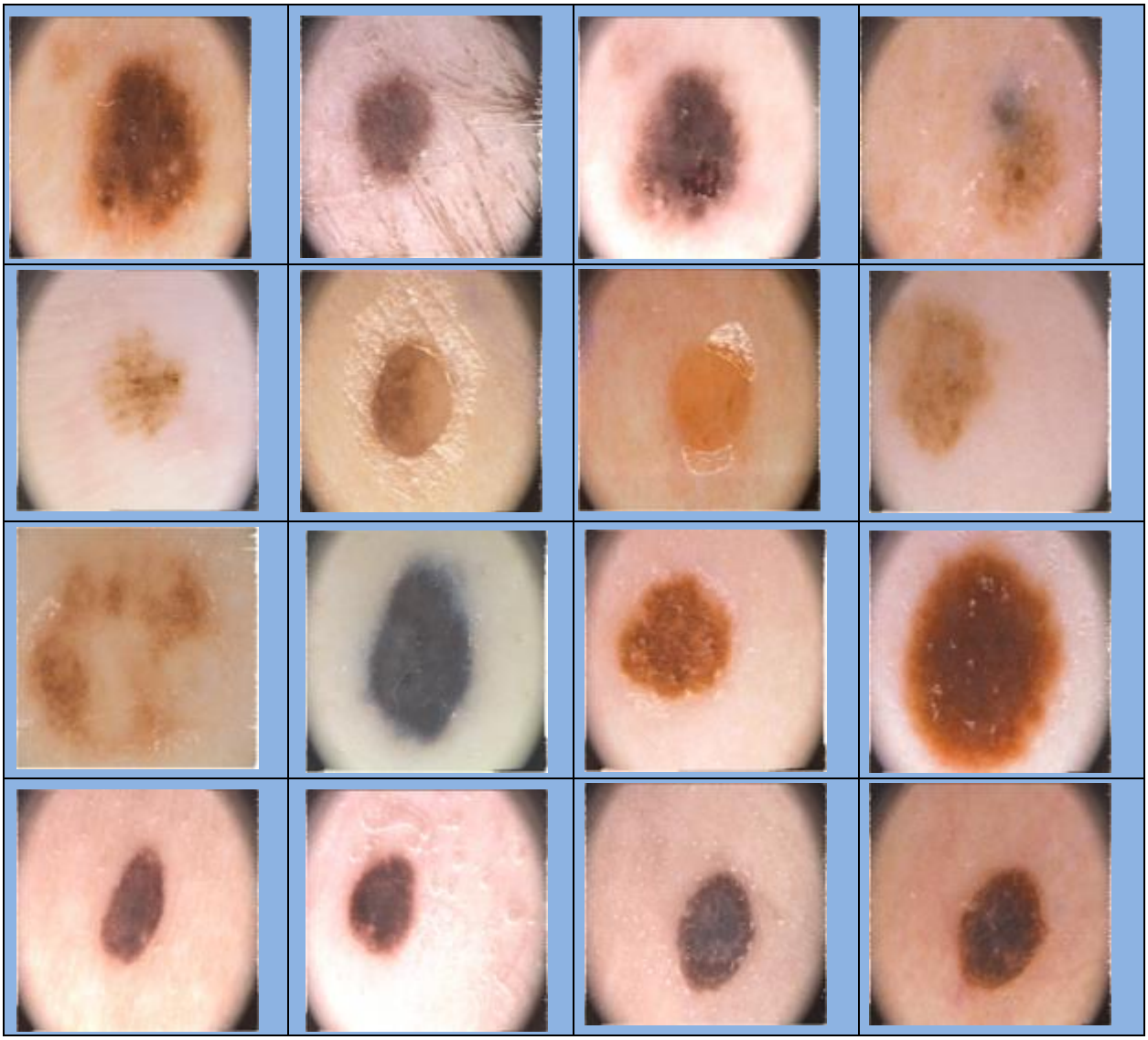


Figure (4.4) Skin Image Preprocessing (Benign Samples)

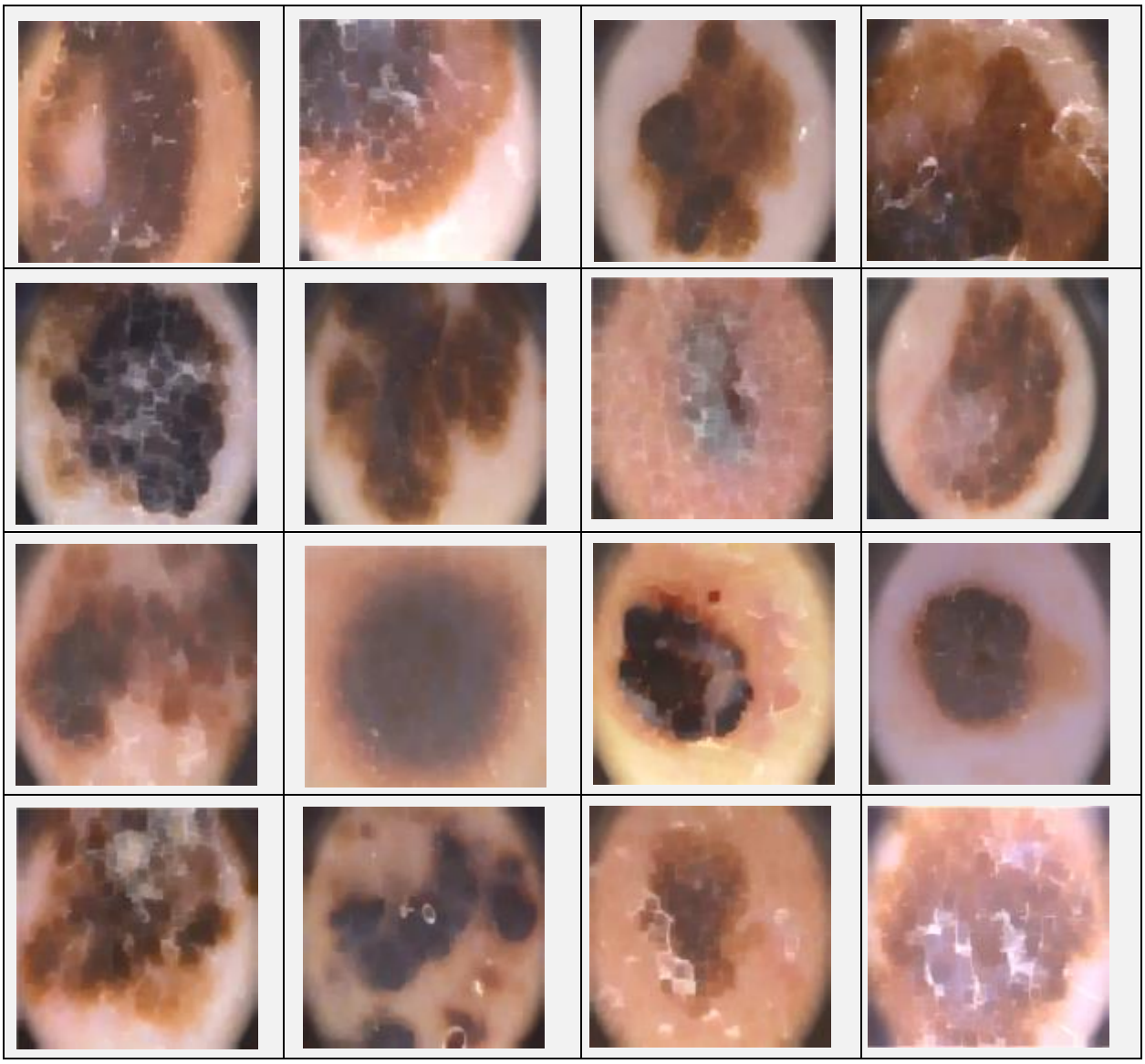


Figure (4.5) Skin Image Preprocessing (Melanoma Samples)

4.2 Results of Skin Image Segmentation Stage

PSO image segmentation process is detailed in section (3.2). This stage consists of a sequence of steps to segment the image in order to reach the region of interest (ROI). The resulted binary image will be inverted in order to make the white area relate to ROI. After this operation is

done, the enhancement of the segmentation is applied to the resulted image by comparing it with ground truth (GT) image of the same image that is obtained from the dataset as shown in figure (4.6 and 4.7).

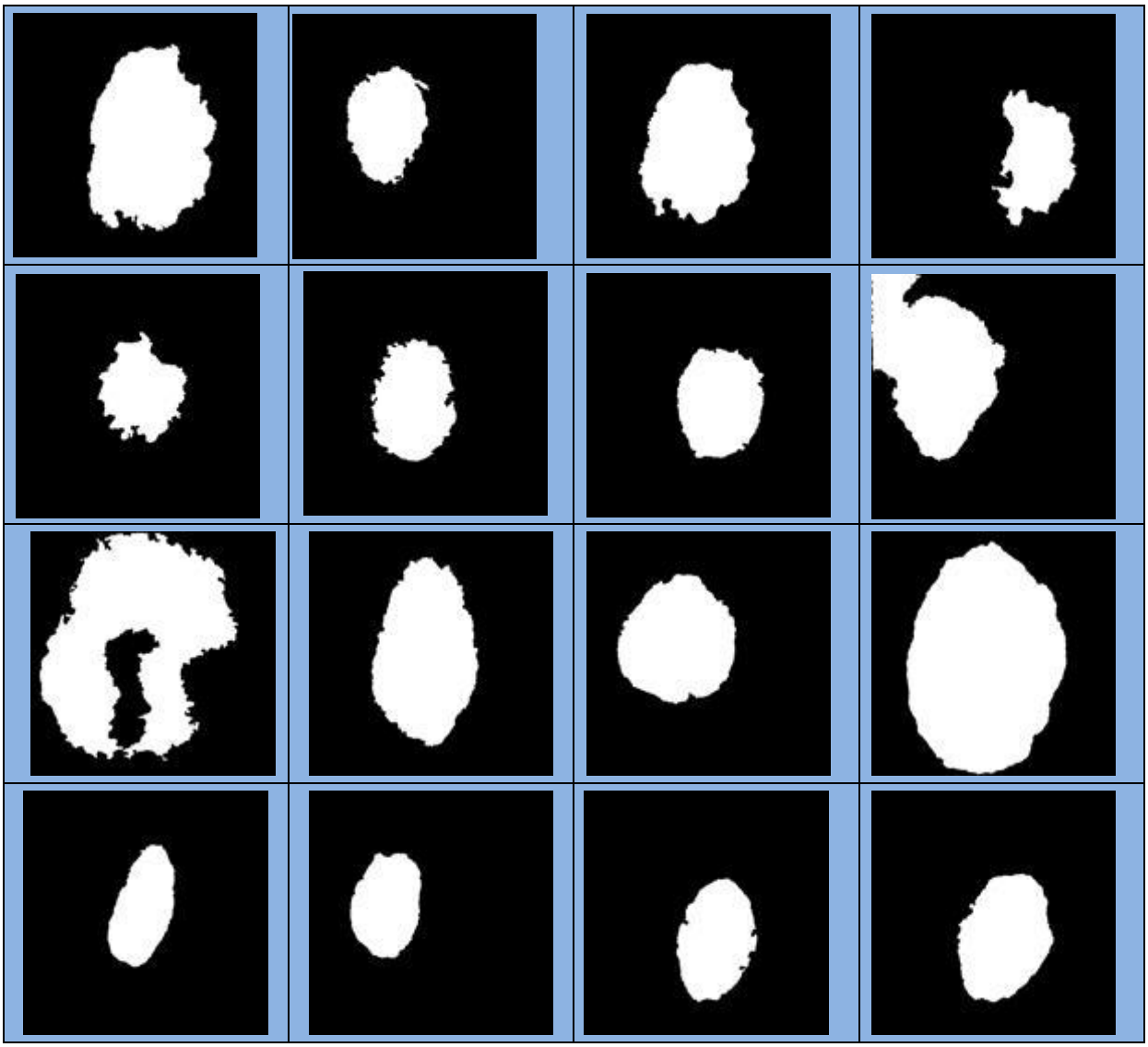


Figure (4.6) Image Segmentation using PSO Algorithm (Benign Samples)

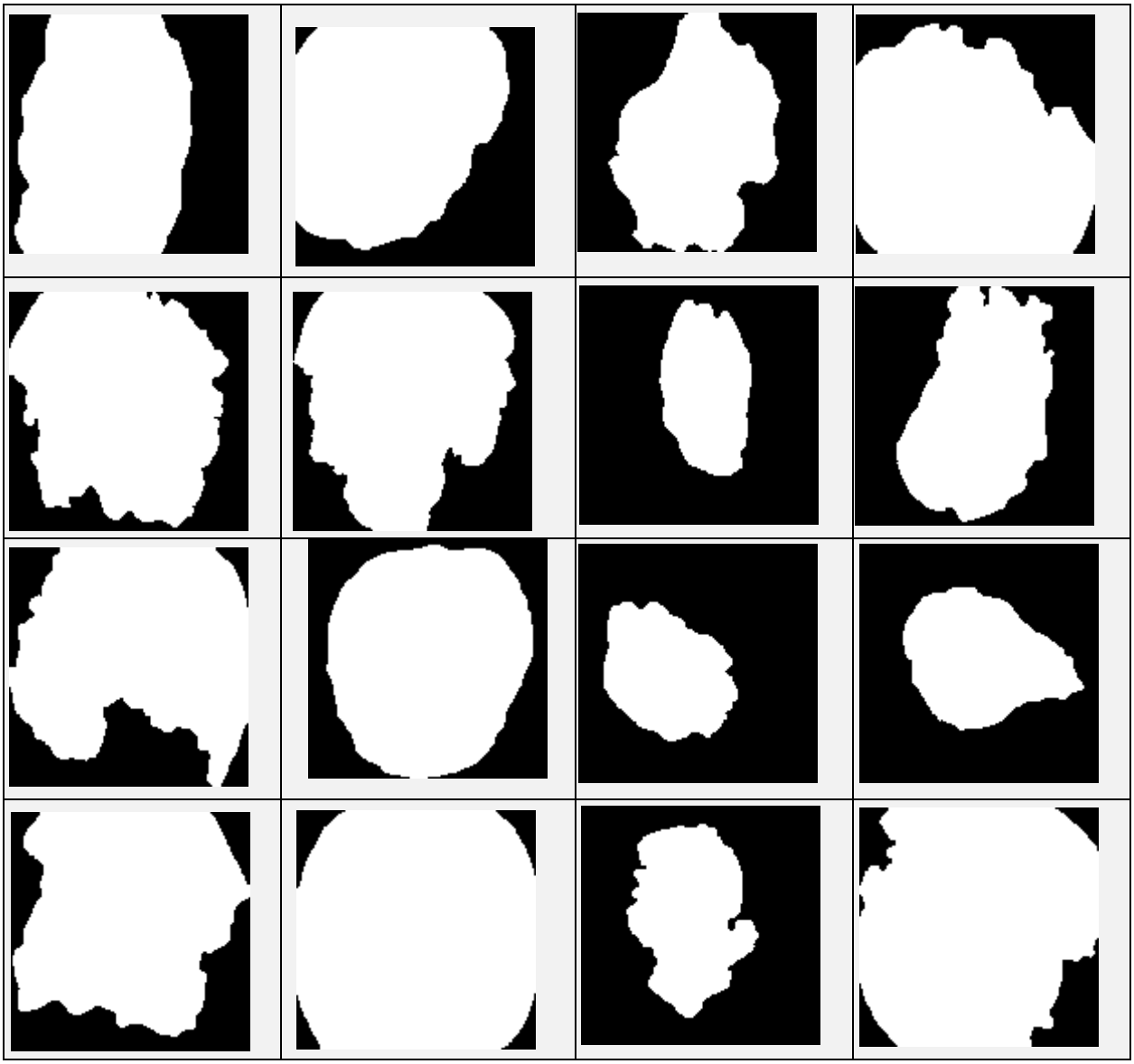


Figure (4.7) Image Segmentation using PSO Algorithm (Melanoma Samples)

4.3 ROI Detection

The executed segmentation technique was an assessment of a group of 200 images which are acquired from PH² dataset.

The detection of ROI in the proposed method is based on the PSO segmentation method. The region of interest is used later for the localization feature extraction. There are four parameters that skin cancer clustering depends on to compute the differences between two images, the ground truth lesion GT (binary image) and segmentation method. These parameters are: Cross error, Hammoude Distance (HM) as shown in Eq. (3.9), True Detection Rate (TDR) as shown in Eq. (3.10), and False Position Rate (FPR) as shown in Eq. (3.11). The outcomes of these parameters are used to evaluate the PSO segmentation in the proposed method. Because the extracted features depend on the ROI, this step is very sensitive, especially the structural features. ROI detection, enhanced ROI, and cropped images for benign samples are shown in figures (4.8, 4.9, and 4.10). The figures (4.11, 4.12, and 4.13) show ROI detection, enhanced ROI, and cropped images for melanoma samples respectively.

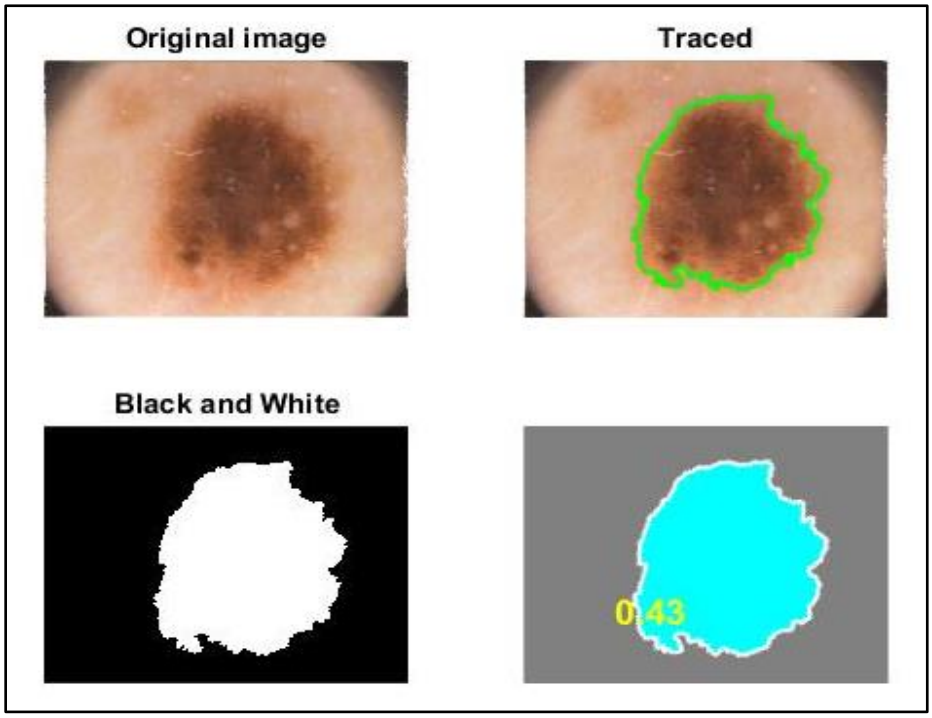


Figure (4.8) ROI Detection using PSO Algorithm (Benign Samples)

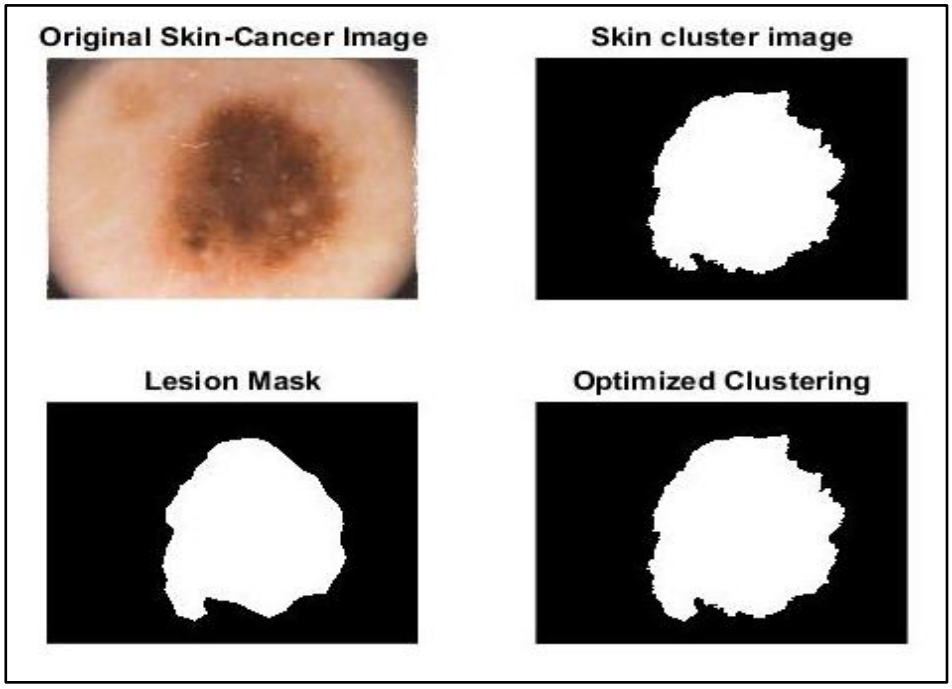


Figure (4.9) Enhancement of ROI (Benign Samples)

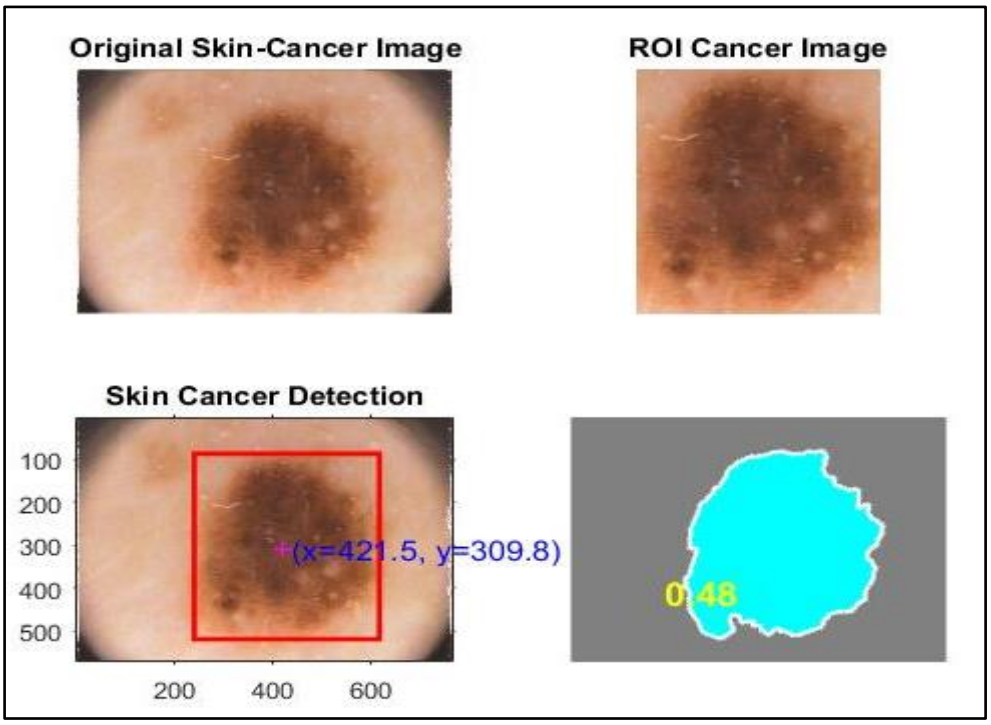


Figure (4.10) Image Crop of ROI (Benign Samples)

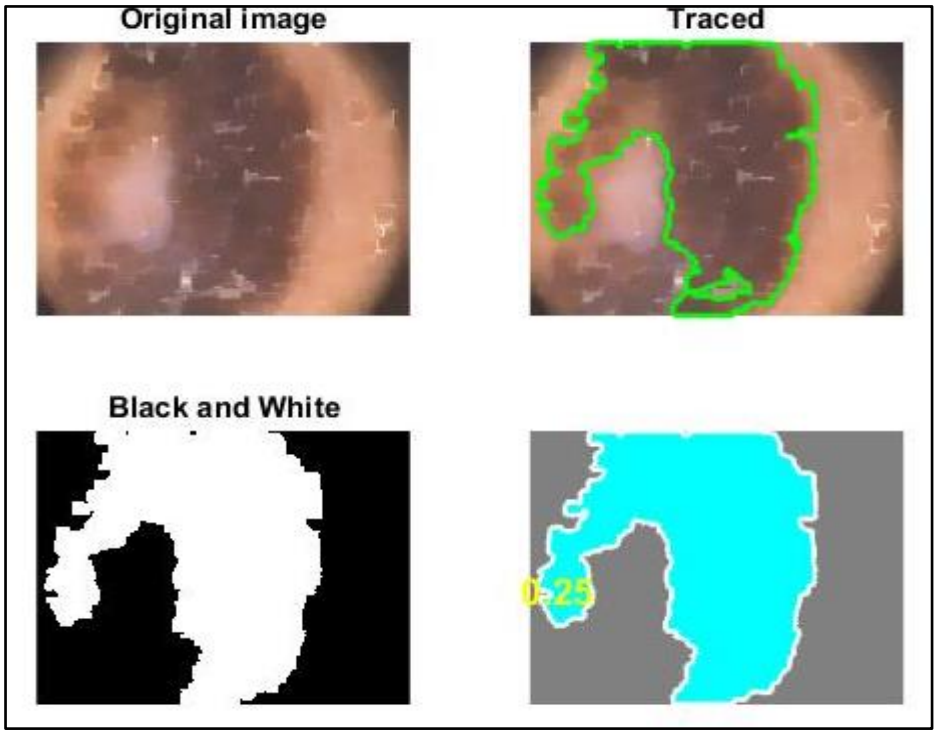


Figure (4.11) ROI Detection using PSO Algorithm (Melanoma Samples)

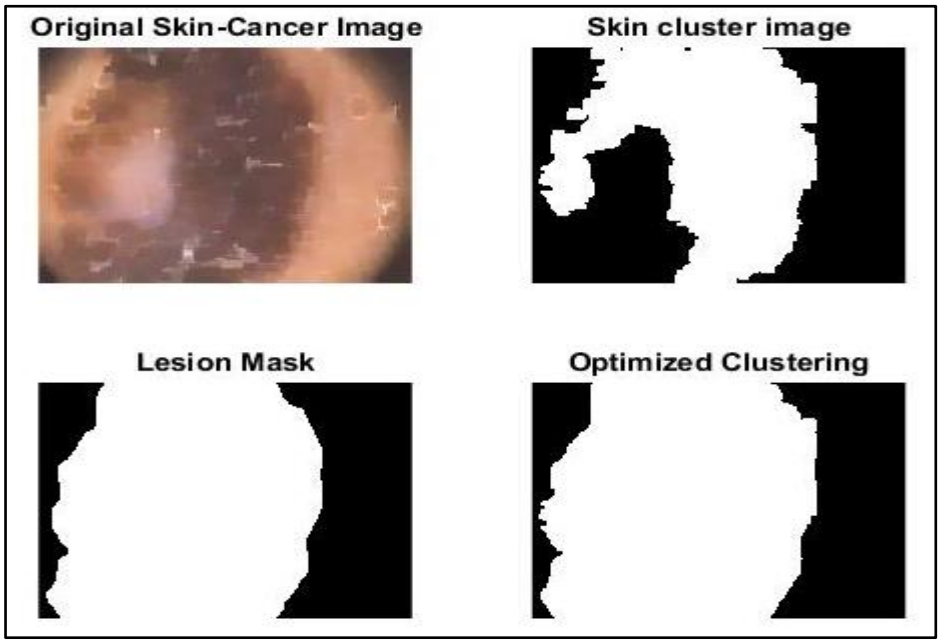


Figure (4.12) Enhancement of ROI (Melanoma Samples)

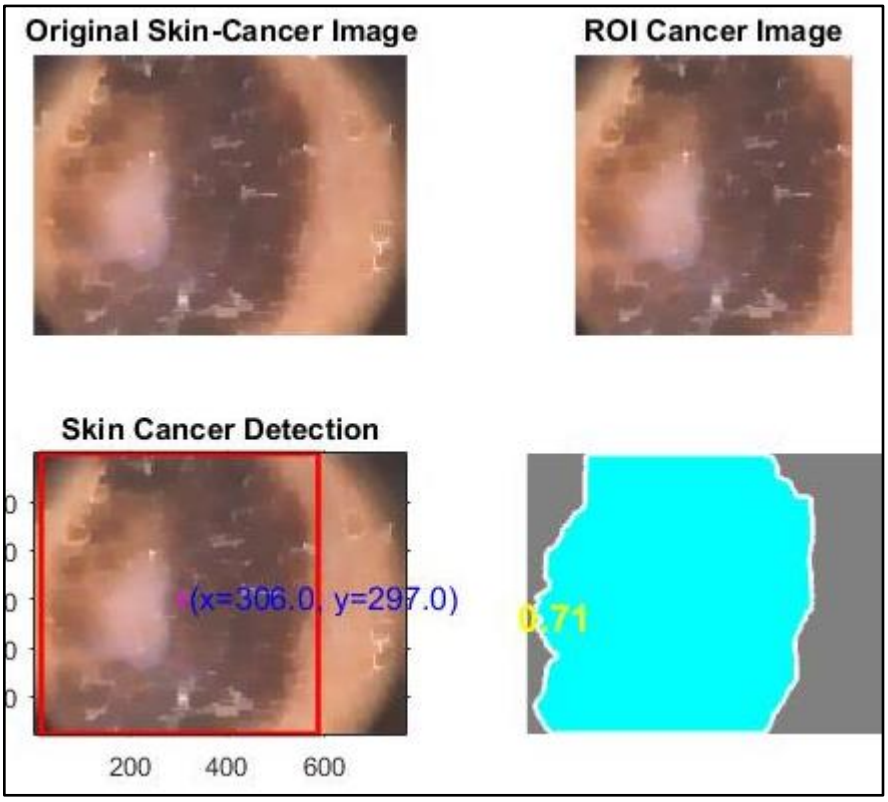


Figure (4.13) Image Crop of ROI (Melanoma Samples)

4.4 PSO Segmentation Evaluation

In the current system, the PSO segmentation method is adopted. In order to evaluate PSO segmentation, a comparison with ground truth lesion (GT) is made to determine our segmentation technique for both benign and melanoma images. The results of this comparison are determined by four parameters (HM, TDR, FPR, and GE), tables (4.1) and (4.2) clarified these results.

.

Table (4.1) PSO Segmentation Algorithm Evaluation for Benign Samples

Segmentation methods	Datasets (Benign skin cancer images) number of images (160)			
	HM (%)	TDR (%)	FPR (%)	GE (%)
PSO	4.257	97.000	5.225	3

Table (4.2) PSO Segmentation Algorithm Evaluation for Melanoma Samples

Segmentation methods	Datasets (Melanoma skin cancer images) number of images (40)			
	HM (%)	TDR (%)	FPR (%)	G E (%)
PSO	1.018	23.0764	0.883	0

4.5 Feature Extraction

In the current study, three kinds of features are extracted: geometric features, color features, and texture features. Total features that are used in the proposed method are 46 features which are clarified in chapter 3 as shown in tables (4.3, 4.4, 4.5, 4.6, 4.7, 4.8, and 4.9).

Table (4.3) Skin Image Features (feature 1 to feature 7)

Image #	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7
1	6023.88	9910.70	10509.46	123.00	0.00	44383.73	17741.26
2	2664.00	3869.02	5918.70	101.05	0.05	35524.84	6413.89
3	4602.38	7425.02	8849.46	127.58	0.00	49024.87	10901.28
4	2711.00	3899.51	6238.30	145.45	0.00	46009.26	12158.93
5	2418.88	3417.37	5058.11	106.89	0.00	51840.73	11002.93
6	2917.50	4354.08	6517.19	105.69	0.68	33887.45	11075.95
7	2644.88	4226.77	6088.19	130.52	0.70	32412.76	16343.04
8	5362.38	8952.58	9137.31	92.32	0.62	48171.21	13880.51
9	9539.38	14855.42	10258.83	144.59	0.00	84455.85	25479.46
10	5041.00	8094.03	10234.46	114.46	0.00	40021.71	984.40
11	4006.88	6368.22	7149.13	98.98	0.00	46864.83	19755.79
12	9030.13	15509.30	12950.39	127.70	0.00	62874.22	21733.04
13	2091.63	3116.45	5471.60	106.25	0.00	30464.43	7267.50
14	2150.13	3230.11	5375.12	95.30	0.00	32840.16	6523.10
15	2648.38	4006.39	6252.54	113.99	0.00	31832.91	5816.39
16	3280.50	5096.97	6914.14	106.16	0.14	27981.90	9806.66

Table (4.4) Skin Image Features (feature 8 to feature 14)

Image #	Feature 8	Feature 9	Feature 10	Feature 11	Feature 12	Feature 13	Feature 14
1	168.11	25.89	107.11	23.71	76.11	19.07	1105.59
2	168.26	21.32	126.94	21.03	122.75	20.78	816.41
3	173.72	40.59	125.07	34.27	116.68	31.54	1241.81
4	193.73	13.44	139.14	12.50	116.73	15.56	320.42
5	213.63	7.24	167.67	12.58	140.59	21.39	172.62
6	186.45	19.53	131.74	22.76	96.76	22.37	288.71
7	204.78	1.28	124.14	5.28	71.23	9.08	52.73
8	187.99	7.52	135.65	11.00	106.46	17.30	177.41
9	190.09	5.14	139.65	8.91	103.37	11.33	125.38
10	119.97	26.43	115.38	24.91	114.69	21.93	720.43
11	193.07	18.61	118.83	21.64	84.14	23.61	538.63
12	170.35	31.11	110.81	34.32	87.97	36.49	1282.91
13	184.51	27.28	130.89	23.75	119.80	20.42	1314.61
14	178.48	33.00	131.56	31.31	125.89	30.00	1744.34
15	156.48	26.55	118.88	20.98	112.85	17.95	1253.31
16	138.14	26.28	88.34	17.48	71.20	13.36	1048.82

Table (4.5) Skin Image Features (feature 15 to feature 21)

Image#	Feature 15	Feature 16	Feature 17	Feature 18	Feature 19	Feature 20	Feature 21
1	549.04	815.81	289.44	666.65	309.45	20.31	0.07
2	216.99	752.45	244.18	732.67	267.11	24.80	0.11
3	612.40	933.43	453.77	846.35	416.55	25.47	0.09
4	242.48	228.52	125.33	411.97	251.67	28.92	0.06
5	118.55	360.26	223.38	845.19	498.71	37.84	0.06
6	252.48	419.69	309.04	465.80	326.70	25.94	0.10
7	45.96	158.13	137.45	309.26	249.74	25.63	0.05
8	126.28	262.33	143.62	530.48	324.31	26.98	0.06
9	82.64	280.53	186.25	375.77	247.37	27.98	0.05
10	525.79	626.84	432.84	463.36	292.27	18.02	0.07
11	179.25	797.84	358.59	1178.00	717.08	24.09	0.11
12	565.36	1366.88	598.58	1906.02	1209.11	22.13	0.07
13	592.78	1110.86	550.82	864.22	416.95	27.67	0.09
14	897.97	1493.51	852.23	1359.03	872.19	27.78	0.12
15	469.43	849.92	378.95	682.88	326.00	22.29	0.13
16	427.49	576.08	277.45	418.38	255.60	14.76	0.08

Table (4.6) Skin Image Features (feature 22 to feature 28)

Image #	Feature 22	Feature 23	Feature 24	Feature 25	Feature 26	Feature 27	Feature 28
1	0.976	0.976	61.259	4.967	0.068	0.218	1.735
2	0.953	0.953	50.753	5.742	0.111	0.243	1.794
3	0.981	0.981	147.532	6.674	0.089	0.169	2.018
4	0.946	0.946	10.730	-1.824	0.057	0.351	1.281
5	0.947	0.947	14.078	-1.675	0.060	0.336	1.344
6	0.946	0.946	36.306	4.718	0.095	0.269	1.623
7	0.780	0.780	3.296	0.738	0.053	0.786	0.581
8	0.925	0.925	7.966	0.781	0.057	0.457	1.148
9	0.936	0.936	6.060	-0.541	0.054	0.378	1.188
10	0.975	0.975	74.721	10.477	0.072	0.263	1.634
11	0.951	0.951	49.857	5.775	0.114	0.241	1.794
12	0.986	0.986	174.821	16.803	0.065	0.226	1.833
13	0.973	0.973	75.246	7.173	0.092	0.249	1.763
14	0.975	0.975	187.636	15.780	0.120	0.172	2.091
15	0.957	0.957	83.258	9.787	0.127	0.225	1.888
16	0.961	0.961	35.348	6.067	0.075	0.338	1.482

Table (4.7) Skin Image Features (feature 29 to feature 35)

Image #	Feature 29	Feature 30	Feature 31	Feature 32	Feature 33	Feature 34	Feature 35
1	0.966	0.966	0.311	20.191	8.697	54.775	1.689
2	0.945	0.945	0.409	24.678	9.727	68.907	1.714
3	0.956	0.955	0.231	25.368	9.631	68.144	1.955
4	0.972	0.972	0.446	28.779	10.662	90.790	1.242
5	0.970	0.970	0.468	37.665	12.215	121.297	1.303
6	0.953	0.953	0.365	25.797	10.020	75.142	1.553
7	0.974	0.974	0.885	25.494	10.106	91.918	0.542
8	0.972	0.972	0.638	26.826	10.322	86.324	1.109
9	0.973	0.973	0.496	27.840	10.506	89.123	1.151
10	0.964	0.964	0.428	17.927	8.152	48.822	1.584
11	0.943	0.943	0.406	23.977	9.589	66.540	1.714
12	0.968	0.968	0.389	22.013	8.915	59.923	1.787
13	0.954	0.954	0.414	27.551	10.206	78.970	1.700
14	0.940	0.940	0.328	27.646	10.079	74.789	2.009
15	0.937	0.937	0.408	22.200	9.135	59.686	1.798
16	0.962	0.962	0.526	14.665	7.440	39.876	1.430

Table (4.8) Skin Image Features (feature 36 to feature 41)

Image #	Feature 36	Feature 37	Feature 38	Feature 39	Feature 40	Feature 41
1	0.068	0.247	-0.809	0.952	0.992	0.999
2	0.112	0.348	-0.715	0.929	0.988	0.998
3	0.089	0.301	-0.794	0.964	0.990	0.999
4	0.057	0.218	-0.771	0.894	0.994	0.999
5	0.060	0.225	-0.778	0.905	0.993	0.999
6	0.096	0.314	-0.732	0.920	0.989	0.999
7	0.054	0.208	-0.517	0.577	0.994	0.999
8	0.057	0.217	-0.748	0.864	0.994	0.999
9	0.054	0.209	-0.765	0.878	0.994	0.999
10	0.072	0.258	-0.794	0.940	0.992	0.999
11	0.114	0.352	-0.711	0.928	0.987	0.998
12	0.065	0.240	-0.830	0.962	0.993	0.999
13	0.092	0.307	-0.759	0.940	0.990	0.999
14	0.120	0.367	-0.748	0.958	0.987	0.998
15	0.128	0.381	-0.700	0.932	0.986	0.998
16	0.075	0.267	-0.756	0.914	0.992	0.999

Table (4.9) Skin Image Features (feature 41 to feature 46)

Image #	Feature 41	Feature 42	Feature 43	Feature 44	Feature 45	Feature 46
1	0.081	0.347	-0.709	0.842	0.792	0.912
2	0.212	0.448	-0.615	0.821	0.788	0.956
3	0.189	0.401	-0.694	0.864	0.791	0.988
4	0.157	0.318	-0.671	0.874	0.794	0.909
5	0.160	0.325	-0.678	0.815	0.793	0.919
6	0.196	0.414	-0.632	0.820	0.779	0.929
7	0.154	0.308	-0.617	0.871	0.794	0.939
8	0.157	0.317	-0.648	0.764	0.794	0.910
9	0.154	0.309	-0.665	0.778	0.794	0.978
10	0.172	0.358	-0.694	0.841	0.772	0.989
11	0.214	0.452	-0.611	0.827	0.767	0.977
12	0.165	0.340	-0.730	0.862	0.793	0.988
13	0.192	0.407	-0.659	0.840	0.780	0.991
14	0.220	0.467	-0.648	0.858	0.777	0.992
15	0.228	0.481	-0.600	0.832	0.786	0.993
16	0.175	0.367	-0.656	0.814	0.792	0.990

4.6 Mutual Information Feature Selection

The mutual information is calculated by weighing each feature vector to the class label. These weights are sorted from high to low, and then higher weight features can be chosen to specify the class label. These operations aim to reduce the complexity of the proposed method. Table (4.10) clarifies feature selection using mutual information.

Table (4.10) Feature Selection using Mutual Information

Feature #	Weight	Feature #	Weight	Feature #	Weight
29	0.48956995	1	0.18624225	8	0.08687406
30	0.48767077	13	0.16393401	10	0.08549911
40	0.48721964	20	0.13526988	24	0.08014811
38	0.48611076	32	0.13229612	25	0.07441735
41	0.46481494	5	0.12131053	9	0.05939861
26	0.46100350	33	0.11594237	27	0.05569418
21	0.45215248	12	0.11400244	19	0.05330576
36	0.45215248	14	0.10763029	28	0.05150652
37	0.45215248	18	0.09989052	31	0.04964335
4	0.35115992	34	0.09907525	35	0.04740629
22	0.31744550	3	0.09489224	7	0.04204114
23	0.31744550	39	0.09017542	16	0.03112867
2	0.24625134	11	0.08954196	17	0.02167799
6	0.20624209	15	0.08884637		

4.7 SVM Training

A suitable number of training samples is needed in SVM classification. Therefore, another part of the dataset called “validation set” was used. After an evaluation of a validation set successfully accomplished, the training proceeds on to the training set. This procedure is called cross-validation which is used to test sets for the final evaluation. This procedure is also called k-fold cross validation, because the training set is split into k smaller sets. Then SVM model is

trained using (k-1) folds as demonstrated in table (4.11) for linear kernel, table (4.12) for polynomial kernel, and table (4.13) for Gaussian kernel.

Table (4.11) Training Linear Kernel

Fold #	Image # 178	Training Fold				Testing Fold
		Sensitivity	Specificity	Precision	Accuracy	Accuracy
fold 1	141:35	0.928	1.000	1.000	0.942	0.971
fold 2	140:36	1.000	1.000	1.000	1.000	1.000
fold 3	141:35	1.000	0.875	0.965	0.972	0.971
fold 4	141:35	0.964	1.000	1.000	0.972	0.971
fold 5	141:35	1.000	1.000	1.000	1.000	1.000
Time Cross Validation		14.78 Sec				

Table (4.12) Training Polynomial Kernel

Fold #	Image # 178	Training Fold				Testing Fold
		Sensitivity	Specificity	Precision	Accuracy	Accuracy
fold 1	143:35	1.000	0.875	0.964	0.971	0.971
fold 2	142:36	0.928	1.000	1.000	0.944	0.944
fold 3	142:36	1.000	0.875	0.965	0.972	0.972
fold 4	142:36	1.000	1.000	1.000	1.000	1.000
fold 5	143:35	1.000	0.857	0.965	0.971	0.971
Time Cross Validation		47.85 Sec				

Table (4.13) Training Gaussian kernel

Fold #	Image # 178	Training Fold				Testing Fold
		Sensitivity	Specificity	Precision	Accuracy	Accuracy
fold 1	143:35	1.000	0.875	0.964	0.971	0.971
fold 2	142:36	1.000	1.000	1.000	1.000	1.00
fold 3	142:36	0.964	0.750	0.931	0.916	0.916
fold 4	142:36	1.000	0.875	0.965	0.972	0.972
fold 5	143:35	1.000	1.000	1.000	1.000	1.00
Time Cross Validation		17.64 Sec				

4.8 SVM Testing

SVM model is tested using: sensitivity, specificity, precision, accuracy, and time for linear kernel, polynomial kernel, and also Gaussian kernel as demonstrated in table (4.14).

Table (4.14) SVM Testing

Globalization	Sensitivity	Specificity	Precision	Accuracy	Time (sec)
Linear kernel	1.000	1.000	1.000	1.000	0.0298
Polynomial Kernel	1.000	1.000	1.000	1.000	0.0301
Gaussian Kernel	1.000	1.000	1.000	1.000	0.0276

4.9 Comparison of the Proposed System with Some Related Work

The results of the proposed system are compared with the results of some related works which used the same dataset (PH^2). The results that are taken in consideration are sensitivity (SE) and specificity (SP). This comparison is shown in table (4.15).

Table (4.15) Comparison of the Proposed System with Some Related work

Ref.	Year	Image	Used Approach	Results
[49]	2011	55	Based on a bank of directional filters and explores color, directionality and topological properties.	SE=80.0% SP=67.5%
[50]	2012	163	Using texture and color features and combining them both.	SE=94.1% SP=77.4%
[51]	2013	176	Using manual and automatic segmentation using K-Nearest Neighbors (KNN).	SE=98% SP=86%
[52]	2013	28	Evaluates the potential of an alternative approach based on the Menzies method presence of one or more of six color classes. Several artificial neural networks are used as classifiers.	SE=93% SP=62%
[26]	2013	148	Developing a global and a local classification Computer-Aided Diagnosis CAD systems, classification is solely based on color features. Using the k-Nearest Neighbors (KNN), and AdaBoost classifier**.**	SE=100% SP=93%
[53]	2013	176	Presents an approach to diagnosing melanomas using Bag-of-features, a	SE = 93% SP = 88%.

			classification method based on a local description of the image in small patches. Using (KNN) and (SVM) classifier.	
[54]	2014	176	Used two approachs, global and local features, used bag of features, compare the role of color and texture features in lesion classification. Used AdaBoost classifier and SVM classifier, and KNN classifier	SE= 96% SP= 80%
Proposed System		200	Using PSO Segmentation and SVM classifier	SE= 100% SP=100%

Chapter 5 Discussion, Conclusions and Suggestions for Further Work

In this study, a computerized system to diagnose melanoma from skin lesion images was designed and implemented using the Hospital Pedro Hispano (PH^2) dataset. Two hundred dermoscopy images of (PH^2) were analyzed to detect melanoma lesions using software code written in MATLAB (2015). Section (5.1) gives discussion points about this work. Conclusions are inferred in section (5.2). Suggestions for further work are described in section (5.3).

5.1 Discussion

In the last decade, amazing progress has been made on the optimization techniques to find more precise and accurate methods of image analysis optimization. Swarm intelligence (SI) algorithms are one of these methods. SI algorithms are part of clever algorithms, which are required these days. PSO is a swarm intelligence algorithm that is used as a segmentation technique.

Figure (4.5) pg (64) showed the successful image preprocessing techniques which were considered in this study, median filter and morphological closing. The comparison results from this study, table 4.14 pg (81), have demonstrated that the diagnosis sensitivity (SE) of the proposed system is 100% and the specificity (SP) is also 100%. Table 4.13 pg (80) show that the precision and accuracy are also 100% for all SVM kernels used in this study. But the consumed time played

an important role in which Gaussian kernel has the best consumed time, 0.028 sec, and the linear kernel 0.03 sec, followed by polynomial kernel computation at 0.03 sec.

These tables (4.3 pg (72), 4.4 pg (73), 4.5 pg (74), 4.6 pg (75), 4.7 pg (76), 4.8 pg (77), and 4.9 pg (78)) clarified geometric, color and texture features by 46 feature's values from 16 sample images. However, not all of these features' values are effective in determining the diagnosis, therefore, table (4.10) pg (79) described selected features using mutual information. This table showed that there were only 27 efficient features: 29, 30, 40, 38, 41, 26, 21, 36, 37, 4, 22, 23, 2, 6, 8, 10, 24, 25, 9, 27, 19, 28, 31, 35, 7, 16, and 17 respectively.

5.2 Conclusions

The results proved that the proposed system can be effectively used to diagnose skin cancer more accurately from skin lesion images. This tool is more convenient for the health care areas where experts in the dermatology may not be available, and it can be used as an automated process for diagnosis of the skin cancer.

This study adopted the PSO segmentation method which was assessed by computing the four parameters. These parameters, Hammoude distance (HM) Eq. (3.9) pg (37), true detection rate (TDR) Eq. (3.10) pg (38), false position rate (FPR) Eq. (3.11) pg (38), and gross error Counts (GE) pg (38), were used to make a comparison between ground truth lesion (GT) images and PSO segmented images. Tables (4.1) pg (71) and (4.2) pg (71) have shown that PSO method is a strong approach to use for segmentation for both benign and melanoma image samples. Furthermore, the final results have proved that the proposed method for the segmentation was robust because the local extracted features from ROI were outstanding.

The results achieved in this study were possible by using PSO method for the segmentation and SVM for the classification, and different types of features being extracted from the images. Multi-types of features (geometric, color, and texture) make the function of the classifier more powerful and efficient to classify the skin lesion image data with high accuracy. Table (4.15) pg (80) demonstrated that the proposed system in this study had good diagnostic results in comparison with seven chosen related works by considering the parameters, sensitivity (SE) and specificity (SP).

5.3 Suggestions for Further Work

This study indicated good approaches to address the key components in a computer-based system for diagnosis of melanoma. This section pursues different possible directions for further advances of this study.

It might be useful to use a larger dataset including various images of the same lesion. These images can be obtained from different imaging instrumentations such as ultra sound, dermoscopy, etc., to consider different aspects of the lesion. This can show the diversity of information about the same tumor such as surface or depth of the lesion and other criteria. Thus, this information would be beneficial to predict the cancer more accurately.

Further research on the orientation of this work may be done by using different segmentation methods, such as ant and bee colonies, or other swarm methods instead of PSO. This would enhance the conventional methods and make them more suitable to the skin cancer images in diagnostic systems. Also, a hybrid of these methods can be used as a new segmentation technique.

Moreover, a new classification technique could be developed to use PSO, ant colonies, bee colonies or other swarm methods instead of SVM

References

[1] Poornima M S and Shailaja K, “Detection of Skin Cancer Using SVM”, International Research Journal of Engineering and Technology (IRJET), Volume 4, Issue 7, July 2017.

[2] Shivangi Jain, Vandana jagtap, and Nitin Pise, " Computer Aided Melanoma Skin Cancer Detection Using Image Processing", Elsevier B.V., This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/), Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015), Procedia Computer Science 48, PP 735 – 740, 2015.

[3] Centers for Disease Control and Prevention, "Skin Cancer Statistics", https://www.cdc.gov/cancer/skin/statistics/index.htm, Page Last Update June 7, 2017.

[4] G.Argenziano, H. P. Soyer, V. De Giorgio et al., "Interactive Atlas of Dermoscopy", (Book and CD-ROM), Edra Medical Publishing and New Media, Milano, Italy, 2000.

[5] Leslie K. Dennis, Marta J. Vanbeek, Laura E. Beane Freeman, Brian J. Smith, Deborah V. Dawson and Julie A. Coughlin, "Sunburns and Risk of Cutaneous Melanoma: Does Age Matter? A Comprehensive Meta-Analysis", Annals of Epidemiology, Volume 18, Issue 8, PP 614–627, August 2008.

[6] Sloot, S., Rashid, O., Sarnaik, A., Zager, J., "Developments in Intralesional Therapy for Metastatic Melanoma", Cancer Control, Vol. (23), No. (1), PP 12–20, 2016.

[7] Argenziano, G.; Soyer, H.P.; Chimenti, S.; Talamini, R.; Corona, R.; Sera, F.; Binder, M.; Cerroni, L.; De Rosa, G.; Ferrara, G.; " Dermoscopy of Pigmented Skin Lesions: Results of a Consensus Meeting via the Internet", Journal of the American Academy of Dermatology, Elsevier, Vol. 48, Issue 5, PP 679–693, May 2003.

[8] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, Yongjun Wang, " Artificial intelligence in healthcare: past, present and future ", Stroke and Vascular Neurology 2017;0:e000101. doi:10.1136/svn-2017-000101, Copyright by BMJ Publishing Group Ltd, 2017.

[9] Pooja Kamavisdar, Sonam Saluja, Sonu Agrawal, "A Survey on Image Classification Approaches and Techniques”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue1, January 2013.

[10] A. Bono, C. Bartoli, M. Baldi Bono A1, Bartoli C, Baldi M, Moglia D, Tomatis S, Tragni G, Cascinelli N, Santinami M. , "Micro-Melanoma Detection- A Clinical Study on 22 Cases of Melanoma with a Diameter Equal to or less than 3 mm”, Tumori, Vol. 90, No. 1, PP 128–131, 2004.

[11] Al. Abadi, N. K.; Dahir, N. S.; Alkareem, Z. A. , " Skin Texture Recognition using Neural Network", Proceedings of the International Arab Conference on Information Technology, Tunisia, PP 1-4, December 16-18, 2008.

[12] Jonathan Blackledge and Dymitiy A. Dubovitskiy, " Texture Classification using Fractal Geometry for the Diagnosis of Skin Cancers ", Proceedings of EG UK Theory and Practice of Computer Graphics, UK, PP 1-8 ,2009.

[13] G. Capdenhourat, A. Corez, A. Bazzano, R. Allonso, and P. Musé, "Toward a Combined Tool to Assist Dermatologists in Melanoma Detection from Dermoscopic Images of Pigmented Skin Lesions", Pattern Recognition Letters, Vol. 32, PP 2187-2196, 2011.

[14] A. Safi, M. Baust, O. Pauly, V. Castaneda, T. Lasser, D. Mateus, N. Navab, R.Hein, and M. Ziai, " Computer Aided Diagnosis of Pigmented Skin Dermoscopic Images", https://link.springer.com/conference/mcbr-cds, International Workshop on medical Content-based Retrieval for clinical Decision Support, Lecture Notes on Computer Science , Vol. 7075, PP 105–115, 2012.

[15] A. Huang, W., Y. Chang, M., Y. Liu, and G.-S. Chen, "Capillary Detection for Clinical Images of Basal Cell Carcinoma", Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Barcelona, Spain, May 2 – 5, 2012.

[16] K. Korotov and R. Garcia, " Computerized Analysis of Pigmented Skin Lesion: A Review ", Artificial Intelligence in Medicine, Elsevier, Vol. 56, Issue 2, PP 69–90, 2012.

[17] L. Ma and R. C. Stauton, "Analysis of the Contour Structural Irregularity of Skin Lesion using Wavelet Decomposition", Pattern Recognition, Elsevier, Vol. 46, Issue 1, PP 98–106, January 2013.

[18] J. L. G. Arroyo and B. G. Zapirain, "Detection of Pigment Network in Dermoscopy Images using Supervised Machine Learning and Structural Analysis", Computers in Biology and Medicine, Elsevier, Vol. 44, PP 144–157, January 2014.

[19] Uzmabano Ansari and Tanujasarode , " Skin Cancer Detection using SVM " , Proceedings of WRFER International Conference, Pune, India, 16th April 2017.

[20] X Josephine Meena1 and S Indumathi, "A Secure IOT Based Skin Cancer Detection Scheme Using Support Vector Machine and Particle Swarm Optimization Algorithm ", International Journal of Engineering Research and Science and technology, Vol. 6, No. 2, April 2017.

[21] Khalid Eltayef, Yongmin Li and Xiaohui Liu, " Lesion Segmentation in Dermoscopy Images Using Particle Swarm Optimization and Markov Random Field", 30th International Symposium on Computer-Based Medical Systems, 23 June 2017.

[22] Macmillan Cancer Support, “A Practical Guide to Understanding Cancer- Understanding Skin Cancer", www.macmillan.org.uk, Booklet 9th edition, registered charity in England and Wales (261017), February 2014.

[23] National Cancer Institute, " What You Need To Know About Melanoma and Other Skin Cancers**",** Booklet from U.S. Department of Health and Human Services – National Institutes of Health, http://www.cancer.gov/espanol, October 2010.

[24] Maryam Sadeghi, " Towards Prevention and Early Diagnosis of Skin cancer: Computer-Aided Analysis of Dermoscopy Images", Ph.D Thesis, School of Computing Science, Faculty of Applied Sciences, Simon Fraser University, summer 2012.

[25] Cigdem Demir and B¨ulent Yener, " Automated Cancer Diagnosis Based on Histopathological Images: A Systematic Survey", Rensselaer Polytechnic Institute, Technical Report, 2005.

[26] Margarida Ruela, Catarina Barata, Jorge S. Marques, “What Is the Role of Color Symmetry in the Detection of Melanomas?”, International symposium (ISVC) : Advances in Visual Computing , pp 1-10, 2013.

[27] Mariam A.Sheha, Mai S.Mabrouk, Amr Sharawy, " Automatic Detection of Melanoma Skin Cancer using Texture Analysis", International Journal of Computer Applications (0975 – 8887), Vol. 42 No.20, March 2012.

[28] Klaus D. Toennies, “Guide to Medical Image Analysis – Methods and Algorithms", Advances in Computer Vision and Pattern Recognition, Springer-Verlag London Limited, 2012.

[29] Azadeh Noori Hoshyar, Adel Al-Jumailya, Afsaneh Noori Hoshyar, " The Beneficial Techniques in Preprocessing Step of Skin Cancer Detection System Comparing" , International Conference on Robot PRIDE 2013-2014 – Medical and Rehabilitation Robotics and Instrumentation, Elsevier Procedia Computer Science 42, PP 25 – 31, 2014.

[30] Alina Sultana, Mihai Ciuc, Tiberiu Radulescu,Liu Wanyu, Diana Petrache, " Preliminary Work on Dermatoscopic Lesion Segmantation ", 20th European Signal Processing Conference (EUSIPCO 2012), Bucharest, Romania, August 27-31, 2012.

[31] Rafael C. Gonzalez and Richard E. Woods, “Digital Image Processing", Pearson Prentice Hall, Third Edition, 2008.

[32] Raman Maini and Himanshu Aggarwal, “A Comprehensive Review of Image Enhancement Techniques”, Journal of Computing, Vol. 2, Issue 3, March 2010.

[33] Geoff Dougherty, “Digital Image Processing for Medical Applications", Cambridge University Press, 2009.

[34] Razi J. Al-azawi, Abbas Abdulazez Abdulhameed, Hussein Majeed Ahmed, " A Robustness Segmentation Approach for Skin Cancer Image Detection Based on an Adaptive Automatic Thresholding Technique", American Journal of Intelligent Systems, Vol. 7 Issue 4, PP 107-112, 2017.

[35] Bonabeau, E., Dorigo, M and Theraulaz, G., "Swarm Intelligence: From Natural to Artificial System", Book from Oxford University Press, New York, USA, 1999.

[36] Shi Y. and Eberhart R.C., "A Modified Particle Swarm Optimizer", Proceedings of the IEEE International Conference on Evolutionary computation, PP 69–73, IEEE Press, USA, 1998.

[37] Ismail Khalil Ali, “Intelligent Cryptanalysis Tool Using Particle Swarm Optimization", Ph.D.Thesis, Univercity of Technology, Department of Computer Science, Baghdad, Iraq, 2009.

[38] Nedjah N .and Luiza de M.,"Swarm Intelligence Systems", Book from Springer Verlag, Berlin Heidelberg , 2006.

[39] Dalia Nabeel Abdul Wadood, “Detection and Localization of Non-Melanoma Skin Cancer Using Texture Analysis", Ms.c Thesis, Baghdad University / College of Science, 2014.

[40] Madhuri A. Joshi, “Digital Image Processing An algorithmic Approach”, Prentice – Hall of India private Limited, New Delhi, 2006.

[41] Azadeh Noori Hoshyar, "Automatic Skin Cancer Detection system", Ms.c Thesis, University of Technology, Sydney, November 2014.

[42] G. N. Srinivasan, and Shobha G., "Statistical Texture Analysis", Proceedings of World Academy of Science, Engineering and Technology, Vol. 36, December 2008.

[43] Paweł Filipczuk, Thomas Fevens, Adam Krzyzak, Andrzej Obuchowicz, " GLCM and Glrlm Based Texture Features for Computer Aided Breast Cancer Diagnosis", Journal of Medical Informatics & Technologies, Vol. 19, 2012.

[44] Zena M. Hira and Duncan F. Gillies, “A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data", Advances in Bioinformatics, Vol. 2015, Article ID 198363, 13 pages, 2015.

[45] Pablo G. Cavalcanti, Jacob Scharcanski , Gladimir V.G. Baranoski , " A two-stage approach for discriminating melanocytic skin lesions using standard cameras" , Expert Systems with Applications, Vol.40, Elsevier Ltd., PP 4054-4064, 2013.

[46] Boser, B.E., Guyon, I.M. and Vapnik, V.N., "A Training Algorithm for Optimal Margin Classifiers", Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92), Pittsburgh, PP 144-152, 27-29 July 1992.

[47] B. Yekkehkhany, A. Safari, S. Homayouni, M. Hasanlou, " A Comparison Study of Different Kernel Functions for SVM-based Classification of Multi-temporal Polarimetry SAR Data", The 1st ISPRS International Conference on Geospatial Information Research, Tehran, Iran,15–17 November 2014.

[48] Genton, Marc G., "Classes of Kernels for Machine Learning: A Statistics Perspective." Journal of Machine Learning Research, Vol. 2, pp. 299-312, 2001.

[49] Catarina Barata, Jorge S. Marques, Jorge Rozeira, " Detecting the Pigment Network in Dermoscopy Images: a Directional Approach", Proceedings of IEEE Conference Eng. Med. Boil. Soc., 5120-3, 2011.

[50] Jorge S. Marques, Catarina Barata, Teresa Mendonc¸"On the Role of Texture and Color in the Classification of Dermoscopy Images", 34th Annual International Conference of the IEEE EMBS, San Diego, California USA, 28 August - 1 September, 2012.

[51] Catarina Barata, Jorge S. Marques, M. Emre Celebi, " Towards an Automatic Bag-of-Features Model for the Classification of Dermoscopy Images: The Influence of Segmentation", 8th International Symposium on Image and Signal Processing and Analysis, September 4-6, Trieste, Italy, 2013.

[52] Ca´tia S.P. Silvaa, Andre´ R.S. Marcalc, " Colour-based dermoscopy classification of cutaneous lesions: an alternative approach", Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, Vol. 1, No. 4, pp. 211–224, 2013.

[53] Catarina Barata, Margarida Ruela, Teresa Mendonça, Jorge S. Marques, " A Bag-of-Features Approach for the Classification of Melanomas in Dermoscopy Images: The Role of Color and Texture Descriptors", Computer Vision Techniques for the diagnosis of Skin Cancer, Chapter in Springer https://link.springer.com/chapter/10.1007/978-3-642-39608-3_3, PP. 49-69, First Online: 27 September 2013.

[54] Catarina Barata, Margarida Ruela, Mariana Francisco, Teresa Mendonça, and Jorge S. Marques, "Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features", IEEE Systems Journal, VOL. 8, NO. 3, September 2014.