

نموذج رقم (1)

إقرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

Melanoma Detection using mobile technologies
and data mining techniques.

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه
حيثما ورد، وإن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل لنيل درجة أو لقب علمي أو
بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

DECLARATION

The work provided in this thesis, unless otherwise referenced, is the
researcher's own work, and has not been submitted elsewhere for any
other degree or qualification

اسم الطالب: Ibrahim Alhalabi

Signature:

Date: 1/4/2015

التوقيع:

التاريخ: 1/4/2015

Islamic University – Gaza
Deanship of Graduate Studies
Faculty of Information Technology



***Melanoma Detection Using Mobile
Technology and Feature-Based
Classification Techniques***

A Thesis Submitted in Partial Fulfillment of the Requirement
for the Degree of Master in Information Technology

Prepared by

Ibrahim Salim AL Halaby
120100907

Supervised by

Dr. Ashraf Alattar

2014/1436



نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحث/ ابراهيم سليم ابراهيم الحلبي لنيل درجة الماجستير في كلية تكنولوجيا المعلومات برنامج تكنولوجيا المعلومات وموضوعها:

تشخيص مرض الميلانوما باستخدام تقنيات تنقيب البيانات والحوسبة النقالة Melanoma Detection Using Mobile Technologies and Data Mining Techniques

وبعد المناقشة التي تمت اليوم الثلاثاء 01 ربيع الأول 1436هـ، الموافق 2014/12/23م الساعة العاشرة والنصف صباحاً، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

.....	مشرفاً ورئيساً	د. أشرف محمد العطار
.....	مناقشاً داخلياً	د. علاء مصطفى الهليس
.....	مناقشاً خارجياً	د. إيهاب صلاح الدين زقوت

وبعد المداولة أوصت اللجنة بمنح الباحث درجة الماجستير في كلية تكنولوجيا المعلومات / برنامج تكنولوجيا المعلومات.

واللجنة إذ تمنحه هذه الدرجة فإنها توصيه بتقوى الله ولزوم طاعته وأن يسخر علمه في خدمة دينه ووطنه.

والله ولي التوفيق،،،

مساعد نائب الرئيس للبحث العلمي والدراسات العليا

.....
.....
.....
أ.د. فؤاد علي العاجز



Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Ashraf Alattar, his expertise, understanding, and patience added a lot to my graduate experience. I appreciate his vast knowledge and skills in many areas. I would like to thank him very much for his support and understanding.

Also, I would like to thank my parents, for supporting me spiritually and financially throughout my study.

Finally, to the most important and unending source of support; my wife and kids. Without their love and understanding I would not have completed this thesis

Abstract

Melanoma is one of the most dangerous types of skin cancer in terms of the ratio of death cases. Probability of death increases when it is diagnosed late. However, it is possible to treat melanoma successfully when diagnosed in its early stages. One of the most common medical methods for diagnosing melanoma is the ABCD (Asymmetry, Border irregularity, Color, and Diameter) method that involves the measurement of four features of skin lesions. The main disadvantage of this method is that estimation error and subjectivity affects the accuracy of diagnosis, especially when performed by non-specialists. Scarcity of specialists makes the problem worse. This has led to the development of computer systems to help in melanoma diagnosis. However, while most computer systems can achieve high accuracy with adequate speed, they have problems in the usability and flexibility. The emergence of smart phones with increasing image capture and processing capabilities has made it more possible to use such devices to perform medical image analysis such as the diagnosis of melanoma. Our research work combines existing melanoma diagnosis method and the image capture and processing capabilities of smart phones to achieve fast, affordable, easily available and highly accurate melanoma diagnosis. In this work, we propose a complete smart phone application to capture, and process an image of the suspicious region of the skin in order to estimate its probability of being melanoma. The system can use historical cases to improve its diagnosis accuracy. The system was tested on 164 sample images. 14 images were not well-captured and could not be diagnosed, while the remaining 150 cases were successfully processed. In

each of these 150 images, the lesion was correctly segmented and their ABCD feature set extracted. Diagnosis accuracy of the analyzed images ranged between 88%-94 with best results using SVM classifier, and worst is the KNN classifier.

Keywords

Medical imaging, melanoma diagnosis, skin lesion analysis, automatic classification, smartphone, Android

ملخص

الميلانوما أحد أخطر أنواع سرطان الجلد من حيث نسبة عدد الوفيات الى حالات الإصابة ،تزداد الخطورة في الحالات التي يتم معالجتها في مراحل متأخرة ، ولكن يمكن علاج الميلانوما بنجاح اذا تم اكتشاف المرض في مراحله الاولى ، لذلك هناك العديد من الطرق المتبعة لتشخيص المرض في المراحل الأولى لتوجيه المريض الى الطبيب المختص مباشرة في حالة الشك في وجود المرض ، اشهر هذه الطرق هي طريقة ABCD، وأهم العقبات التي تواجه هذه الطريقة هي عدم دقة التنفيذ من الاشخاص غير المتخصصين ، لأن اعتماد الطريقة على خصائص مثل الحجم واللون والشكل يجعلها عرضة للكثير من التقدير والنسبية في التشخيص مما يفقد هذه الطريقة الكثير من الدقة في النتائج ، لذلك تم خلال الاعوام السابقة العمل على العديد من الانظمة المحوسبة التي تسعى الى المساعدة في التشخيص وتقليل نسبة الخطأ، وأتمتة التشخيص بحيث لا يخضع للتقدير الشخصي ، ومع بداية ظهور الاجهزة الذكية والتي تجاوز استخدامها حدود التواصل ليتم استخدامها في التقاط الصور بدقة عالية ومعالجة البيانات بكفاءة والتواصل مع شبكة الانترنت، كان اتجاه للاستفادة من مرونة الأجهزة الذكية ودقتها في التقاط الصور ومعالجتها من اجل توفير انظمة لتشخيص الأمراض المختلفة، وكان نصيب من هذه الابحاث والتطبيقات لتسهيل عملية تشخيص الميلانوما، ورغم توفر العديد من الأنظمة والابحاث الا أن النتائج لا تزال في المراحل الاولى حيث هناك العديد من العقبات، فلا تزال دقة هذه الانظمة لا تصل الى المستوى المطلوب مما يجعلها في بعض الحالات خطر على حياة المريض اذا تم تشخيص الحالات المصابة بشكل سلبي، كما ان امكانات معالجة الصور وتخزين البيانات وتصنيفها يعتبر من المجالات الجديدة والتي لم يتم اختبارها بشكل كافي في الابحاث السابقة ولم يتم التعامل مع الامكانات الجديدة للهواتف الذكية ، وخلال فترة قريبة كان مثلا تحليل كميات كبيرة من البيانات على الهواتف الذكية ومعالجتها وتصنيفها من الامور الغير ممكنة والتي اصبحت الان من الامور الممكنة في ظل تطور وحدات التخزين والمعالجة بشكل كبير للهواتف الذكية.

لذلك جاء هذا العمل لمواصلة الجهد المبذول في توفير حل لمشكلة تشخيص مرض الميلانوما بشكل دقيق وفعال ومرن باستخدام الهواتف الذكية ، حيث تم استخدام امكانات الهواتف الذكية في التقاط الصور ومعالجتها بالإضافة لإمكانية تخزين المعلومات واستخدامها في التصنيف وتوقع الحالات الجديدة المصابة بالمرض، وتم بناء واختبار نظام كامل لتحقيق ذلك ، وقد كانت نتائج العمل مُرضية جدا حيث تم فحص البرنامج على عينة مكونة من 164 صورة حيث نجح البرنامج في مرحلة معالجة الصورة في معالجة 150 صورة وعزل منطقة الآفة من أصل 164 صورة كما ذكرنا، علما ان نسبة النجاح في معالجة الصور تم تحسينها لتجاوز الأخطاء اثناء التقاط الصور من

خلال استخدام واجهة تفاعلية للمستخدم ، اما في مرحلة التصنيف علي 150 صورة الناتجة من معالجة الصور فكانت دقة نتائج التصنيف بين 88-94% حسب نظام التصنيف المتبع ، وهي نتائج جيدة و يمكن البناء عليها في استخدام الموبايل في التشخيص الأولى لمرضى الميلاнома.

CONTENTS

1. INTRODUCTION.....	1
1.1 Melanoma Skin Cancer.....	1
1.1.1 How skin lesion develops into Melanoma.....	1
1.1.2. Signs and approaches for detection.....	2
1.2. Research Problem.....	3
1.2.1. Shortcomings in ABCD self-examination.....	3
1.2.2. Research Objective.....	4
1.3. Research Format.....	5
2. RELATED WORK AND CURRENT TECHNOLOGY.....	6
2.1. Related research.....	6
2.2. Related applications.....	7
3. METHODOLOGY.....	9
3.1. Research task.....	9
3.2. Proposed system architecture.....	10
3.3. Image Processing Operations.....	11
3.2.1. Image capture / selection.....	11
3.2.2. Image preprocessing.....	12
3.2.3. Image segmentation.....	13
3.2.4. Maximum Contour Extraction.....	16
3.2.5. Isolating the Lesion Area.....	17
3.2.6. Finding the centroid and the symmetry axis.....	18
3.2.7. Real time interactive image capture algorithm.....	20
3.3. Feature Extraction.....	21
3.3.1. Asymmetry.....	21
3.3.2. Border.....	22

3.3.3. Color.....	23
3.3.4. Diameter.....	24
3.4. Classification.....	24
3.4.1. KNN classification algorithm	26
3.4.2. SVM classifier.....	27
3.4.3. Neural networks classifiers	28
3.4.4. Decision tree classifier	29
4. IMPLEMENTATION.....	30
4.1. Software	30
4.2. Hardware.....	31
4.3. Dataset.....	31
4.4. Experiments procedures.....	32
5. RESULTS AND EVALUATION.....	33
5.1. Image processing results	33
5.2. Feature Extraction	40
5.3. Classification.....	44
6. CONCLUSION AND FUTURE WORK	50
REFERENCES.....	53
APPENDICES.....	57
Appendix A: Data Analysis.....	57
Appendix B: Application screenshot.....	60

LIST OF TABLES

Table 5-1 KNN classification results 44

Table 5-2 KNN Confusion Matrix, split the dataset to 66.0% training and the remainder testing 45

Table 5-3 KNN Confusion Matrix with cross-validation method..... 46

Table 5-4 SVM classifier Confusion Matrix 47

Table 5-5 Neural Networks classifiers Confusion Matrix..... 47

Table 5-6 Decision Tree classifier Confusion Matrix 47

LIST OF FIGURES

Figure 1-1: Skin with abnormal Melanoma area [3] 2

Figure 3-1: Basic melanoma detection system architecture 10

Figure 3-2: Doctor Mole application 12

Figure 3-3: Flood-fill algorithm From the left, 1: select a seed point, 2: move outwards to neighbors, 3: stop when the region is filled 18

Figure 3-4: The principle of the nearest neighbor (NN) classifier. 26

Figure 3-5: SVM Linearly separable data 28

Figure 3-6: SVM with many possible separating hyperplanes..... 28

Figure 5-1: Results of using noise removal filters..... 34

Figure 5-2: Otsu image segmentation method..... 35

Figure 5-3: Mean segmentation method “A” The image before any processing, “B” Image with Mean Shift Segmentation and Median Blur filtering, “C” Otsu Segmentation 36

Figure 5-4: Selecting the best lesion contour 37

Figure 5-5: Lesion centroid and semi axis 38

Figure 5-6: Using Flood fill algorithm to isolate the lesion A the input image, B the contour of the lesion, C flood-fill the contour, D the input image masked with the flood-fill image..... 38

Figure 5-7: Melanoma lesion image capture interface 39

Figure 5-8: Lesion X,Y Asymmetry samples 40

Figure 5-9: Lesion Border property..... 42

Figure 5-10: Color Mean and standard deviation analysis 43

Figure 5-11: Lesion size properties 43

Figure 5-12: Decision Tree visualization 48

Figure 5-13 Comparison between the four classifiers 49

Figure 7-1 Features calculated from the Lesion ROIs..... 57

Figure 7-2 Example of classification results using ANN classifier..... 59

Figure 7-3: The application screenshot 60

Abbreviations

ABCD	Asymmetry, Border irregularity, Color, and Diameter
CAD	Computer-Aided Diagnostic
ROI	Region Of Interest
AR	Augmented Reality
CI	Compactness Index
HCR	Hull/Contour Ratio
RGB	Red, Green, Blue
HIS	Hue, Intensity, Saturation
HSV	Hue, Saturation, Value
KDD	Knowledge Discovery in Databases
KNN	k-Nearest Neighbors
SVM	Support Vector Machine
ANN	Artificial Neural Network
DT	Decision Tree
MMH	Maximum Marginal Hyperplane
ML	Machine Learning
SDK	Software Development Kit
TP	True Positives
FP	False Positives
TN	True Negatives

1. Introduction

The first part of this chapter introduces Melanoma; how melanoma develops and spreads, and how it is typically detected. In the second part, the research problem is presented and specific research objectives are defined. The final section of this introduction outlines the format of this thesis.

1.1.Melanoma Skin Cancer

Melanoma is the most serious type of skin cancer, it accounts for less than 2% of skin cancer cases but causes a large majority of deaths, more than 75 percent of all skin cancer deaths [1], American Cancer Society's estimates for melanoma in the United States for 2014 shows that about 76,100 new melanomas will be diagnosed (about 43,890 in men and 32,210 in women), and 9,710 are expected to die of melanoma [2], this shows the importance of finding new methods and technology to diagnose the melanoma.

1.1.1. How skin lesion develops into Melanoma

The skin cancer "Melanoma" begins in skin cells called melanocytes. Melanocytes are the cells that make melanin, which gives skin its color. Melanin also protects the deeper layers of the skin from the sun's harmful ultraviolet (UV) rays. When people spend time in the sunlight, the melanocytes make more melanin and cause the skin to tan, this also happens when skin is exposed to other forms of ultraviolet light (such as in a tanning booth). If the skin receives ultraviolet light, the melanocytes may begin to grow abnormally and become cancerous. This condition called Melanoma as shown in Figure 1-1, Melanoma cancer grows in the outer layer where melanocytes are located on the skin, then at later stage spread to other parts of the body such as bones and lungs. Therefore, melanoma is lethal if not detected in the early stage; this awareness has led to increase interesting of early-stage diagnosis solutions.

The issue of providing early-stage melanoma detection solutions become increasingly important, where the malignant melanoma is increasing rapidly worldwide, including in countries with historically low incidence rates, this increase is occurring at a faster rate than for any other neoplasm [1], Population-based strategies to control the disease have largely focused on primary prevention and early detection.

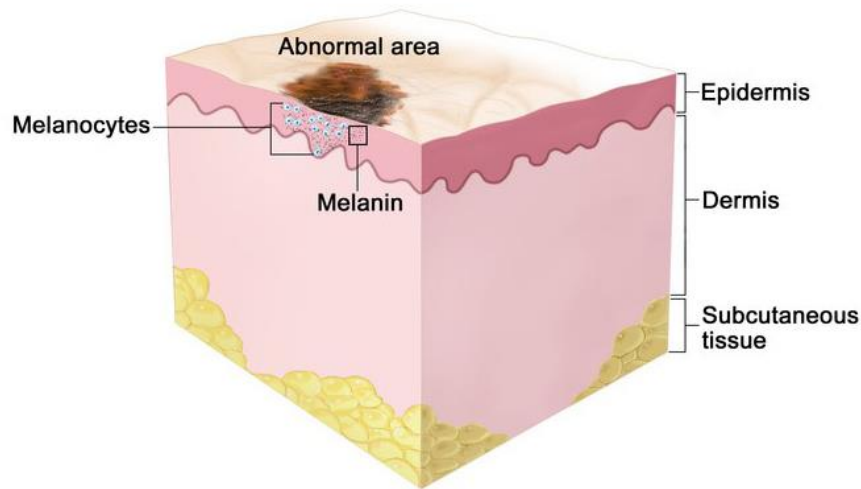


Figure 1-1: Skin with abnormal Melanoma area [3]

1.1.2. Signs and approaches for detection

This research focuses on early melanoma detection stage, this stage based on skin self-examination, the examination methods include many reliable and standardized techniques that improve the early detection rate, including the ABCDE (Asymmetry, Border Irregularity, Color, Diameter and Evolution) method [4], and the seven-point checklist [4], these techniques have proven it's useful in clinical situations, those two methods are the most commonly used, and there are some other methods such as Menzies [5] method and CASH [6], which is considered less common and important, The seven-point checklist method needs high quality images, captured with special medical cameras, also need specialist doctor decision, so it's not suitable to use with automated mobile application, the most common is ABCD method, which tracks the most important warning signs of

Melanoma, and helps in distinguish between a normal and abnormal lesions. ABCD method depends on some basic melanoma characteristics that can observed with naked eye or through images taken with the camera, the method name ABCDE stands for the following [2]

- A is for **Asymmetry**: One-half of a mole does not match the other.
- B is for **Border**: The edges are irregular, ragged, notched, or blurred.
- C is for **Color**: The color is not the same all over and may include shades of brown or black, or even patches of pink, red, white, or blue.
- D is for **Diameter**: The spot is larger than about $\frac{1}{4}$ inch, but melanomas can be smaller than this.
- E is for **Evolving**: The mole is changing in size, shape, or color.

Most references use the name ABCD, but in other references they added E feature, which represents evolution and lesion changes during long intervals, we will focus on the ABCD features, at least in this stage.

1.2. Research Problem

This section first discusses the shortcomings of the ABCD self-examination technique, then presents the objectives of our research work

1.2.1. Shortcomings in ABCD self-examination

While ABCD is the most widely used method for evaluation of skin lesions, the main problem when patients are not sure about the features, especially that the result depends entirely on the perspective and self-interpretation of the user. ABCD features are susceptible to subjectivity, this decrease the specificity and the accuracy in diagnostic results. Even with many brochures and websites that explain the signs and examination method, still not easily to decide if the mole is normal or not. Therefore, we need to develop support systems to help the user; these systems support the process of self-examination,

give accurate results about the seriousness of the situation, and determine the need to see doctor or trained health professional. Based on the above introduction, we can conclude the following shortcomings in melanoma detection:

1. Lack of immediate medical expertise
2. Subjectivity in the interpretation of Melanoma features
3. Cannot always rely on the existence of reliable communication networks
4. Poor accuracy of existing mobile based melanoma diagnosis applications

The problem, therefore, is how to overcome the above shortcoming and limitations using modern technology?

1.2.2. Research Objective

The discussion in the previous section shows the importance of this issue, and leads to the motivation and objective of the work, which is,

“Implementing and examining new system to enhance the melanoma self-examination, and automate the diagnostic process using mobile technology”

The system will be able to achieve the following goals:

1. Melanoma detection mobile-based image-processing method.
2. Machine learning capabilities to help with melanoma detection.
3. Ability to improve and update the decision support by having a skin features dataset used in training phase.
4. Showing the melanoma risk factor, with the possibility of showing the similar cases to the patient, to help the patient as much as possible in the decision to consult a specialist doctor.

5. Improving the efficiency and accuracy of existing systems by working to improve the image processing and classification methods.

We will discuss some other suggested solutions that discuss melanoma diagnosis including image processing, features extraction and classification in mobile platform. The next chapter discusses some existing solutions and related work, and shows the weaknesses, strengths and shortcomings in the automate diagnosis solutions that use mobile technology.

1.3. Research Format

The remainder of this thesis is organized as follows: Chapter two provides an overview of related research work and existing diagnosis technologies. Chapter three presents and explains our proposed method and the various image processing operations and classification techniques it uses. Chapter four discusses the implementation details. Chapter five presents the results of the system, finally, chapter six concludes the thesis with some suggestions for future work.

2. Related Work and Current Technology

2.1. Related research

The references and previous research in mobile based melanoma diagnoses can be divided into three fields:

- Enhancing melanoma detection and classification techniques.
- Using mobile in medical systems and development of Computer-Aided Diagnostic (CAD) tools.
- Enhancing the image processing methods to extract melanoma features.

The research in the first field determines the best machine learning systems and data mining techniques, second field determines the methods to build telemedicine system and the possibilities to overcome mobile limitation in data processing and storage using methods and algorithms commensurate with the mobile specification. The third field is the most important in this research, aims to find the most accurate image processing techniques to capture the image and turn it into a set of features that can be used in diagnosis disease.

Abuzaghleh, et al. [7], presented smart-phone based system named SKIN cure to assist melanoma early detection. They proposed a framework to analyze and classify the images into normal and melanoma, and alert the user at real-time to seek medical help urgently. The framework compared the performance of two classifier techniques, one level classifier and two-level classifier, concluded that the two-level classifier outperforms the one level classifier. The paper needs more focus on defining and extracting features, for example, asymmetry and irregularity, to improve the accuracy of classification.

Wadhawan, et al, [8], introduced SkinScan portable library for melanoma detection on handheld devices, the library implemented with C / C++. They showed that the most computational intensive and time consuming algorithms of the library, namely image segmentation and image classification, can achieve accuracy and speed of execution

comparable to a desktop computer. These findings demonstrate that it is possible to run sophisticated biomedical imaging applications on smart phones and other handheld devices, which have the advantage of portability.

Ramlakhan et al. [9], presents a prototype of an image-based automated melanoma recognition system on Android smartphones, the system consists of three major components: image segmentation, feature extraction based on ABCD method, and classification. Experimental result showed that the system was not highly efficient, achieved an average accuracy of 66.7%, with average malignant class sensitivity of 60.7% and specificity of 80.5%. The paper presented two systems for the detection of melanoma cases in dermoscopy images using texture and color features.

Karargyris et al. [10], have worked on an advanced image-processing mobile application for monitoring skin cancer. The authors presented an application for skin prevention using an inexpensive apparatus (microscope) and a smartphone (iPhone). These two components standalone are sufficient to capture highly detailed images for use by experts with medical background. Additionally, an advanced software framework for image processing backs the system to analyze the input images. The main research goal was to demonstrate how smartphones could turn into powerful and intelligent machines and help large populations without expertise in low-resource settings. Their images database was small, and consisted of only 6 images of normal cases and 6 images of suspicious case.

2.2.Related applications

In the related technology and applications, several applications are available that offer skin self-examination and that aid the user in classifying skin lesions to benign or melanoma. Based on previous review to this application [11], the performance of smartphone applications in assessing melanoma risk is highly variable, and 3 of 4 smartphone applications incorrectly classified 30% or more of melanomas as benign.

The review show that this applications are not subject to any sort of validation or regulatory oversight. Despite disclaimers that these applications are intended for educational purposes, they have the potential to harm users who may believe mistakenly that the evaluation given by such an application is a substitute for medical advice. This risk is of particular concern for economically disadvantaged and uninsured patients. Because a substantial percentage of melanomas are detected initially by patients, the potential effect of such applications on melanoma detection patterns is particularly relevant.

We can conclude the following shortcomings in melanoma detection research and applications, There is a problem in the accuracy of the results, and we can't rely on existing solutions in the diagnosis of the melanoma, even some of the proposed solutions constitute a danger to the patients, so we need more research to improve the accuracy, other systems and solutions have a good results in certain dataset, but when using a mobile camera results we have inaccurate classification, this systems based on high-resolution images are not always available, other systems were good in the diagnosis of some of the melanoma features, but failed in the other, for example, the results were good in determining the color of the lesion, but completely failed to identify properties such as symmetry and irregularity.

Through a review of previous results, and after discussing several systems, we find that we need more work on improving the accuracy of the systems, through further research.

3. Methodology

The significance of early melanoma diagnosis motivated the development of computer-aided detection/classification (CAD) systems, the scientific community continues to work toward improving the diagnostic performance and clinical integration of mobile CAD technology, For this reason, we consider that reliable CAD systems for automated detection/classification of melanoma lesions, will be very useful and helpful to supply a valuable “second opinion” to melanoma self-examination.

We made a detailed review of techniques and systems of mobile melanoma detection related mobile CAD systems, this review included methods and techniques, these systems typically involve a hierarchical approach, first applying image preprocessing steps to enhance suspicious structures in the image and then employing feature extraction and finally the classification to better classify these features.

This work is focused to testing and examining new system to improve the following fields: image preprocessing and image enhancing, accurate melanoma segmentation, features vectors extraction and the classifiers accuracy to reduce classification errors. Our intention is to build a more robust mobile system and implemented it on mobile environment to expand the system possibilities in flexible image capture and storage, this system will facilitate the analysis of melanoma lesions from mobile images and we consider it the needed support to design, develop and evaluate more reliable and robust CAD systems.

3.1. Research task

The first research stage is determination of the related systems of the melanoma detection systems, theoretical background of image processing algorithms, image segmentation, and feature extraction and classification techniques, then design of melanoma detection architecture based on the previous study, finally we implementing and testing the system and validate of the results obtained by developed system. Based on related work in the previous chapter, we suggests a general system architecture, during the thesis we will use

this architecture and work to improve and examine the suitable algorithms to find the most accurate results.

3.2. Proposed system architecture

State of the art of development and evaluation of melanoma detection using mobile technology and feature-based classification techniques, based on digital image processing, features extraction and classification techniques. Some examples of developed methods with interesting results, in which is inspired this theses architecture are outlined in the related work. Figure 3-1 shows the proposed system architecture.

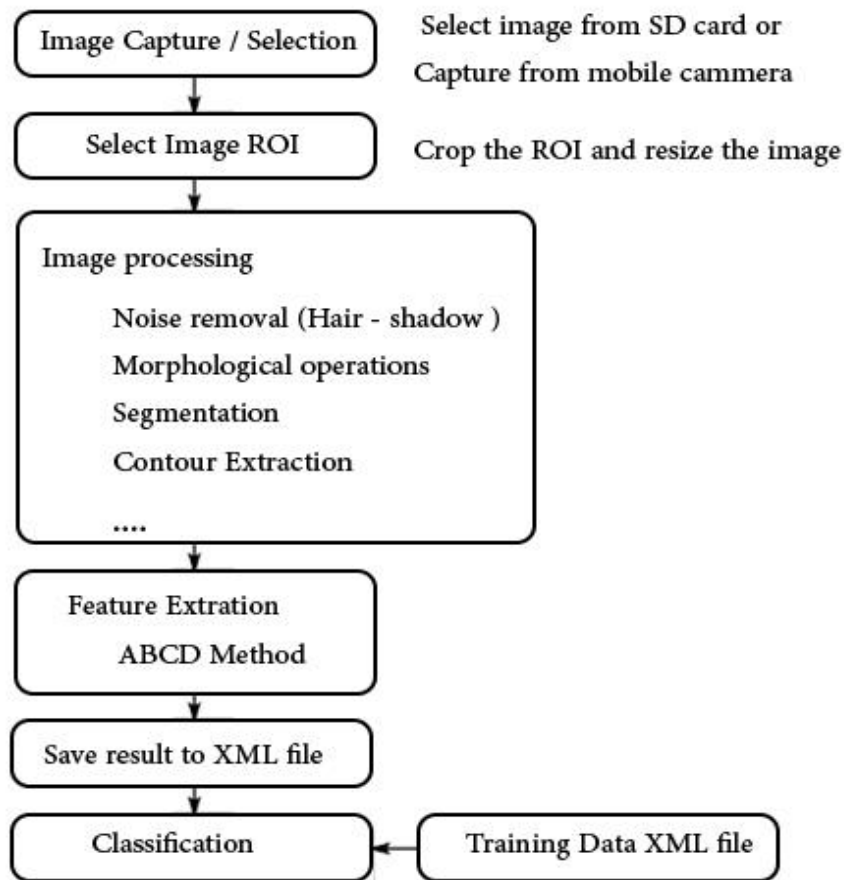


Figure 3-1: Basic melanoma detection system architecture

The theses objective is testing a set of melanoma detection algorithms on a mobile environment, melanoma detection algorithms are based on digital image processing, feature extraction and classification techniques, and we can reduce the insufficiency in melanoma self-examination on mobile CAD platforms if we can find the best appropriate algorithm, Therefore, the aim of the research is to work on each architecture part to improve accuracy. The proposed research applied using experimental methods and developed from quantitative point of view, the population of the proposed research is formed by 164 images from the public database in medical website [12], the same source used in related work, so we can compare and evaluate the results. The population represents a sample of possible normal or melanoma lesions that could be appearing in the algorithms evaluation process. Next we will give a brief description of the proposed method of this used in examined in the system architecture, including the main three parts: image processing, features extraction and the classification.

3.3. Image Processing Operations

The image processing includes many of the details, from image capture using a mobile smartphone, image preprocessing, then isolate the lesion area and find the contour, and finally the use of various methods of image processing to extract the features based on ABCD method.

3.3.1. Image capture / selection

We assume using two methods to capture and select the image. The first method, selecting the image directly from the SD Card or any storage device connected to the mobile, the second, capturing the image form the mobile camera.

There is no approved or clear method to identify the skin lesion areas using mobile application, but some applications, for example, "Doctor Mole" in Google play store shown in Figure 3-2 uses efficient virtual reality technique to capture the skin lesion in real time. So, we need to find an effective way to identify and isolate the lesion from the skin.

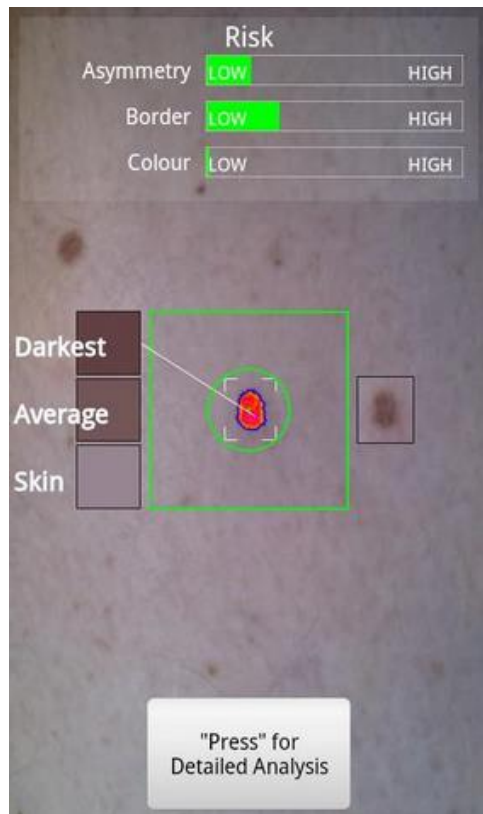


Figure 3-2: Doctor Mole application

3.3.2. Image preprocessing

The image cropped to square image with 300 pixel width to reduce the image processing time in the mobile later. Then we must reduce the noise, especially the unwanted small structures, such as fine hair, this can be done using simple noise removal filters like Gaussian filter [13], however, this method has the disadvantage of blurring the region boundaries. The best results have been obtained using Median filter [13], this kind of filters take every pixel and replace it with the median pixel value computed in a given neighborhood that depends on the image resolution. The principal drawback of these methods is that they are applied to all the pixels into the images, whether they are noisy or not. Morphological filter has also been evaluated, it strongly affects the lesion borders and

should be avoided and used after the lesion segmentation task. We need to review and compare the different approaches and may merge them to get proper results.

3.3.3. **Image segmentation**

The goal of this process is separating lesion from the surrounding skin. Image segmentation is one of the basic techniques of image processing and computer vision. In general, image segmentation is defined as a process of partitioning an image into homogenous groups such that each region is homogenous, but the union of no two adjacent regions is homogenous [14]. Also, the representation of an image is changed into something that is more meaningful and easier to analyze and make it easier to select the region of interest. Image segmentation is considered as the first step in medical image analysis applications, and also one of the most critical tasks of image analysis. Accurate image segmentation significantly helps in the process of extracting information and feature measurement (represented by data), even help with the following higher-level tasks such as object classification. In our case, melanoma detection depends on the skin lesions features; any mistake in the image segmentation to find the lesion boundary may dramatically affects the lesion features such as size, regularity, roundness and other characteristics.

It has been noted that none of the developed segmentation algorithms are generally applicable to all images, and different algorithms are not equally suitable for particular applications. Therefore, the goal of the research at this stage is to find and test an effective and appropriate method of the existing image segmentation approaches.

There are too many algorithms in image segmentation field, two state-of-the-art techniques in skin lesion segmentation were discussed in some important related work:

- Otsu's method
- Mean shift segmentation Method

OTSU'S THRESHOLDING METHOD:

If we assume that the lesion boundary to some extent clearly defined and distinct from the background, then using a simple thresholding approach like Otsu's method is sufficient to convert the image to a binary and separate the lesion from the background, Otsu's method is a non-parametric approach for global histogram thresholding [15]. It was developed to calculate optimal threshold from image histogram. One of the advantages in this method is focusing on nonparametric approaches, that support the process of automate the skin lesion detection with the mobile application without user thresholding level selection or adjustment, so Otsu method is one of the best automatic thresholding methods, the basic principle in Otsu method is to split the image into two classes which are the objects and the background. The automatic threshold is obtained through finding the maximum variance between the two classes

Otsu's thresholding Algorithm:

- Let $I = [1, L]$ is the range of grayscale levels of image $f(x, y)$ and p_i is the probability of each level
- The number of pixels with gray level i is denoted f_i giving a probability of gray level i in an image as $p_i = \frac{f_i}{N}$
- The automatic threshold t that divides the range into two classes which are $C_0 = [1, \dots, t]$ and $C_1 = [t+1, \dots, L]$, the gray level probability distributions for the two classes are $C_1 \rightarrow \left[\frac{p_1}{\sum_{i=1}^t p_i} + \frac{p_2}{\sum_{i=1}^t p_i} \dots \frac{p_t}{\sum_{i=1}^t p_i} \right]$ $C_2 \rightarrow \left[\frac{p_{t+1}}{\sum_{i=t+1}^L p_i} + \frac{p_{t+2}}{\sum_{i=t+1}^L p_i} \dots \frac{p_L}{\sum_{i=t+1}^L p_i} \right]$
- means for classes C_1 and C_2 are $\mu_1 = \frac{\sum_{i=1}^t i \cdot p_i}{\sum_{i=1}^t p_i}$ And $\mu_2 = \frac{\sum_{i=t+1}^L i \cdot p_i}{\sum_{i=t+1}^L p_i}$
- Let μ_T be the overall mean of the whole image. Obviously, by summing the parts, it is easy to show that $\mu_T = \beta_1 \mu_1 + \beta_2 \mu_2$ where $\beta_1 = \sum_{i=1}^t p_i$ and $\beta_2 = \sum_{i=t+1}^L p_i$
- From statistics, it is clear that the total sum of the probabilities is always equal to one $\beta_1 + \beta_2 = 1$

- Otsu defined the between-class variance of two classes C1 and C2 as

$$\rho^2 = \beta_1(\mu_1 - \mu_t)^2 + \beta_2(\mu_2 - \mu_t)^2$$

- The optimal threshold t is the value that maximize the between-class variance σ^2

$$t = \text{Max} \{ \sigma^2(t) \}, 1 \leq t < L$$

MEAN SHIFT SEGMENTATION METHOD

In the previous section, we assumed that the lesion is clearly defined and this is not the case always. Therefore, a more accurate segmentation algorithm is proposed, called pixels clustering, which has the following advantages: first it's based on pixel color, intensity, and location, or a weighted combination of these factors not only on the color intensity. Second, we have more than two regions as in the out's method, this may help in the case of blurred edges or noisy skin. One of the simplest clustering algorithms is K-means algorithm [16]. K-means is an iterative technique that is used to partition an image into K clusters, pick K cluster centers, either randomly or based on some heuristic. K-means algorithm is not appropriate for our case, we are looking for an automated non-parametric algorithm. On the other hand, mean shift segmentation [17], used for clustering the feature vectors associated with each pixel (e.g., color and position), but one of the most important difference between K-means and Mean shift is the initialization of the cluster number k , Mean shift a non-parametric algorithm does not assume anything about number of clusters (as required in k-means).

Mean shift segmentation Algorithm:

Without discussing the complex mathematical proof to Mean Shift algorithm, the algorithm can be summarized as follow:

1. For each point x_i , Compute the mean to the feature vectors associated with each pixel within the kernel, this step depend on the kernel size, we can use Gaussian kernel, and the only parameter we need to initialize is the kernel size.

2. Compute mean shift vector, the distance between the point and the mean, $m(x_i^t)$.
3. Move the point by $m(x_i^t)$.
4. Repeat until convergence
5. All points associated with the same stationary point belong to the same cluster.

3.3.4. Maximum Contour Extraction

After image segmentation, morphological close operation could be used to remove the small gaps within the lesion, then the contours are ready for extraction. " A contour is a list of points that represent, in one way or another, a curve in an image. This representation can be different depending on the circumstance at hand" [18]. The contours are a useful tool for shape analysis and object detection and recognition, they are used to find the lesion shape properties, such as the lesion centroid, borders, roundness and regularity. The OpenCV contour extraction implementation was based on border following algorithms algorithm proposed in [19], the contours created are sequences of vertices (i.e. Points).

If the image has more than one contour, the largest contour area could be selected in most cases, where the largest contour is assumed to be the lesion boundary. This assumption is not always correct, in some failed segmentation cases, for example, with effect of light reflections and shadows, the size of the contour is larger than two-thirds the entire image, but the second largest area is mostly the lesion. Our proposed solution to this problem is to filter the contours areas, which are larger than two-thirds or smaller the 10 pixels are likely to be a noise and can be neglected. The return value from the contour extraction process is a list of contours; each of them includes sequences of vertices. The following algorithm is used to select the lesion contour from the contours list if the first greater than two-thirds the entire image:

1. Find all the contours in the image, store them in a list;

2. Sort the contours in the list and remove both the contours smaller than 10 pixel and larger than two-thirds of the image (the two-thirds value may be changed after testing later);
3. Find the maximum contour area;
4. And store the contour points sequence in a list to be used later.

We need to find both the contour "border" and the lesion area "fill"; the border is used to analyze the lesion shape features like the irregularity, and the "fill" to analyze color variance and symmetry. The next step is to isolate the lesion area from the rest of the skin, after the border was identified.

3.3.5. Isolating the Lesion Area

To isolate the lesion region from an input image, the contour area is filled with solid color and then used as a mask to the image. Flood-fill algorithm is used [13] to fill the contour area, the algorithm goes as follows:

1. Select a seed point inside a region as shown in Figure 3-3(a);
2. Move outwards from the seed point as shown in Figure 3-3(b),
 - a. If pixel is not set, set pixel,
 - b. Process each neighbor of pixel that is inside the region;
3. And stop when no more seed points as shown in Figure 3-3(c).

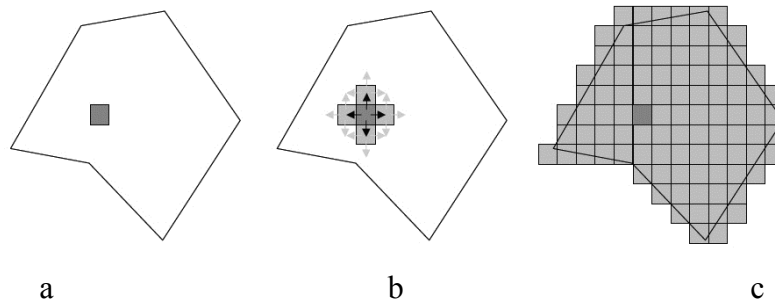


Figure 3-3: Flood-fill algorithm
From the left, a: select a seed point, b: move outwards to neighbors, c: stop when the region is filled

After finding the lesion contour and isolate it from the image, the next step is to find the lesion center, axis, orientation and other shape features. The next section discusses using of moments to analyze and extraction of the center, semi-axes and orientation of the lesion.

3.3.6. Finding the centroid and the symmetry axis

The moments are features of the object, which allow a geometrical reconstruction of the object. Moment function has been widely used in image analysis, since calculated moment of a digital image not only describes the shape of the image's global features generally, but also provides a mass of different types of geometric feature information about this image, such as size, location, direction, shape and so on [20].

The concept of algebraic invariable moment was proposed by Hu in his first paper about the application of 2-D Moment in pattern recognition [21]. In order to understand the characteristics of moment, some definitions are given as follows:

To any non-negative integer p and q , the definition of $(p + q)$ order moment on the plane R of two-dimensional image model $f(x, y)$ is [20]:

$$m_{p,q} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy$$

For digital images, this equation can be written as

$$m_{p,q} = \sum_x \sum_y x^p y^q f(x,y)$$

The first order moments contain information about the area $m_{0,0}$ and center of gravity of the object, the center (X_c, Y_c) ;

$$X_c = \frac{m_{1,0}}{A} = \frac{m_{1,0}}{m_{0,0}}$$

$$Y_c = \frac{m_{0,1}}{A} = \frac{m_{0,1}}{m_{0,0}}$$

Where:

$$A = m_{0,0} = \sum_x \sum_y f(x,y)$$

$$m_{0,1} = \sum_x \sum_y x f(x,y)$$

$$m_{1,0} = \sum_x \sum_y y f(x,y)$$

The Orientation \emptyset can directly be derived from moments by the formula:

$$\emptyset = \frac{1}{2} \arctan \frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}}$$

Where $\mu_{1,1}$, $\mu_{2,0}$, $\mu_{0,2}$ can be derived from the first and second order moments:

$$\mu_{1,1} = \frac{m_{1,1}}{m_{0,0}} - \left(\frac{m_{1,0}}{m_{0,0}} * \frac{m_{0,1}}{m_{0,0}} \right)$$

$$\mu_{2,0} = \frac{m_{2,0}}{m_{0,0}} - \left(\frac{m_{1,0}}{m_{0,0}} \right)^2$$

$$\mu_{0,2} = \frac{m_{0,2}}{m_{0,0}} - \left(\frac{m_{0,1}}{m_{0,0}}\right)^2$$

3.3.7. Real time interactive image capture algorithm

We have discussed the image capture in section 3.1.1, and showed that the image can be selected either from external storage SD card or captured from the mobile camera. There are some points that need to be discussed when using a mobile camera, for example, we assumed in the previous sections that the selected image from the SD card is ready to be processed, and the lesion region of interest ROI is cropped. If the image contains a large area of the skin in addition to the lesion, it is possible to get a lot of noise, shadows and reflection, thus the image processing becomes very difficult. On the other hand, choosing a very small image may leads to lose some parts of the lesion, so when the image is captured from the camera we need a mechanism to automatically select and crop the lesion region, this mechanism must be processed easily by the user, while ensuring full determination of the lesion area and reduction the amount of skin around the lesion.

Augmented Reality (AR) can be used to assist the user in determining the success of the process of taking picture in real time, this solution is used in doctor mole application in section 3.1.1. In this research, the proposed algorithm to select the lesion from the mobile captured image can be summarized as follows:

1. The user focus the mobile camera to lesion area.
2. Capture image with the mobile default mobile image, at frame rate 15 fps.
3. Cropped the centered 300 pixel square, work within it to reduce heavy computation.
4. Do the image preprocessing, segmentation, thresholding, contour extraction as described in previous sections.
5. Filter the noise contours area.
6. Find centroid of each contour.

7. Select the nearest contour center to the 300 pixel square center.
8. Highlighted the contour area and enable the capture button then show message to the user to click on it.
9. Select smallest rectangle that include this contour area.
10. While the user not pressed the capture button, or no contour discovered within the target area, return to the first step.
11. Crop the selected area from the image and send it to more details analysis.

When the user clicks the capture button, the region automatically cropped to the smallest rectangular around, the lesion contour is filled using flood-fill and masked. Now we have small image with specified lesion.

The proposed solution needs more testing to the frame rate value and the captured image area to have best performance.

3.4.Feature Extraction

The next step after selecting both the lesion area and contour is the feature extraction depending on the method of ABCD. This section will discuss each feature in details; Asymmetry, Border, Color, and Diameter.

3.4.1. Asymmetry

In ABCD approach for the evaluation of asymmetry, the lesion is divided into four segments using the two axes which pass through the lesion centroid [4]. Many automated algorithms achieve the ABCD rule and evaluate asymmetry across axes within the lesion [21]. The results in section 3.1.6 are used to achieve the symmetry which define the centroid and symmetry axis, the previous method depends mainly on the lesion boundaries without consideration to the internal color. Knowing that the symmetry definition is measured based on both border and internal color, but from observation to a large number of lesions

and the testing on a melanoma dataset, internal lesion color and details do not carry a lot of impact to determine the lesion center and axis, and might cause significant delay.

Based on the results in section 3.1.6, we find the lesion centroid, symmetry axes and the lesion orientation. The first step rotate the lesion to fit both the horizontal and vertical axis, and the symmetry obtained by overlapping the two halves of the lesion along the principal axes of inertia, and dividing the non-overlapping area differences of the two halves by the total area of the lesion. An asymmetry index is computed for both axes using the following definition:

$$\text{Asymmetry index} = \frac{\Delta A}{A}$$

Where ΔA is the nonoverlapping area between the original and reflected masks, and A is the area of the original mask.

3.4.2. Border

There are different measures to calculate border irregularity, such as shape compactness or circularity (sometimes it is expressed as shape factor) [22], and usually associated with the ratio $C = \frac{P^2}{A}$, where C is the value of shape compactness, A is the shape area and P is the shape perimeter. For a circle is simple and is given by $C = \frac{A}{P^2}$ the factor 4π used to normalize and ensures that C equals 1 (the lowest possible value) for a circle so we have the compactness index equation.

$$\text{Compactness index} = \frac{p^2}{4 \cdot \pi \cdot A}$$

The reciprocal $1/C$ is sometimes called the Roundness of an object [23].

$$\text{Roundness} = 4\pi \frac{A}{p^2}$$

Similarly, we use the ratio of the lesion area to the circle area with the same perimeter. This formula used in [22-23-24] as a measure of shape irregularity, where A is the area within the lesion contour and P is the perimeter of the lesion contour. The maximum value of the compactness index is 1, and occurs when the lesion contour is a circle. The value of compactness index decreases as the lesion contour becomes more irregular.

Another border irregularity was computed in [24], where irregularity is defined by the formula:

$$\text{Irregularity} = \frac{P}{A}$$

This ratio depends on the size of the skin lesion, therefore, the calculation of the irregularity is more correct as the ratio:

$$\text{Irregularity} = \frac{P}{\text{Geatest Diameter}}$$

Another border related feature calculated in [25] is the Hull/Contour Ratio (HCR), this is defined as the ratio of lesion convex hull to the perimeter. The convex hull of a contour is the smallest polygon that encloses it [26]. Lesion with notched, ragged edges should have more convexity defects than a smooth lesion.

3.4.3. Color

One of the early signs of melanoma is the lesion color and color variance, melanoma cells are often colorful around brown, or black, depending on the production of the melanin at different depth in the skin. Color description is mainly statistical parameters calculated from different color channels. In [24] the RGB (Red, Green, blue) and the HIS (Hue, Intensity, Saturation) were analyzed, and the color variation was calculated by measuring standard deviation, while the lesion average color was computed from both RGB and HIS. In [27] the RGB (Red, Green, blue) and HSV (hue-saturation-value) were analyzed. Therefore, the minimum, maximum, average and standard deviations of the RGB and HIS measured.

3.4.4. Diameter

Melanoma tends to grow larger than common moles, especially those with 6mm diameter. Because of the irregular forms of the lesion, the minimum bounding circle is used to find the greatest diameter, and the minimum bounding rectangular is used to find the width and height of the lesion region ROI, where all images are assumed to be scaled to 1:1. Scaling task is not easy, since many of the dataset images used in this thesis were selected from the Internet without information about the image scale. Besides, the captured images from the camera depend entirely on camera zoom and distance, which are unknown values and difficult to estimate. On the other hand, it is important to use the size property in the classification task. Therefore, the size property has been used in both feature extraction and classification process, and has been included in the classification model.

3.5. Classification

These sections discuss the classification concepts and methods. Classification is a shared concept used in many fields, such as data mining, machine learning and expert system. Before starting the review of the used classification methods, we will discuss some general concepts related to the classification, such as data mining and machine learning.

Machine learning *investigates how computers can learn (or improve their performance) based on data* [28]. Machine Learning field is a branch of Artificial Intelligence, in which machines are prepared to imitate intelligent skills of humans. Machine learning is programming computers to optimize a performance criterion using example data or past experience [29]. Machine learning can be used when we do not know an exact algorithm to solve a task, but we can provide enough data to learn from (e.g. hand-written text recognition).

A main research area is for computer programs to automatically learn to recognize complex patterns and make intelligent decisions based on data. For example, a typical machine learning problem is to program a computer so that it can automatically recognize handwritten postal codes on mail after learning from a set of examples.

There are many similarities between data mining and machine learning. For classification and clustering tasks, machine learning research often focuses on the accuracy of the model. In addition to accuracy, data mining research places strong emphasis on the efficiency and scalability of mining methods on large data sets, as well as on ways to handle complex types of data and explore new and alternative methods.

Data mining *is a process to extract the implicit information and knowledge which is potentially useful and people do not know in advance, and this extraction is from the mass, incomplete, noisy, fuzzy and random data* [30]

Data mining use automated data analysis techniques to uncover previously undetected relationships among data items by analysis of data stored in a data warehouse. Three of the major data mining techniques are regression, classification and clustering. Data Mining, also popularly known as Knowledge Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. While data mining and knowledge discovery in databases (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process.

So we can look at our system as a Machine learning system, where a melanoma dataset available and we need to build a classification system based on this data to classify the new cases.

Another concept we are discussing is **Supervised Learning**. Basically, the supervision in the learning comes from the labeled examples in the training data set, for example, in melanoma case, the melanoma training examples are labeled as benign or cancer, which supervise the learning of the classification model.

Methods of feature extraction and selection described in the previous sections. In order to classify each image based on those extracted features, a classification method is required to distinguish between melanomas and benign lesions. **Classification** *“is a form of data analysis that extracts models describing important data classes. Such models, called*

classifiers, predict categorical (discrete, unordered) class labels” [28]. Four different classifiers will be discussed in the next section: k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Artificial Neural Network ANN and the Decision Tree classifier. The goal is to find the best classifier from previous methods and implement it into our system, selection was based on related work that we discussed in chapter 2.

3.5.1. KNN classification algorithm

KNN is one of the simplest and rather trivial classifiers, which memorizes the entire training data and finds a group of k training set that are closest to the test object, and bases the assignment of a label on the predominance of a particular class in this neighborhood [31]. There are three key elements of this approach: a) a set of labeled objects, e.g., a set of stored records; b) a distance or similarity metric to compute distance between objects; and c) the value of k, the number of nearest neighbors. To classify an unlabeled object, the distance of this object to the labeled objects is computed, its k nearest neighbors are identified, and the class labels of these nearest neighbors are then used to determine the class label of the object (see Figure 3-4)

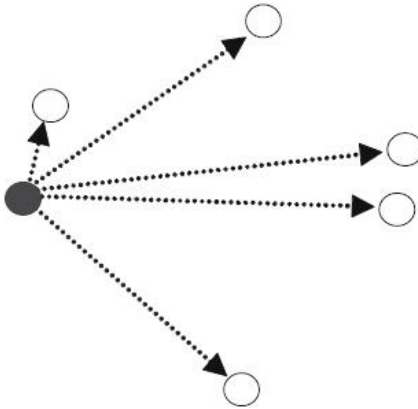


Figure 3-4: The principle of the nearest neighbor (NN) classifier.

Similarity or **Closeness** between objects is defined in terms of a distance metric, such as Euclidean distance. The Euclidean distance between two points or tuples [28], says

$$x_1 = (x_{11}, x_{12}, x_{13}, \dots, x_{1n})$$

$$\text{And } x_2 = (x_{21}, x_{22}, x_{23}, \dots, x_{2n})$$

Then the distance between x_1 and x_2 is

$$\text{dist}(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}$$

Few years ago, this method could not be applied on mobile applications due to the limitations of mobile memory and processing units, but as a result of the great development on the processing and storage components in mobile devices nowadays, good results in terms of speed and accuracy are expected to prove the possibility of using KNN method.

3.5.2. SVM classifier

The second classification technique is Support Vector Machines (SVM), one of the most robust and accurate methods among all well-known algorithms [31]. It has a theoretical foundation, requires only a dozen examples for training, and is insensitive to the number of dimensions.

In a two-class learning task, the aim of SVM is to find the best classification function $f(x)$ to distinguish between members of the two classes in the training data. The metric for the concept of the “best” classification function can be realized geometrically. $f(x)$ that passes through the middle of the two classes, separating the two. Once this function is determined, new data instance x_n can be classified by simply testing the sign of the function $f(x)$; x belongs to the positive class if $f(x) > 0$ as shown in Figure 3-5.

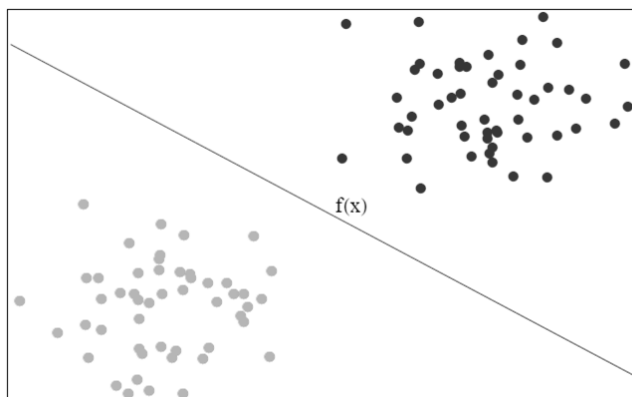


Figure 3-5: SVM Linearly separable data

SVM additionally guarantee is that the best function is found by maximizing the margin between the two classes. Intuitively, the margin is defined as the amount of space or separation between the two classes. This is why (during the learning or training phase) the SVM searches for the largest margin, that is, the maximum marginal hyperplane (MMH).

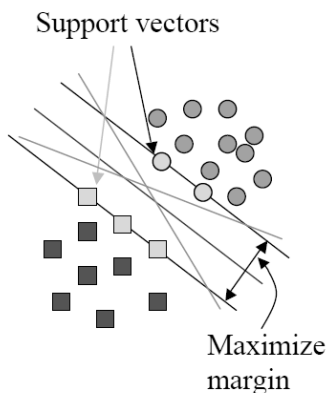


Figure 3-6: SVM with many possible separating hyperplanes.

3.5.3. Neural networks classifiers

An Artificial Neural Network (ANN) is an information processing paradigm inspired by the way biological nervous systems, such as the brain and process information. The key element of this paradigm is the novel structure of the information processing system that composed a large number of highly interconnected processing elements (neurons) working

in unison to solve specific problems. ANNs, like people, learn by example. ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons, which is true for ANNs as well.

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be considered as an "expert" in the category of information that has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.

The model used is multilayer perceptron; it is a feed-forward [32] artificial neural network model that maps sets of input data onto a set of appropriate output. It is a modification of the standard linear perceptron as it uses three or more layers of neurons (nodes) with nonlinear activation functions, and is more powerful than the perceptron as it can distinguish data that is not linearly separable, or separable by a hyper-plane.

3.5.4. Decision tree classifier

The final classification model tested is J48 classifier, it's a simple decision tree for classification, was chosen because it is easy to implement and apply to mobile platform. J48 creates a binary tree, and with this technique, a tree is constructed to model the classification process. Once the tree is built, it is applied to each tuple in the database and results in classification for that tuple. While building a tree, J48 ignores the missing values; for example, the value for that item can be predicted based on what is known about the attribute values for the other records. The basic idea is to divide the data into ranges based on the attribute values for that item found in the training sample. J48 allows classification via either decision trees or rules generated from them.

4. Implementation

This chapter presents the implementation of this research, it describes both the software tools and hardware equipment used to build and test both the internal modules and the end-user system, then it shows the experiments and procedures for the proposed methodology.

4.1. Software

An appropriate android programming library is required for image processing and machine learning fields. **OpenCV** (Open Source Computer Vision Library) is one of the best library in android platform. It is an open source computer vision and machine learning software library [33] that was built to provide a common infrastructure for computer vision applications. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. OpenCV is a free library that support windows and also has C++, C, Python, Java and MATLAB interfaces [34]. On the other hand, more than one option is available, such as ml4android [35] that attempts to use a wide range of algorithms. The Android version of Weka [36] is a collection of machine learning algorithms for data mining tasks [37] that is also used for machine learning. The algorithms can either be applied directly to a dataset or called from java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. It is well-suited for developing new machine learning schemes, and is also used in classification testing.

Weka has been adopted in this research for the following reasons; first, it's one of the most popular machine learning tools with implementation to most ML algorithms. Second, two versions of Weka are available; desktop and android library, which can be used for testing and in mobile systems, respectively. Finally, its java based and open source that make it easy to integrate with android platform. The implementation in this research goes through the following steps, first, testing the image processing algorithms using the windows version of the OpenCV before using the android version of the library, and second, using

the windows version of Weka to test the best classification algorithm to be used in mobile systems using Weka for android library.

4.2.Hardware

Android platform has been chosen in this research as mobile devices are promising and has a wide range of ideas in this field. Android is the fastest growing mobile platform; millions of devices (namely, mobile phones, tablets, etc.) run Android [38]. It has two main features; first, most of the applications are available at a single place, which is natively supported by the platform where anyone can make his application publicly visible while users know where they can find it. Second, it is a platform for which the development is supported very well; it uses one of the most common programming languages Java, and there are lots of tutorials and support especially for its SDK (Software Development Kit).

The system will be tested using eclipse IDE with virtual simulator, and for more accurate results, the final version of the system will be tested on several android based device with different storage and CPU capabilities, such as galaxy tab 3, galaxy note and nexus tab.

4.3.Dataset

The date set used includes 164 images collected from the web site in [12], the same source used in related work. The dataset used is balanced, include 48.7% labeled as melanoma and the rest 51.3% benign lesion, this dataset used in feature extraction and classification. Another dataset captured with mobile camera to prove the success of the mobile image capture and ROI selection, firstly, the dataset features have been extracted and stored in a database using a java program in Eclipse environment, the used set stored in CVS Excel file then converted to REFF Weka file to test the classification. After the success in the image processing and the classification testing, depending on the results and the features database; implemented a mobile application to prove the proposed method, android OpenCV have been used in the image processing, and the Weka android version API in classification.

4.4.Experiments procedures

As previously noted, first phase described the implementation of a program to extract the image features, we have 12 feature in the dataset, and then the extracted features stored in a database. To complete this, we need to programing a tool for image processing according to the procedures in the proposed method. Next the Weka can be used to test the different classification techniques and finding the best and most accurate one.

Then to prove the success of the image processing and classification in the mobile environment, we need to implement the program of Mobile uses the best image processing methods as well as the best way found in the classification testing, eclipse environment with OpenCV and Weka APIs are used in this application.

The next procedures summarize the experiments and implementation work

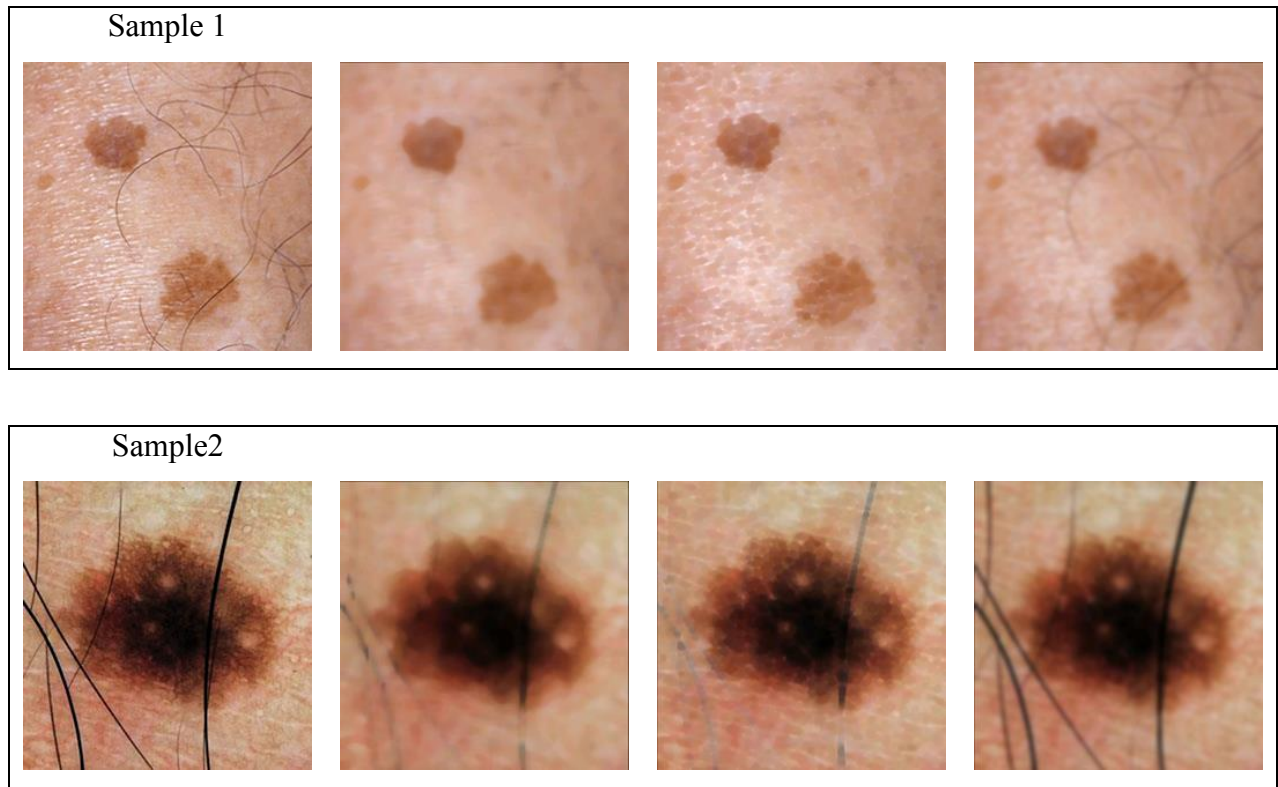
1. Select the image data set
2. Test and find the best image processing methods (noise removal – segmentation – automatic ROI selection – features extraction), as we discussed in the proposed method
3. Implement application to extract the image features and store them in the database
4. Using Weka to find the most accurate classification technique with database in step2
5. Implement mobile application based on the results in step 2, 3, 4. The dataset from step 3 used as training dataset classification technique.

5. Results and Evaluation

This section presents the experimentation of the proposed methodology, where both image processing operations and different classification techniques will be discussed. The data set used includes 164 images from the internet [12] where 14 images of them failed in image processing stage, so the actual number of images used in classification was 150. All images were captured by a normal camera with up to 5 megapixel; neither special equipment was used nor a high resolution camera customized for medical purposes.

5.1. Image processing results

Initially, the image is processed as mentioned earlier to remove the unwanted noise. We have two main sources of noise, the light reflection and the hair. The former source is the most annoying problem, due to uneven lighting in most of the captured images especially in hands and faces. Figure 5-1 shows the noise removal filters results; "A" is the original image, "B" when using Median Blur the reflections are significantly reduced and the fine hair completely removed, "C" use Close Morphological Operation the hair removed but the reflections increased, "D" the Gaussian Blur with bad effects on the lesion boundary. Median filter shows the best results where both hair noise and light reflection are significantly reduced.



A Original

B Median

C Morphological

D Gaussian

Figure 5-1: Results of using noise removal filters

After the noise removal phase, the image is converted from RGB mode to HSI, then the intensity channel is extracted and converted to gray scale, and finally, segmentation algorithm is used to isolate the melanoma region from the skin. Otsu method tested on some images from the dataset, it fails to extract the lesion ROI in more than 30% of the image as shown in Figure 5-2. Otsu thresholding method is inappropriate due to the following: first, it is very sensitive to noise even after using noise removal algorithms like median blur filtering, and the texture of the skin increased this problem as shown in Figure 5-3. Second, the problem of skin shadows; Otsu algorithm assumes there are two clear and separated regions while this is not the case with hand and face images which have shadows in the corners. In an attempt to overcome these problems, Mean Shift

Segmentation algorithm is tested and the Otsu method used to binarize the resulting image. As shown in Figure 5-3, the output of the Mean Shift Segmentation algorithm is RGB image with reduced color. Using Otsu thresholding on the segmented image produce good results, because most of the shadow and skin texture problems are resolved by the segmentation step.

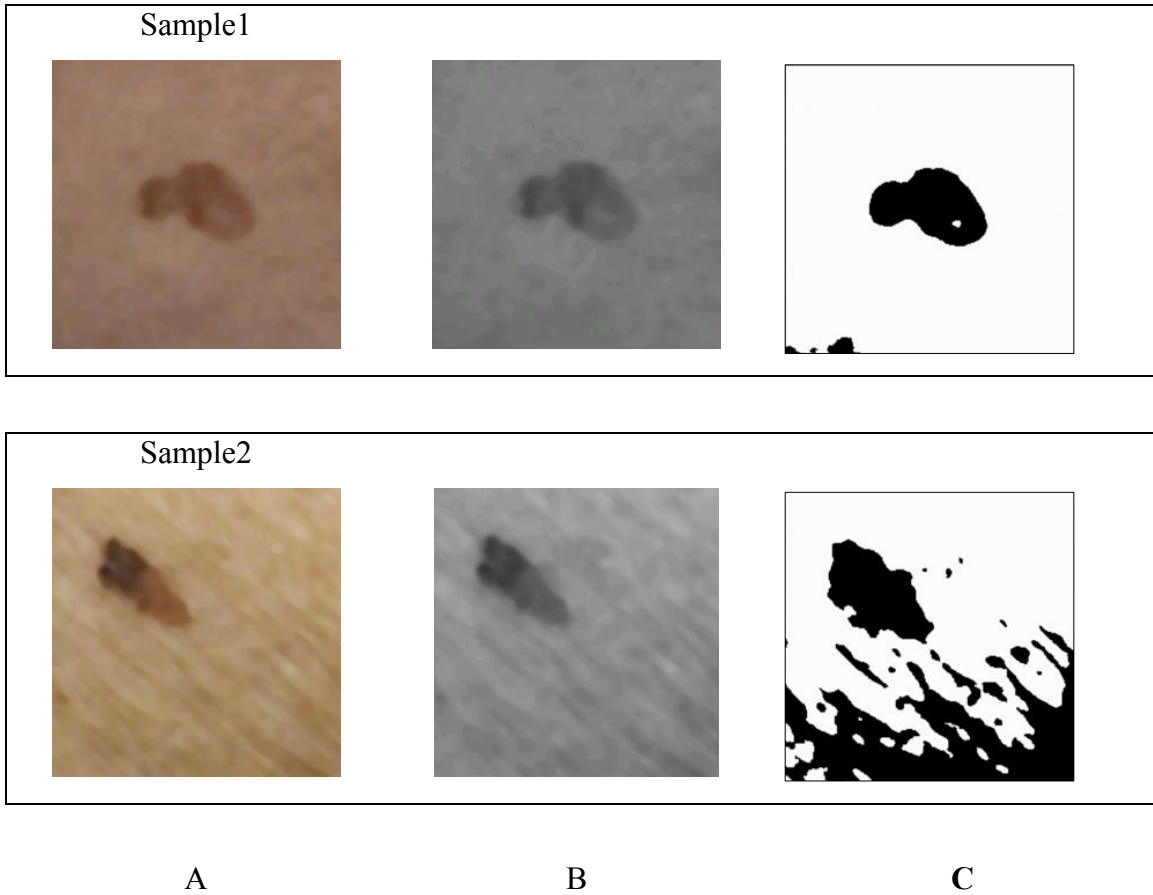


Figure 5-2: Otsu image segmentation method

“A” Median Blur Filtering, “B” extract Hue channel from HSI image, “C” Otsu Segmentation

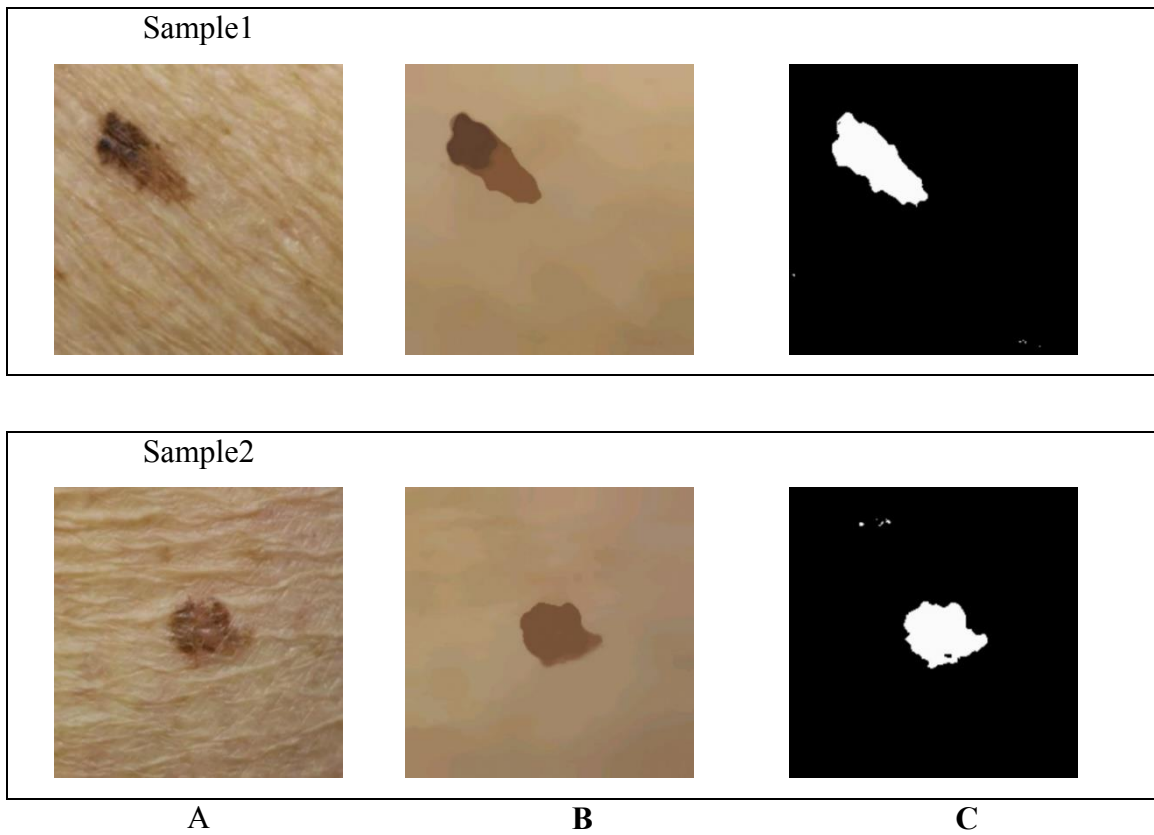


Figure 5-3: Mean segmentation method
“A” The image before any processing, “B” Image with Mean Shift Segmentation and Median Blur filtering, “C” Otsu Segmentation

The next step after the segmentation process is to find the best lesion contour; there are two problems in the contours results, first, the presence of more than one lesion or noisy areas in the image as shown in A, B, C from Figure 5-4. Second, the large contours resulting from shadows and light reflection shown in D, E, F from Figure 5-4.

The proposed solution for these problems is filtering the contours areas as discussed in section 3.1.5, the contours larger than two-thirds or smaller the 10 pixels are removed. In in Figure 5-4. A and D show the images before segmentation and contour extraction. B shows the image with three noisy contours while E shows lesion contour with large shadow area contour. C preset the image after applying the contour selection algorithm as discussed

in section 3.1.5, and select the largest contour area after removing the noise, and finally, F show the selection biggest contour after removing shadow and reflection area.

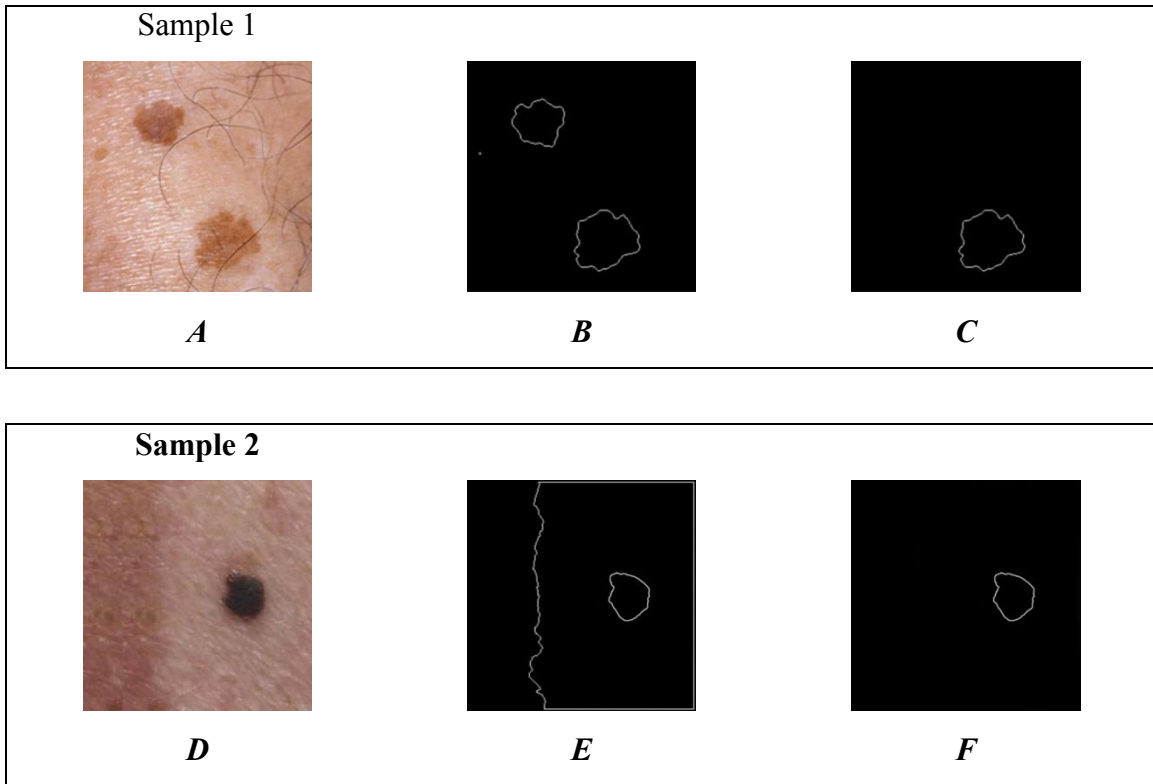


Figure 5-4: Selecting the best lesion contour

The noise filtering to contours with smaller than 10 pixels or greater than two-thirds of the image size depends on testing the dataset to determine the nature and size of the contours generated by the noise and light reflection problems, so that leads to the best possible results, and this value depends directly on the size of the images used in the system.

Next, it is easy to determine both the centroid and the semi axis in the extracted contour in the previous step. Firstly, as discussed in section 3.1.7, calculate the moments to the contour and find the centroid and orientation from the moments, then use the centroid and orientation (namely, semi axes) to find the first and second symmetry axis, Figure 5-5 show the centroid and axis extraction.

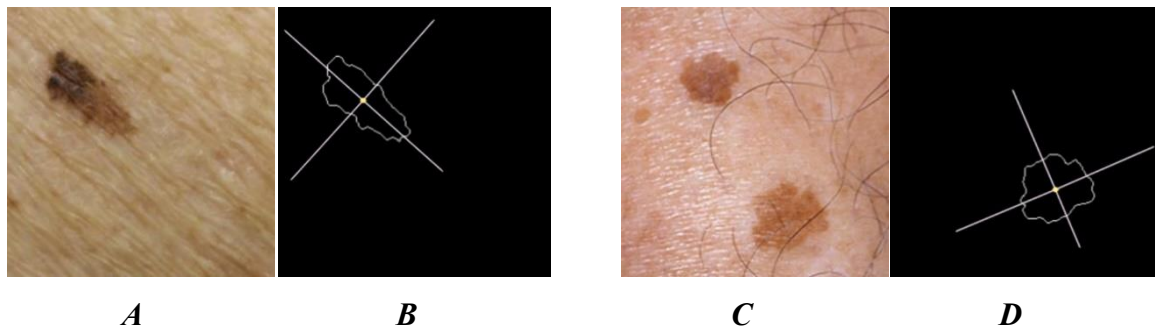


Figure 5-5: Lesion centroid and semi axis

As shown in Figure 5-6, Flood fill algorithm is implemented starting from the lesion centroid to get the mask image, and the resulting image shown in C was used as a mask to the input image to get the isolated lesion in D.

The border image “Contour image” B, will be used to find features related to the lesion borders shape and irregularity. On the other hand, the isolated lesion shown in D used to find a symmetry about the axes that found previously and analyze both the lesion colors and size.

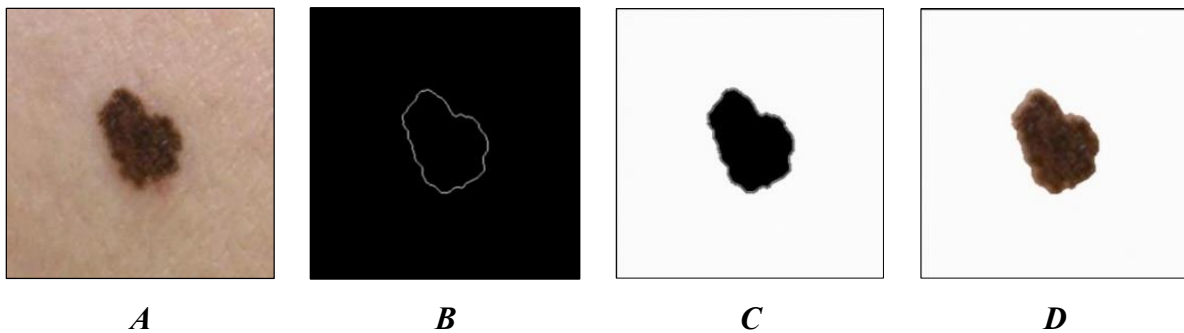


Figure 5-6: Using Flood fill algorithm to isolate the lesion

A the input image, B the contour of the lesion, C flood-fill the contour, D the input image masked with the flood-fill image

Up to this stage and before discussing the process of melanoma feature extraction based on the ABCD, the image processing operations were used to determine the edges of the lesion and isolated it from the rest of the skin. This part discusses the user interface to facilitate

images capturing, and also to overcome the various problems arising while an image is captured.

Figure 5-7 shows the implementation and the user interface related to the algorithm discussed in section 3.1.8. The system automatically detects the lesion, highlights it and enables the capture button. In case of more than one lesions, the nearest one to the circle in the middle is selected, any lesion greater than the circle area or smaller than 10 pixel is removed. Finally the smallest rectangular around the lesion is selected and ROI is cropped. Several modifications are added to the interface to get the required accuracy, such as camera auto-focus, camera zoom and image capture frame rate 15 fps to work in real time. In addition, filtering of noise is improved by isolating areas outside the center of the image and remove areas of small or very large size.

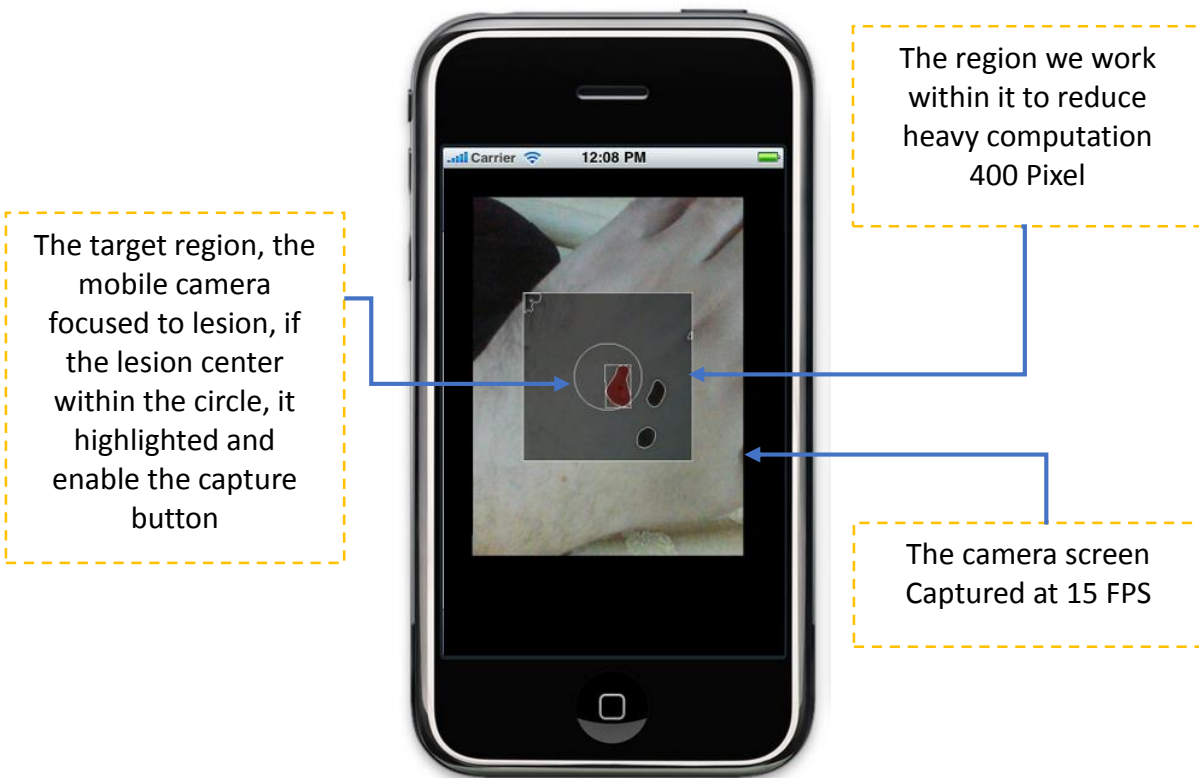


Figure 5-7: Melanoma lesion image capture interface

5.2.Feature Extraction

Depends on the results of the previous section, the borders of the lesion and the required ROI are determined, then, the ABCD melanoma features are extracted. The first feature is **Asymmetry**, which can be extracted using the approach in section 4.2.1, the lesion firstly rotated to the X symmetry axis which passes through the lesion centroid, X asymmetry obtained by overlapping the two halves of the lesion, and dividing the non-overlapping area by the total lesion area. Finally, repeat the same process for the Y axis to find the Y asymmetry index. Results are shown in Figure 5-8.








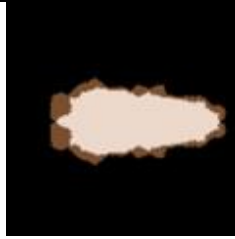

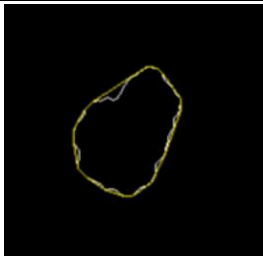

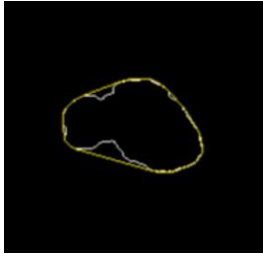
Image	Rotated Image	X axis symmetry	Y axis symmetry	Asymmetry index
				<p>X-axis = .510 Y-axis = .281</p>
				<p>X-axis = .397 Y-axis = .609</p>

Figure 5-8: Lesion X,Y Asymmetry samples

The second feature is the **Border irregularity** that has been previously discussed in section 4.2.2, this feature, as shown in Figure 5-9, includes convex hull, HCR, roundness and irregularity. Note that some of these features may be ineffective in the classification

process, and some of them did not give good results, however, all related features are calculated and the process of determining their impact is left to the classification stage.

Image	Convex hull	HCR	Roundness	Irregularity
		.920	.719	2.88
		.892	.620	2.81




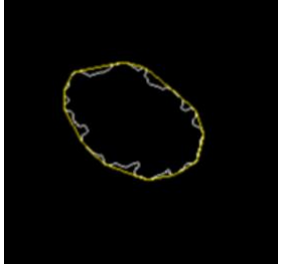
		.897	.529	2.63
		.791	.523	3.22

Figure 5-9: Lesion Border property

The third feature is the Color variance, the means and standard deviations of the lesion RGB values computed. Both the source image and the mask image were used, and the subset of pixels from the source image that correspond to pixels in the mask are also calculated. The color variance sample results are shown in Figure 5-10.







Image			
Mean Color			
R Mean	86.6	116.5	51.8
B Mean	57.9	81.8	46.5
G Mean	43.7	56.2	49.7
R Std.	25.1	24.5	25.9
B Std.	21.7	21.0	23.3
G Std.	19.2	16.7	22.9

Figure 5-10: Color Mean and standard deviation analysis

The final feature is the **Diameter**, all images are assumed to be scaled to 1:1 to simplify the computation. It is not easy to implement algorithms to resize the images to 1:1 scale due to the lack of information (namely, actual size) for images from the Internet, and even the images captured from the mobile camera depend on the focus and zoom. Therefore, only images dataset with known size is used, then the properties related to Diameter ,such as minimum bounding rectangle, smallest enclosing circle, area, diameter to the smallest circle, contour perimeter and width and height to the minimum bounding rectangle are computed as shown in Figure 5-11.




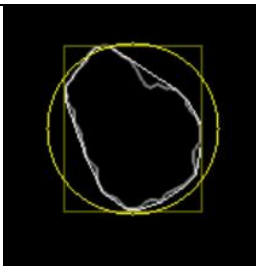

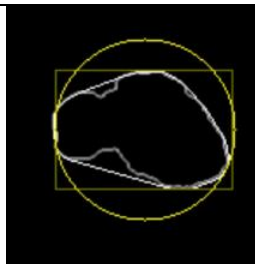
Image			
Boundaries			
Area	4599.5	3069.5	4378.0
Diameter	98.3	102.4	105.9
Perimeter	283.4	269.8	297.9
Radius	49.2	51.2	52.9
Width	78	82.0	101.0
Height	94	81.0	68.0

Figure 5-11: Lesion size properties

5.3. Classification

The following section presents the classification process of skin lesions using both the proposed system and the methods described in Chapter 3.3. The results of feature extraction process have been stored in a file; the features include 12 attributes for each record, 73 instances are classified as melanoma, 77 instances as benign case and the rest 14 instances are not classified (namely, failed in previous image processing stage); a WEKA program is used for testing the classifiers. The used classifiers are k-Nearest Neighbors (KNN), support vector machines (SVM), Neural Networks (ANN) and Decision Tree (DT), where sensitivity, specificity and accuracy of results are calculated for each classifier.

This section presents the implementation of the final results. The research work attempted to achieve the following: a) improving the process of image capturing, overcoming the problems related to noise and light reflection, and c) improving the results by using more than one way. However, there is a percentage of error in the image processing results, where the system failed to identify 14 out of 164 images in the dataset. The percentage of finding the lesion contour boundary is 91.5%.

K-Nearest Neighbors (KNN) is the simplest classifier used throughout this thesis. Table 5-1 presents the KNN classification results, where the dataset is spited to 66.0% for training and 34% for testing, the K value is tested for the values 3,4,5 and the best result achieved with k=3.

Table 5-1 KNN classification results

Correctly Classified Instances	45	88.2353 %
Incorrectly Classified Instances	6	11.7647 %

The performance assessment of the classifiers used in this study is based on a comparison between the label assigned and the results achieved. Thus, the applied measures include a count of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN), where:

- a) TP: is number of images correctly classified as melanomas,
- b) FP: is number of images misclassified as melanomas,
- c) TN: is number of images correctly classified as nevi, and
- d) FN: is number of images misclassified as nevi.

The explanation of these measures could be simplified by the concept of confusion matrix that can illustrate the number of correct predictions in relation to the expected label. Each column represents the label in a predicted class.

Table 5-2 KNN Confusion Matrix, split the dataset to 66.0% training and the remainder testing

		Predicted Class		
		Melanoma (1)	Normal (2)	
Actual Class	Melanoma (1)	TP 20	FP 4	
	Normal (2)	FN 2	TN 25	
		Sensitivity = 0.909	Specificity = 0.862	Accuracy = 0.88

Table 5-2 presents a confusion matrix to the KNN algorithm, where the Sensitivity, Specificity and Accuracy results were calculated as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = 20/(20+2) = \mathbf{0.909}$$

$$\text{Specificity} = \frac{TN}{TN + FP} = 25/(25+4) = \mathbf{0.862}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = (20+25)/(20+2+25+4) = \mathbf{0.882}$$

Table 5-3 presents the results of the same previous experiment using cross-validation method instead of splitting the dataset into training and testing.

Table 5-3 KNN Confusion Matrix with cross-validation method

		Predicted Class		
		Melanoma (1)	Normal (2)	
Actual Class	Melanoma (1)	TP 68	FP 5	
	Normal (2)	FN 6	TN 71	
		Sensitivity = 0.92	Specificity =0.93	Accuracy = 0.93

The previous performance assessment has also been implemented for other classifiers discussed previously, namely, SVM, ANN and DT, where the results for these classifiers are shown in Table 5-4, Table 5-5 and Table 5-6, respectively, for more accuracy the cross-validation is used

Table 5-4 SVM classifier Confusion Matrix

		Predicted Class		
		Melanoma (1)	Normal (2)	
Actual Class	Melanoma (1)	TP 23	FP 1	
	Normal (2)	FN 2	TN 25	
		Sensitivity = 0.92	Specificity =0.96	Accuracy = 0.94

Table 5-5 Neural Networks classifiers Confusion Matrix

		Predicted Class		
		Melanoma (1)	Normal (2)	
Actual Class	Melanoma (1)	TP 23	FP 1	
	Normal (2)	FN 4	TN 23	
		Sensitivity = 0.85	Specificity =0.96	Accuracy = 0.90

Table 5-6 Decision Tree classifier Confusion Matrix

		Predicted Class		
		Melanoma (1)	Normal (2)	
Actual Class	Melanoma (1)	TP 20	FP 4	
	Normal (2)	FN 2	TN 25	
		Sensitivity = 0.91	Specificity =0.86	Accuracy = 0.88

As described in Figure 5-12, the visualization of the decision tree shows the greatest impact of certain features, such as the color (especially, the red standard deviation and the green mean) and the border irregularity. All the 12 extracted features are tested and used in the DT.

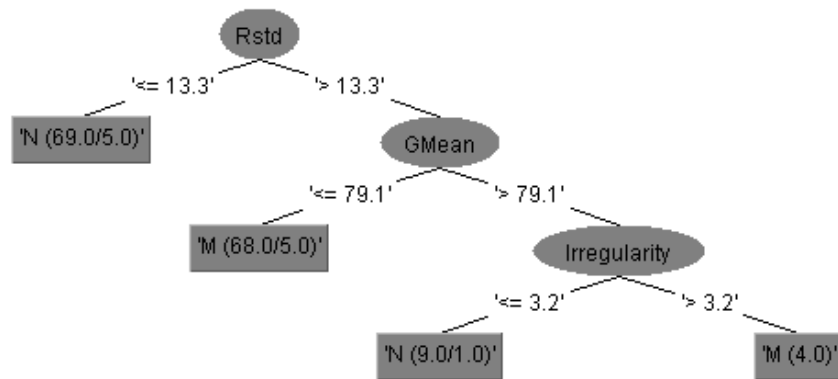
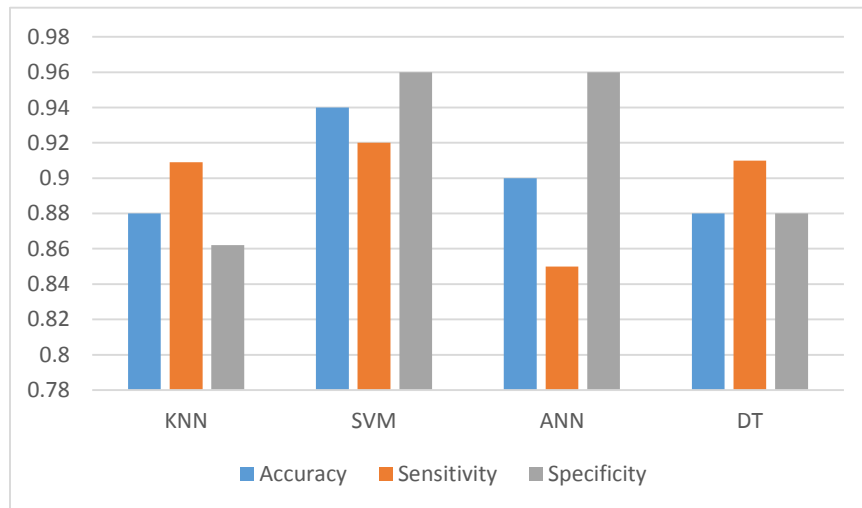


Figure 5-12: Decision Tree visualization

Figure 1-1 presents a comparison between the four classifiers in an attempt to perform a final assessment of the best results achieved.



	Accuracy	Sensitivity	Specificity
KNN	0.88	0.91	0.86
SVM	0.94	0.92	0.96
ANN	0.90	0.85	0.96
DT	0.88	0.91	0.88

Figure 5-13 Comparison between the four classifiers

Based on the previous graphs, the best results achieved are with the SVM classifier, and the results look convergent in the various classifiers. The accuracy with KNN, DT, ANN and SVM classifiers are 88%, 88%, 90% and 94%, respectively. The implementation phase of the system depends on both SVM and KNN classifiers, despite SVM classifier shows the best accuracy. The KNN classifier is adopted for the final system, because SVM classifier gave only a certain percentage of the risk of the disease, while it is also necessary to show similar cases to be more convincing to the user. KNN classifier used in the system to identify the nearest cases fast and easily.

Last but not least, the system is implemented using the OpenCV image processing library, and the Weka for android machine learning library as well as other methods and algorithms that were previously discussed in the research, the system was examined on a dataset of images to verify the results reliability.

Previous results are considered well compared to other related work in the same field, the success in image processing (ROI selection and features extraction) is 91%, and in the final classification 94% in the best case and 88% in the worst case, the proposed system have been implemented in different virtual and real mobile environments to test the accuracy, flexibility and usability.

6. Conclusion and Future Work

This chapter presents the main conclusion and the future work of this research. The future work is pointed out as opportunities for improving the system by handling obstacles faced in the current work.

In this work, we developed an automatic system for preliminary diagnosis of melanoma based on the well-proven commonly used ABCDE medical procedure. The proposed system combines the ABCDE method with the image capture and processing capabilities of smart phones to achieve fast, affordable, easily available and highly accurate melanoma diagnosis. The system includes multiple modules for handling various steps of the method such as: image capture, noise removal, lesion segmentation, features extraction and, finally, classification (diagnosis).

In the presented method, after capturing the image using the smart device's camera, the region of interest is automatically determined using automatic lesion detection and selection in real time. The next step is noise removal where the extracted region of interest is cleaned from hair and light reflection is removed. Next, the lesion contour is selected using two segmentation algorithms Otsu's and Mean shift segmentation. These two techniques were selected after experimenting with various segmentation alternatives in order to get the best and most accurate results. In the final steps of the method, the ABCDE melanoma features were extracted and the KNN classification method applied to reach a diagnosis. The proposed solution used 12 different Melanoma features.

The KNN classifier is selected after examining a number of other classifiers. Using the Weka data mining application, the following classifiers are tested KNN, SVM, ANN, and Decision Trees (DT). The test data set included 164 sample images. 14 images are not well captured and failed in image processing (lesion selection or features extraction), and, thus, could not be used in the classification step. The remaining 150 cases are successfully

processed and their ABCDE feature sets input to the classifiers. SVM scored highest accuracy of 94%, while KNN scored minimum accuracy of 88%.

The mobile application is built using the Eclipse environment with OpenCV and Weka APIs, and implemented the image processing methods and KNN classifier. The final application is tested on the same training dataset.

Having described our work above, a number of conclusions and comments should be made here. First, taking into consideration that the system performs its analysis in real time, and that it plays a role for only initial diagnosis (not final), the achieved performance level is more than adequate. The results are promising and successful, compared with the related work that used the same database. The results show that color features play a decisive role in the classification of skin lesions.

Second, the implemented system has been published as an android application on the Google App Store. This provides a good opportunity to examine the application further and get feedback comments and observations from the different users.

Third, during this research project we faced the following obstacles:

- The differences in the mobile camera properties and specifications, for example some smart phones does not support the auto-focusing and flashlight, which led to some problems with shadows and reflections.
- The large storage space required for OpenCV libraries in the application package when using Static initialization development approach. On the other hand, using of Async initialization development approach requires installing OpenCV libraries from the Google play store, which results in some problems in both the usability and permission of the application.
- Most images used in the features extraction stage are from the Internet without information about the patients, such as age, lesion location, family medical history

and the sex of the patient, which affects the classification accuracy. The future work will enhance the data to include all this information in the classification model.

- Lack of information about image size from the Internet, namely, scale and zoom, which affect the accuracy of the final classification process.

Based on the above obstacles we can suggest the following opportunities for future work:

- Improve the classification process by adding more patient details (age - sex - lesion position - family medical history) beside the features that have been extracted from the images.
- Tracking and evaluating the lesion size over time, for example, storing the lesion size during varying periods of time and compared the changes
- Make the system suitable to various mobile devices. That needs to study more characteristics of various types of mobile phones

References

- [1] "American academy of dermatology," [Online]. Available: <http://www.aad.org/media-resources/stats-and-facts/conditions/skin-cancer>. [Accessed 2014].
- [2] "Melanoma Skin Cancer," [Online]. Available: www.cancer.org/melanoma-skin-cancer-pdf. [Accessed 2014].
- [3] "http://www.cancer.gov/cancertopics/pdq/treatment/melanoma/HealthProfessional/page3," <http://www.cancer.gov/>, [Online]. Available: http://www.cancer.gov/cancertopics/pdq/treatment/melanoma/HealthProfessional/page3#figure_327_e. [Accessed 1 2015].
- [4] R. P. Braun, H. s. Rabinovitz and M. Oliviero, "Dermoscopy of pigmented skin lesions," *Journal American Academy of Dermatology*, vol. 52, pp. 109-121, 2005.
- [5] S. Menzies , C. Ingvar , K. Crotty and W. McCarthy, "Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features," *Arch Dermatol*, 1996.
- [6] J. Henning, . S. Dusza, S. Wang, D. Polsky, H. Rabinovitz , A. Kopf and . A. Marghoob, "The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy," *J. Am. Acad. Dermatol*, 2007.
- [7] O. Abuzagheh, B. Barkana and M. Faezipour, "SKINcure: A real time image analysis system to aid in the malignant melanoma prevention and early detection," *Image Analysis and Interpretation (SSIAI),IEEE*, pp. 85-88, 2014.
- [8] T. Wadhawan, N. Situ, K. Lancaster and G. Zouridakis, "SkinScan©: A portable library for melanoma detection on handheld devices," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 33-136, 2011.
- [9] K. Ramlakhan and S. Yi, "A Mobile Automated Skin Lesion Classification System," *23rd IEEE International Conference on Tools with Artificial intelligence (ICTAI)*, pp. 138-141, 2011.
- [10] A. Karargyris, O. Karargyris and A. Pantelopoulos, "DERMA/care: An advanced image-processing mobile application for monitoring skin cancer," *IEEE 24th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 1-7, 2012.
- [11] B. Joel A. Wolf, B. Jacqueline F. Moreau, M. Oleg Akilov, D. Timothy Patton, M. Joseph C. English III, M. Jonhan Ho and L. K. Ferris, "Diagnostic Inaccuracy of Smartphone Applications for Melanoma Detection," *JAMA Dermatol*, Vols. 149(4):422-426, 2013.

- [12] K. Lloyd and J. Lloyd , "The Lloyd Dermatology and Laser Center," [Online]. Available: <http://www.lloyd-derm.com/searchresults.php?search=images>. [Accessed 2014].
- [13] R. C. Gonzalez and R. E. Woods, Digital Image Processing, MA: Addison-Wesley, 1992.
- [14] K. N. Mody, "survey of image segmentation and classification using markov," *International Journal of Computer Science Engineering and Information Technology Research (IJCEITR)*, vol. Vol. 3, no. Issue 5, pp. 47-56, 2013.
- [15] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. vol. 9, no. no. 1, pp. 62-66, January 1979.
- [16] M. Pavani and S. Balaji, "An approach for segmentation of medical images using pillar K-means algorithm," *International Journal of Computer Trends and Technology (IJCTT)*, vol. volume4, no. Issue4, pp. Page 636-641, April 2013.
- [17] P. Meer and O. Comaniciu, "Mean Shift : A Robust approach towards feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. vol 24, no. No 5, May 2002.
- [18] A. Kaehler and G. Bradski, Learning OpenCV, O'Reilly Media, October 2013.
- [19] S. Suzuki and K. be, "Topological Structural Analysis of Digitized Binary Images by Border Following," *CVGIP, Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32-46, 1985.
- [20] "Wikipedia: Image moment," [Online]. Available: http://en.wikipedia.org/wiki/Image_moment. [Accessed 2014].
- [21] P. Schmid-Saugeon, J. Guillod and J.-P. Thiran, "Towards a computer-aided diagnosis system for pigmented skin lesions," *Computerized Medical Imaging and Graphics*, vol. 27, p. 65–78, 2003.
- [22] R. S. Montero and E. Bribiesca, "Compactness and Circularity Measures," *International Mathematical Forum*, 2009.
- [23] T. Maintz, in *Digital and Medical Image Processing*, 2005, p. 195.
- [24] I. Maglogiannis, S. Pavlopoulos and D. Koutsouris, "An Integrated Computer Supported Acquisition, Handling, and Characterization System for Pigmented, Skin Lesions in Dermatological Images," *IEEE Trans Inf Technol Biomed*, vol. 9, p. 13, 2005.
- [25] K. Ramlakhan and Y. Shang, "A Mobile Automated Skin Lesion Classification System," in *23rd IEEE International Conference on Tools with Artificial Intelligence*, 2011.
- [26] T. T. Godfried, "<http://cgm.cs.mcgill.ca/~beezer/cs507/3coins.html>," [Online]. Available: <http://cgm.cs.mcgill.ca/~beezer/cs507/3coins.html>. [Accessed 2014].

- [27] A. H. Bhuiyan, I. Azad and K. Uddin, "Image Processing for Skin Cancer Features Extraction," *International Journal of Scientific & Engineering Research*, vol. 4, p. 6, 2013.
- [28] Jiawei Han, Micheline Kamber and Jian Pei, *Data Mining: Concepts and Techniques*, Elsevier.
- [29] E. Alpaydin., *Introduction to Machine Learning*, Cambridge, London: MIT press, 2009.
- [30] H. Sahu, S. Shurma and S. Gondhalakar, "A Brief Overview on Data Mining Survey," *International Journal of Computer Technology and Electronics Engineering (IJCTEE)*, vol. 1, no. 3, Hemlata Sahu, Shalini Shurma, Seema Gondhalakar.
- [31] W. Xindong, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, p. 37, 4 December 2007.
- [32] A. Minin, "The Neural Network analysis," [Online]. Available: www5.in.tum.de/JASS/JASS06/presentations/Minin_handout.doc. [Accessed 2014].
- [33] "OpenCV: Computer vision library," [Online]. Available: <http://opencv.org/about.html>. [Accessed 2014].
- [34] "OpenCV for Android SDK," [Online]. Available: <http://opencv.org/platforms/android.html>. [Accessed 2014].
- [35] "ml4android: machine learning for android," [Online]. Available: <https://code.google.com/p/ml4android/>. [Accessed 2014].
- [36] "Weka for Android: Android library," [Online]. Available: <https://github.com/rjmarsan/Weka-for-Android>. [Accessed 2014].
- [37] "Weka : Data Mining Software in Java," [Online]. Available: <http://www.cs.waikato.ac.nz/ml/weka/>.
- [38] K. Restivo, R. Llamas and M. Shirer, "International Data Corporation (IDC)," <http://www.idc.com/getdoc.jsp?containerId=prUS24108913>, 16 5 2013. [Online]. Available: <http://www.idc.com/getdoc.jsp?containerId=prUS24108913>. [Accessed 5 9 2014].
- [39] I. Alhalabi, "Skin Doctor Application," [Online]. Available: <https://play.google.com/store/apps/details?id=skinDoctor.med.melanoma>. [Accessed 7 4 2014].
- [40] "Mobile weka: Data mining tool WEKA to mobile platform," [Online]. Available: <https://code.google.com/p/mobileweka/>. [Accessed 2014].
- [41] J. Figura, "Machine Learning for Google Android," *Institute of Formal and Applied Linguistics*, 2012.

- [42] Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski and Lukasz A. Kurgan, *Data Mining, A Knowledge Discovery Approach*, Springer, 2007.
- [43] M.-K. Hu, "Visual pattern recognition by moment invariants," *Information Theory*, *IRE Transactions on Information Theory*, vol. 9, no. 179-187, p. Feb, 1962.

Appendices

Appendix A: Data Analysis

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	RMean	GMean	BMean	Rstd	Gstd	Bstd	Xsymmetr	Ysymmetr	HCR	Roundnes	minEnclos	Irregularity	Classification	
2	70.9	42.1	39.4	26.6	17.3	14.8	0.34	0.48	0.898	0.66	0.4	2.99	M	
3	58.9	53.5	54.7	24.4	23.9	24.7	0.33	0.4	0.82	0.55	0.46	3.1	M	
4	80.2	52.2	38.8	16.5	13	11.1	0.49	0.29	0.92	0.72	0.41	2.87	M	
5	59.5	48.6	48.1	22.4	19.3	19.8	0.35	0.19	0.92	0.71	0.48	2.7	M	
6	97.2	69.4	76.6	16.7	12.1	8.6	0.64	0.76	0.81	0.51	0.39	3.4	M	
7	89	60.2	61.7	22.21	20.88	21.71	0.437	0.397	0.78	0.47	0.52	3.17	M	
8	88	65.3	60.6	17.3	14.6	14.5	0.31	0.34	0.78	0.52	0.3	3.64	M	
9	103.9	38.4	24.8	19.8	12.7	10.2	0.272	0.762	0.875	0.55	0.55	2.85	M	
10	95.4	57.4	31.6	20.5	18.4	16.9	0.83	1.82	0.783	0.33	0.65	3.25	M	
11	94.1	59.4	45.4	22.6	18.9	15.1	0.64	0.37	0.89	0.96	0.33	3.09	M	
12	92.1	74.5	81.4	36.04	31.3	38.3	0.42	0.41	0.78	0.45	0.55	3.16	M	
13	101.5	68.6	41.5	35.3	21.4	16.7	0.42	0.29	0.93	0.74	0.43	2.77	M	
14	126.4	99.1	60.7	37.7	29.6	31.6	0.36	0.29	0.85	0.59	0.47	2.98	M	
15	127.3	83.4	52.9	33.2	20.9	23.9	0.24	0.21	0.87	0.67	0.27	3.27	M	
16	143.2	77.8	44.5	42.8	21.2	16.6	0.31	0.35	0.93	0.77	0.35	2.89	M	
17	109	67.8	32.2	15.8	13.7	13.2	0.5	0.37	0.87	0.55	0.56	2.82	M	
18	108.6	73.3	24.7	23.5	20.1	13.9	0.499	0.996	0.73	0.34	0.54	3.67	M	
19	84.9	56.9	25.9	25.9	17.4	14.4	0.41	0.47	0.83	0.49	0.53	3.08	M	
20	139.2	73.1	30.3	18	17.3	15.3	0.66	0.73	0.75	0.4	0.43	3.71	M	
21	116.6	79.1	60.3	13.6	14	13.9	0.48	0.52	0.83	0.57	0.42	3.18	M	
22	119.9	70.7	60	8.1	8.5	7.6	0.36	0.67	0.93	0.7	0.48	2.7	M	
23	43.3	33.5	33.2	17.4	9.72	8.5	0.193	0.48	0.92	0.72	0.41	2.8	M	

Figure 0-1 the 12 Features calculated from the Lesion ROIs

Table 0-1 the extracted features description

	The feature	Description
1	RMean	The Mean for the Red color in the RGB image mode
2	GMean	Mean Green for the color in the RGB image mode
3	BMean	Mean Blue for the color in the RGB image mode
4	Rstd	The stander deviation for the Red color in the RGB image mode
5	Gstd	The stander deviation for the Green color in the RGB image mode
6	Bstd	The stander deviation for the Blue color in the RGB image mode
7	Xsymmetry	The x-axis symmetry , the nonoverlapping area around X axis
8	Ysymmetry	The y-axis symmetry , the nonoverlapping area around Y axis
9	HCR	Hull convex ratio
10	Roundness	The lesion roundness
11	minEnclosingCircle	The lesion minimum enclosing circle
12	Irregularity	The lesion border Irregularity

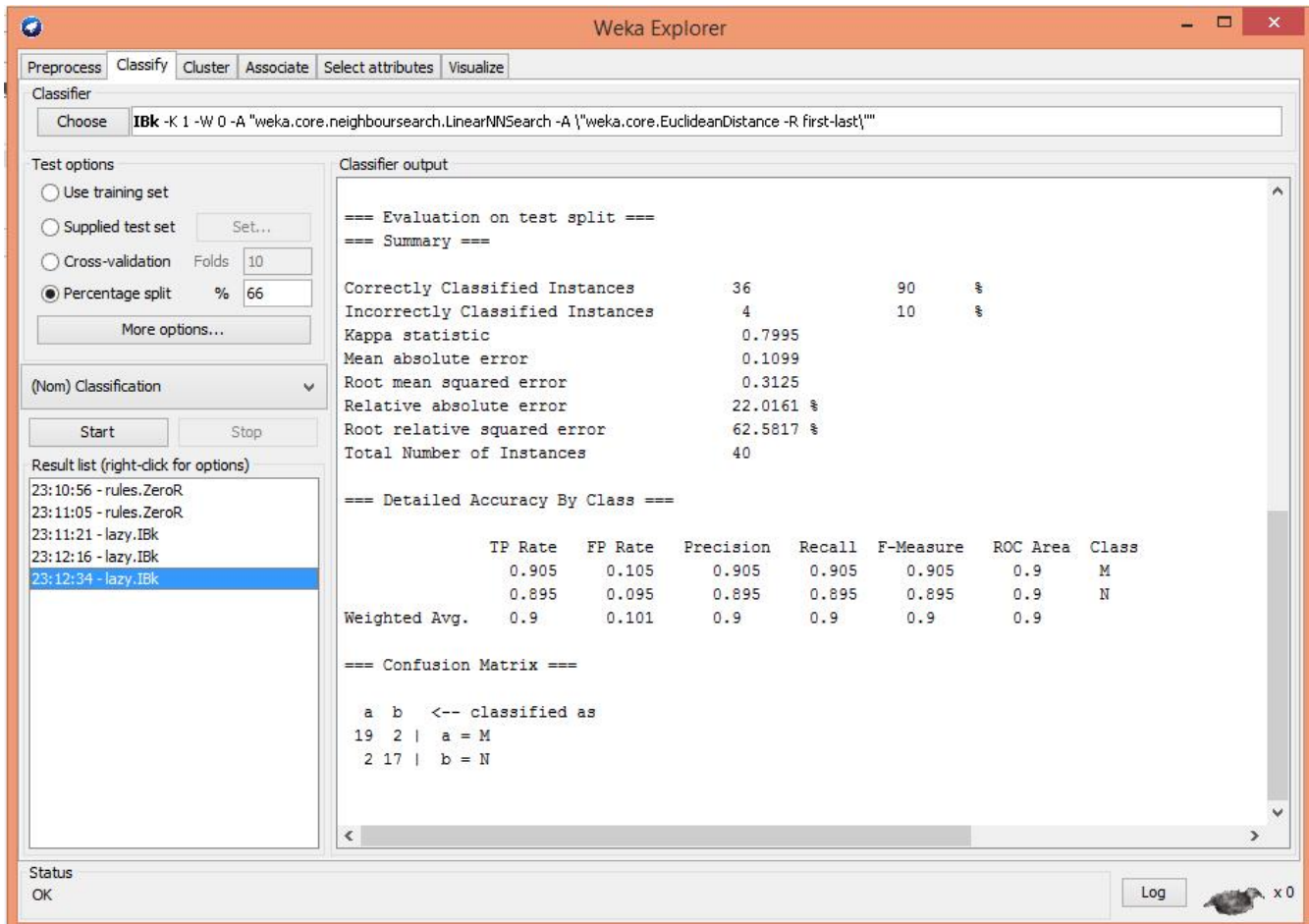


Figure 0-2 Example of classification results using ANN classifier

Appendix B: Application screenshot

The application is shown in Figure 0-3, and also available on the Google store [39] .

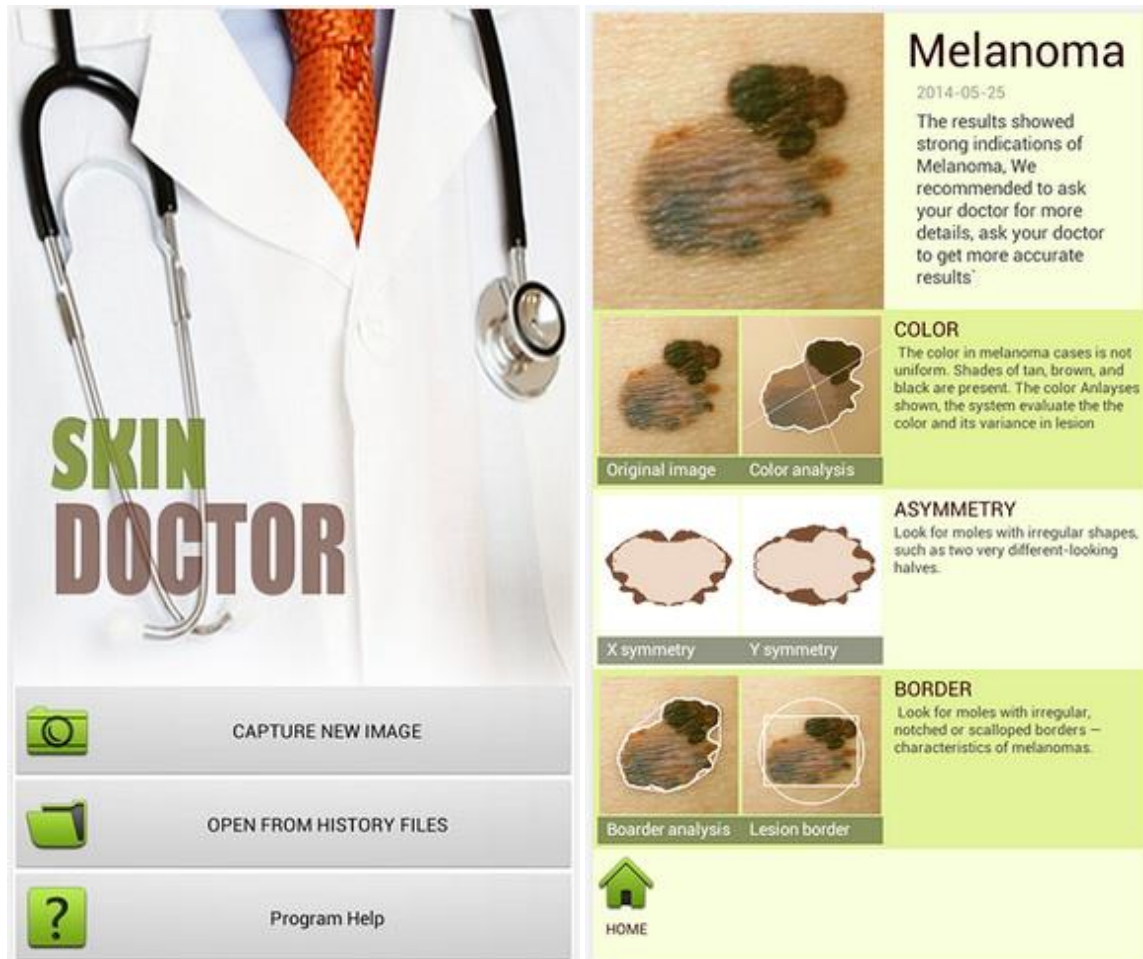


Figure 0-3: The application screenshot

First image show the application main interface and the second show the classification results