Ministry of Higher Education and
Scientific Research
University of Diyala
College of Science
Department of Computer Science

Deep Learning Method for Classification of Skin Cancer Disease

A Dissertation
Submitted to the Department of Computer Science\ College of Sciences\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science

By

Ohood Fadhil Alwan

Supervised By

Prof. Dr. Dhahir Abdulhadi Abdullah

2020

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

﴿يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنكُمْ
وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ﴾

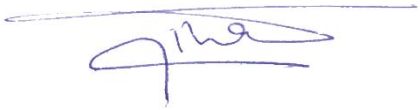
صدق الله العظيم

(سورة المجادلة :الآية11)

(Supervisor's Certification)

We certify that this research entitled "***Deep learning methods detection of disease from images*** was prepared by ***Ohood Fadhil Alwan*** Under our supervisions at the University of Diyala Faculty of Science Department of Computer Science, as a partial fulfillment of the requirement needed to award the degree of Master of Science in Computer Science.

(Supervisor)

Signature: 

Name: **Dr. Dhahir Abdulhadi Abdulah**

Date:2/7/2020

Approved by University of Diyala Faculty of Science Department of Computer Science.

Signature:

Name: **Dr. Taha Mohammad Hassan**

Date:

(Head of Computer Science Department)

Dedication

I would like to dedicate this work to:

To my father and mother, may God have mercy on them

To My husband Nashwan

For his unlimited love, support,

endurance and encouragement

To my candle, my children

Ahmed and Amina.

To my brothers and sisters

To everyone who helped me from a friend or fellow…

Acknowledgements

My thanks are above all to God Almighty, who guided my steps towards the path of knowledge and without his help and blessing. This thesis would not be advanced or you will see the light.

I express my sincere appreciation to my supervisors ***Prof. Dr. Dahir Abdul Hadi Abdullah*** *for providing his ideas, inspiration and continuous support for me during my studies.*

I am very grateful to all members and professors of the Computer Science Department at Diyala University for their general support, especially ***Lecturer Hazem Noman, Adel Abed****,*

Finally, I would have never been able to finish my message without the help of friends ***Zaidoon, Mohammad Flaih and Mustafa Flaih****. and especially mention my friends (****Noor Hassan, Shaimaa Salem and Zeinab Yassin****) and support my family and my husband.*

Thank you all!

Ohood

Abstract

Skin cancer is an abnormality in skin cells caused by mutations in cells Deoxyribonucleic Acid (DNA). Most deaths from skin cancer are caused by the malignant type. Therefore, one of the last types of cancer is considered a treatment that can detect the disease early by biopsy examining, so the best solution for improving the diagnosis of skin cancer is early detection. Computer-Aided Diagnosis (CAD) is one of the widely used imaging techniques for detection and classification of skin cancer. The automatic detection and classification of image is considered very important for tumors skin and very challenging task for medical images. This thesis presents a proposed system for classification of skin cancer after its detection with the help of deep learning mechanisms and machine learning algorithms, where several steps are used in the form of stages, which are include, the image acquisition stage, image pre-processing, and the classification stage. The used dataset is obtained from the ISIC (International Skin Image Collaboration) Archive, it contains 3297 images. There are 1497 image cases of malignant skin caser type, and 1800 images cases for benign. In preprocessing stage, hair removal algorithm is using. The First proposed model depends on Convolutional Neural Networks (CNN) classifier. The second proposed model uses Naïve Bayes (NB) classifier. While the third proposed model relays on Support Vector Machine (SVM) classifier. And each model with applying preprocessing algorithm and without applying. The results show that the first proposed model using (CNN) without preprocessing had average accuracy 85.00%, while with preprocessing had accuracy 69.99%. The second proposed model using (NB) without preprocessing had average accuracy 70.15%, while with preprocessing had accuracy 69.69%. The third proposed model using (SVM) without preprocessing had Achieve accuracy 76.81%, while with preprocessing had accuracy 77.12 %.

List of Contents

Subject	Page No.
List of Contents	I
List of Abbreviations	IV
List of Tables	V
List of Figures	VI
Chapter One: General Introduction	
1.1 Introduction	1
1.2 Skin cancer diagnoses	2
1.3 Overview of Deep learning Techniques	4
1.4 Related Works	5
1.5 Problem Statement	9
1.6 Aims of the Thesis	9
1.7 The Organization of the Study	10
Chapter Two: Theoretical Background	
2.1 Introduction	11
2.2 Image Preprocessing	11
2.2.1 Hair Removal	12
2.2.2 Image Enhancement	14
2.3 Artificial Neural Network (ANN)	16
2.3.1 Neural Network Model	16
2.4 Convolutional Neural Networks (CNN)	18
2.4.1 Basic Structure of CNN	19
2.4.2 Training a Network	26
2.4.3 Back Propagation Algorithm	27
2.5 Naïve Bayes Algorithm	30
2.6 Support Vector Machine	32

2.6.1 Linear SVM	*33*
2.6.2 Non-Linear SVM	*34*
2.7 Evaluation Measures	*35*
Chapter Three: Proposed System Design	
3.1 Introduction	*37*
3.2 The Proposed System Design	*37*
3.3 Image Acquisition Stage	*39*
3.4 Image Preprocessing Stage	*39*
3.4.1 Hair Removal Algorithm	*40*
3.4.2 Normalization for Image	*43*
3.5 Classification Stage Using First Model (CNN) Algorithm	*43*
3.5.1 Design Convolution Neural network (CNN) Structure	*46*
3.5.2 CNN Training	*52*
3.5.3 CNN Testing	*56*
3.6 Classification Using Naïve Bayes Algorithm	*51*
3.6.1 NB Training	*58*
3.6.2 NB Testing	*59*
3.7 Classification Stage Using (SVM) Algorithm	*61*
3.7.1 SVM Training	*63*
3.7.2 SVM Testing	*64*
Chapter Four: Experiments Rustles and Discussion	
4.1 Introduction	*66*
4.2 Implementation Environment	*66*
4.3 Evaluation of Skin Cancer Systems	*67*
4.4 Dataset Acquisition (Skin Cancer Images)	*68*
4.5 Result of Image Pre-processing	*70*
4.5.1 Morphological Close Operation Results	*71*
4.5.2 Median filter Results	*71*
4.6 Evaluation of First Proposed System	*72*
4.7 Results of Second Proposed Model (NB)	*87*
4.8 Results of Third Proposed Model (SVM)	*92*

4.9 Result Analysis for the Proposed Models	*97*
4.6 Comparison to the Related Works	*99*
4.1 Introduction	*66*
4.2 Implementation Environment	*66*
Chapter Five: Conclusions and Suggestions for Future Work	
5.1 Conclusions	*100*
5.2 Suggestions for Future Work	*101*
References	
References	*102*

List of Abbreviations

Abbreviations	*Description*
ABCD	Asymmetry, Border, Color, Diameter
AC	Accuracy
AI	Artificial Intelligence
ANN	Artificial Neural Network
CAD	Computer-aided diagnosis
CNN	Convolutional Neural Network
Conv2D	Convolutional Tow – Dimension
DNA	Deoxyribonucleic Acid
ES	Error Signals
FN	False Negative
FP	False Positive
FS	Function Signal
IM	Input Image
ISIC	International Collaboration Skin Imaging
ISBI	International Symposium on Biomedical Imaging
ML	Machine Learning
MAP	Maximum A Posteriori
MLP	Multi-Layer Perception
NB	Naïve Bayes
NBC	Naïve Bayes Classification
NNs	Neural Networks
RBF	Radial Basis Function
ReLU	Rectifier Linear Unit
SVM	Support Vector Machine
SE	Structuring Element

List of Tables

Table No.	*Description*	*Page*
Table (2.1)	Type of Activation Function	22
Table (4.1)	Confusion Matrix	67
Table (4.2)	Distribution of Number Skin Cancer Images Dataset	68
Table (4.3)	Proposed Design of CNN Layers (conv2D, Pooling, Full connected)	73
Table (4.4)	Accuracy and loss for each training in 10-Epoch(without)	76
Table (4.5)	Accuracy and loss for each training in 30-Epoch(without)	77
Table (4.6)	Accuracy and loss for each training in 10-Epoch(with)	80
Table (4.7)	Accuracy and loss for each training in 30-Epoch(with)	82
Table (4.8)	Difference between CNN with and without processing	84
Table (4.9)	Naïve Bayes Accuracy without preprocessing	89
Table (4.10)	Naïve Bayes Accuracy with pre processing	91
Table (4.11)	Difference between NB with & without processing	91
Table (4.12)	SVM Accuracy without pre processing	93
Table (4.13)	SVM Accuracy with pre processing	95
Table (4.14)	Difference between SVM with and without processing	95
Table (4.15)	Average performance measure of the proposed models, CNN, Naïve Bayes, SVM and without preprocessing	97
Table (4.16)	Average performance measure of the proposed models, CNN, Naïve Bayes, SVM and with preprocessing	97
Table (4.17)	Comparison of classification accuracy with earlier studies	99

List of Figures

Figure No.	Description	Page
Figure (1.1)	Unaffected skin and Affected skin	1
Figure (1.2)	Skin cancer Incidence and Death.	2
Figure (2.1)	Skin Cancer with Hair Effects	12
Figure (2.2)	Flowchart of the Opening and Closing Processes	13
Figure (2.3)	Example of Closing Process	14
Figure (2.4)	Structure Element with Disk Shape	14
Figure (2.5)	Example of Median Filter	15
Figure (2.6)	Biological Neuron	16
Figure (2.7)	Neural Network Model.	17
Figure (2.8)	The General Structure of CNN System	18
Figure (2.9)	Convolutional Layer	20
Figure (2.10)	An example of convolution operation	21
Figure (2.11)	An Example of ReLU Transformation	23
Figure (2.12)	Two Classic Pooling Methods	23
Figure (2.13)	Max 3 × s3 Pooling Layer to Minimize the Spatial Size Image	24
Figure (2.14)	Schematic representation of an MLP	25
Figure (2.15)	Dropout Neural Network	27
Figure (2.16)	Illustration of Directions of Signal Flows	28
Figure (2.17)	The SVM hyperplane between two classes	33
Figure (3.1)	Block Illustration of General Proposed System	38
Figure (3.2)	Samples of Different Skin Cancer Images	39
Figure (3.3)	Block Diagram of the First Proposed Model	44
Figure (3.4)	Structure of the CNN algorithm	46
Figure (3.5)	The input layer (input image).	47
Figure (3.6)	Max pooling layer	49
Figure (3.7)	Block Diagram of the of Second Model Naïve Bayes	58
Figure (3.8)	Naïve Bayes Parameters during the training process	59
Figure (3.9)	Block Diagram of the of Support Vector Machine	62
Figure (3.10)	SVM Parameters during the training process	64

Figure (4.1)	Benign Skin Cancer Images	69
Figure (4.2)	Malignant Skin Cancer Images.	70
Figure (4.3)	Original image before pre processing	71
Figure (4.4)	Image after pre processing	72
Figure (4.5)	Accuracy and Loss Validation Change Against Training Epochs using CNN Model(10-Epoch) without preprocessing.	75
Figure (4.6)	Accuracy and Loss Validation Change Against Training Epochs using CNN Model(30-Epoch) without preprocessing	76
Figure (4.7)	confusion matrix for the CNN training without Preprocessing CNN training. Left: In 10- Epoch; Right: In 30- Epoch.	78
Figure (4.8)	Accuracy and Loss Validation Change Against Training Epochs using CNN Model(10-Epoch) with preprocessing	80
Figure (4.9)	Accuracy and Loss Validation Change Against Training Epochs using CNN Model(30-Epoch) with preprocessing	81
Figure (4.10)	confusion matrix for the CNN training with Preprocessing CNN training. Left: In 10- Epoch; Right: In 30- Epoch	83
Figure (4.11)	Result of Conv2D and Pool 2D Layers in CNN	86
Figure (4.12)	Skin Cancer Images with Labels	87
Figure (4.13)	The Confusion Matrix for Naïve Bayes without preprocessing	89
Figure (4.14)	The Confusion Matrix for Naïve Bayes with preprocessing	90
Figure (4.15)	Scatter Plot for Dataset	92
Figure (4.16)	The Confusion Matrix for SVM without preprocessing	93
Figure (4.17)	The Confusion Matrix for SVM with preprocessing	94
Figure (4.18)	Number of Support Vectors	96
Figure (4.19)	Illustration of the average accuracies of the proposed models.	97

Chapter One

Chapter One

General Introduction

1.1 Introduction

The skin is a vital organ that covers the entire outside of the body, forming a protective barrier against pathogens and injuries from the environment. But because it is located on the outer part, the skin is prone to disease. One of these diseases is known as skin cancer. Skin cancer is an abnormality in skin cells caused by mutations in cells Deoxyribonucleic Acid (DNA). One of the most dangerous types of skin cancer is melanoma cancer. It is a skin malignancy derived from melanocyte cells; the skin pigment cells that produces melanin. Because these cells are still able to form melanin, melanoma is mostly brown or black colored [1].

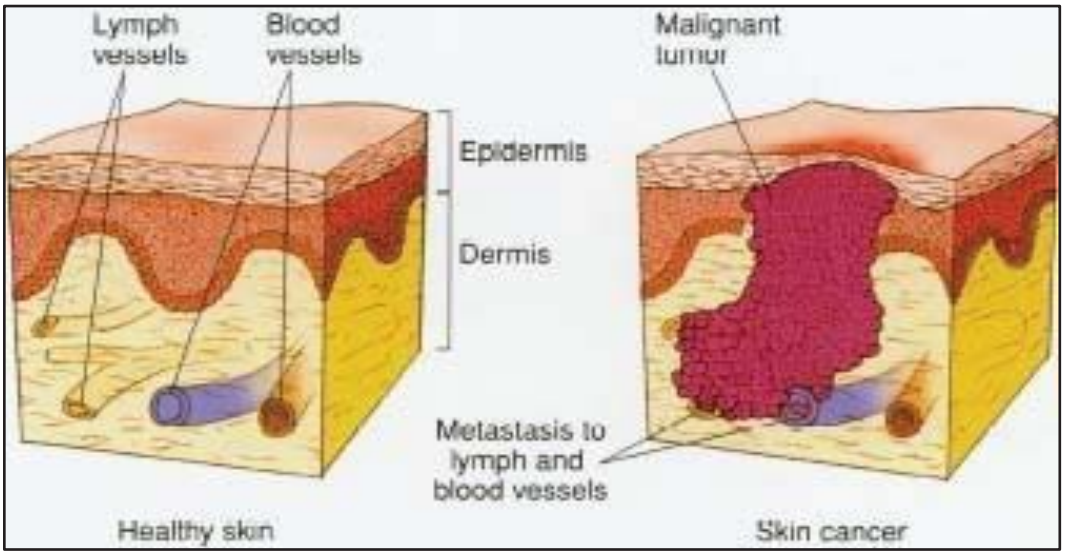


Figure (1.1): (left) Unaffected skin, (right) Affected skin [1].

More than 5,400 people worldwide die every month from malignant skin cancer, and estimates and statistics indicate that the number of new cases of melanoma cancer diagnosed in 2020 will increase by about 2%. The number of skin cancer deaths is expected to decrease by 5.3%. in 2020. Of these, 60,190 cases will be men and 40,160 cases will be women. In the past decade (2010-2020), the number of new diagnostic melanoma cases diagnosed annually increased by 47 % [**2**]. Skin cancer affects the men and the women at different ages, as shown in the figure.

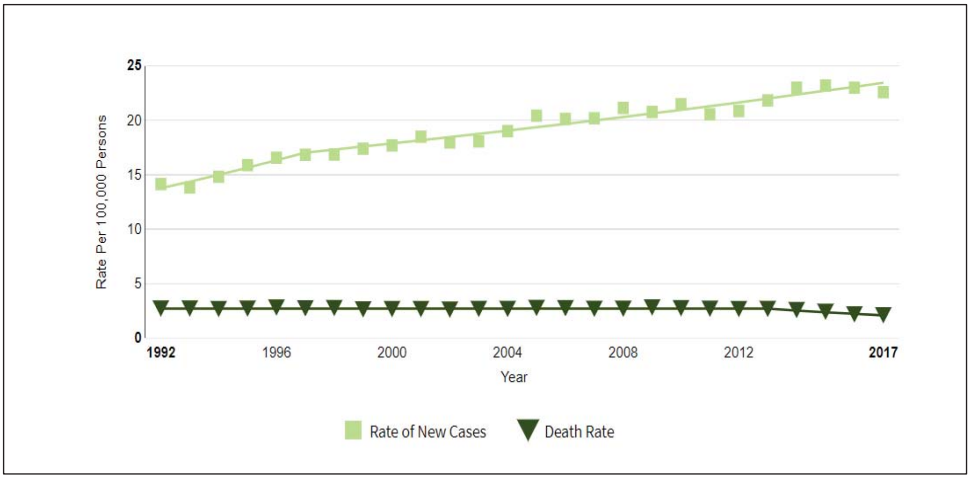


Figure (1.2): Skin Cancer Incidence and Death [2].

1.2 Skin Cancer Diagnoses

1.2.1 Traditionally Diagnoses

It is difficult to distinguish between types of skin cancer (melanoma and benign moles) at the beginning of its appearance is, even for experienced doctors **[3]**. The use of traditional methods to diagnose the disease by physical examination and biopsy. The biopsy is removed part or all of this spot and sent to the laboratory and the results may take a week to come through. This physical diagnosis is expensive, time-consuming, and may produce the wrong result for some reason. Therefore, sophisticated equipment and algorithms are required to

assist decision makers. Various methods in dermatology such as "ABCD" (Asymmetry, Border irregularity, Color patterns, and Diameter) rule and the seven - points checklist **[4].**

1.2.2 Computer-Aided Diagnosis (CAD)

The concept of using computer vision to solve the task of identifying skin cancers arose recently. Automated pigmented lesion analysis has become an important research topic trying to improve or develop the diagnosis of computer-assisted skin cancer [**5**]. Medical image processing is an area of proven expansion and an interdisciplinary field of research and interest and various fields, computer science, engineering, applied mathematics, statistics, physics, medicine, and biology. Computer-assisted diagnostic treatment has occupied a remarkable space in the clinical routine and with the recent advances in high technology and the introduction of different methods and techniques leads to more challenge in the mechanism of dealing with the huge number of images. It provides a high-quality information that helps in diagnosing the disease [**6**]. The introduction of artificial intelligence methods as a method that helps doctors in diagnosing has become an increasing trend in dermatology. These methods generally utilize some procedure of machine learning (ML), which is a branch of Artificial Intelligence (AI) including approaches that enable machines to make the predictions based on their prior information and experiences [**7**].

1.3 Overview of Deep Learning Techniques

Machine and deep learning methods performance an important role to train computer systems as a professional prediction and decision making could be used. Machine learning is the field of study that give the computers the ability to learn without the need for complicated programs. Deep learning is one of the ways that gives the ability to understand the world, by arranging ideas and bringing intelligence to the computer, as extracting patterns and processing them becomes easy and it is one of the branches of machine learning **[8]**.

Automatic learning techniques that applied to images directly are not well efficient because they neglect or ignore the structure and composition of the image. Therefore, a deep learning solution is the place of automatic learning in many of the image processing tasks because it has the advantage of extracting features, which is part of the learning process **[9]**.

The request of computers to identify several features that can distinguish between the required data is under the idea of the basis of the work of many deep learning methods as it lies in the transfer of the image between different layers to give the result of a specific disease. These models or methods are used in processing big data to reach a size Interest required. The convolutional neural network is considered one of the most important models of deep learning in the field of image classification, that outperforms many automated machine learning algorithms **[10]**.

1.4 Related Works

In this section the study reviews some of various styles and techniques that can used for detection skin cancer are presented:

- **Park, D. C. 2016 [11]** proposed a model to explain how the naive Bayes Classifier could classify image of skin cancer and can be formed as the maximum posteriori of decision-making rule. The researcher relied on taking the advantage of concepts of the Naive Bayes probability classifier, in order to reduce the training time for the algorithm. The number of images in the dataset 800 images divided into four class and each class contains 200 images. The proposed classifier reached accuracy 77.2 %.
- **Shoieb, D. et al 2016 [12]** Presented model for diagnosing the skin cancer by applying deep learning approaches. Enhanced segmentation is a stage the model applies to identify malignant skin cancer while the researcher used a network (CNN) to extract features from the images. The model was built on a multi-layered linear with SVM that was trained by features extracted from a(CNN) network. Despite the experimental results obtained by the system with accuracy 94%, but dataset that obtained from normal camera it faced additional effort in the pre-processing stages. Whereas, the total dataset 337 image 80% for training and 20% for testing.
- **Nasr-Esfahan. et al 2016 [13],** Applied a two-layer CNN was trained for the distinction of melanoma against benign nevi) built on clinical pictures. Only (136) images from dataset were used to train the model and the test dataset contained 34 images. The images were all from the public image archive of the “Department of Dermatology”. The proposed method after preprocessing stage and tested model achieve accuracy of 81%,

sensitivity of 81%, and a specificity of 80%. The tested images were very limited. However, the result can be improved when increased it.

- **Mustafa, S. et al 2017 [14]** Suggested a system which can make distinction between the skin lesions using machine learning techniques such as SVM model and with use ABCD rule where he used color space by experimenting with luminance to increase the visualization for Grab Cut segmentation of image, dataset that used 200 images 100 as benign and 100 as malignant .The algorithm can discover the optimum line to separate the two classes with accuracy 80% but with low sensitivity 71% and specificity 55%. The proposed method faces a problem with small dataset for training algorithm.

- **Codella, N. C. et al 2017 [15]** Designed a system based on machine and deep learning techniques to detect and classify skin lesions (benign and malignant) using dataset released by the International Skin Imaging Collaboration (ISIC) for the 2016 International Symposium on Biomedical Imaging (ISBI 2016), where dataset was splatted into 900 for training and 379 for testing, and the researcher relied on multiple models of deep learning deep fully convolutional-Net architecture residual networks, convolutional neural networks, segmentation using to extracting features with the help of machine learning algorithms. Proposed model achieved classification accuracy76%.

- **Lopez, A. et al 2017 [16]** Focus on the classification of skin cancer (benign or malignant), how to detect it early, and introduce deep learning methods to solve these problems. The researcher used the convolutional neural network model with VGG structure where that transfer learning model was used. The proposed method was tested on the "International

Symposium- on Biomedical Imaging" (ISBI) 2016 data set (346 for training and 150 for testing). The model obtained a 78.66% classification rate.

- **Md Ashraful 2018 [17]** Selected four approaches of the convolutional neural network such as SENet154, PNASNet-5-Large, InceptionResNetV2, InceptionV4to test the model in the classification of skin cancer images obtained from the (ISIC) 2018 Challenge data set. It contains more than 10015 pictures, after the pre-processing stage and testing the models, the results showed an accurate classification to PNASNet-5-Large with 76%, But it faced the problem of unbalanced data with a big change in all images make it difficult to generalize these features of skin lesions.

- **Mohan, K. et al. 2019 [18]** Classification of skin cancer was discussed using naive Bayes classifier with shearlet transformation factors with three coefficients. Treated melanoma images for rank feature then applied naive bayes for classification. The results showed that the system achieved accuracy of 90% at levels 3 through 100. By using the PH^2 dataset contains 100 dermoscopic RGB images with melanocytic lesions with resolution is768x560 pixels. The researcher also explained that when applied shear let transformation on images and with other coefficients increases the complexity of the calculations and the required more time.
- **Sanket .K. Chandra J 2019 [19]** The researcher suggested a model to classify skin cancer in three ways, including SVM, KNN, Ensemble, used in preprocessing stage a hybrid method that are starts with Wiener filter to remove noise the unwanted regions and then applied median filter to remove hair and after that used watershed algorithm with morphological operation for segmentation to extract features from the images. The researcher relied on two types of data PH^2 contains of 200 images used for

training and testing phase and ISIC dataset contains more than 30000 images of several types of cancer. SVM model has better performance with accuracy 92% when compared to other methods like KNN, Ensemble algorithms. But he did not mention the time of the algorithm in training

- **Refianti R. et al 2019 [20]** Design a skin cancer image classification system to examine endoscopy. Convolutional neural network (CNN) with LeNet-5 geometry was used as a proposed method for the system in the classification of image data, as the number of testing data reached to 44 images. A classification accuracy of 93% for training and 100% for testing, which is a percentage that the model might expect in overfitting. Because of the small number of datasets, which consists only 176 images in the 100 epochs.
- **Albahar, M. A. 2019 [21]** Relied in proposed model on a technique that depends on the engineering of the convolutional network. The network is consisting of two convolution layers followed by one of max pooling layer with dropout layer to treat the overfitting, then the fully connection layer as it contains 128 neurons. The idea in this model is to include a regulator on each convolution layer to control the values of weights, which is a matrix of filter applied to each input, the model training 5600 images, but the proposed system faced problem of choosing an appropriate λ value that is difficult, because it is a continuous value and several attempts to select it are costly and takes time.

1.5 Problem Statement

Skin cancer is a disease that requires early detection to determine it, whether it is benign or malignant. Using neural networks has shown outstanding results, with high flexibility in different environmental conditions, but its limitations in image classification processes have led us to use deep learning and machine methods to solve this problem as it has achieved impressive results in the field of medical image classification because early detection leads to rapid treatment.

1.6 Aim of the Thesis

The main aim of this thesis is to design a system to detect and classify skin cancer with different models, with applied proposed systems, the doctor can train the system on some known data and then apply this method to classify skin cancer. These models are:

1- Design and implement such a powerful structure by using Convolutional Neural Network (CNN) structure as the first classification approach.
2- The second model is Naïve Bayes (NB) that used for classification skin cancer for benign or malignant.
3- Support Vector Machine (SVM) is the third approach for classification approach Skin cancer of benign or malignant.

The objective of utilizing more meaningful information to improve skin cancer detection and help doctors and physicians in the clinical diagnosis with accurate detection of disease and giving reliability in decision-making and rapid detection of skin cancer.

1.7 The Organization of the Study

This thesis consists four chapters in addition to chapter one that was alredy discussed here and it is organized as follows:

Chapter Two describes the pattern classification system, design medical images analysis system, the concept of skin cancer with its types and overview of the method used to analysis and categorize skin images with their characteristic.

Chapter Three presents the details of the proposed detection and classification algorithms that are used to design the proposed system and the implementation of each one.

Chapter Four gives the experimental results obtained from the implementation of proposed system.

Chapter Five discusses results, conclusions and lists a number of suggestions for future studies.