

SKIN CANCER DETECTION USING CNN

A Project Report Phase - II

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering Degree in Electronics and Communication Engineering

by

NAGAPRASANTH.B (39130306)

NANDHAKUMAR.B (39130317)



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
SCHOOL OF ELECTRICAL AND ELECTRONICS**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI – 600119

April 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY (DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report submitted for final year project is the Bonafide work of **NAGAPRASANTH.B (39130306)** and **NANDHAKUMAR B (39130317)** who carried out the project entitled "**SKIN CANCER DETECTION USING CNN**" under our supervision from NOVEMBER 2022 to APRIL 2023.

Internal Guide

Dr.S. LAKSHMI, M.E, Ph.D.

Head of the Department

Dr. T. RAVI M.E., Ph.D.,

Submitted for Viva voce Examination held on 24.04.2023

Internal Examiner

External Examiner

DECLARATION

We, **NAGAPRASANTH.B (39130306)** and **NANDHAKUMAR B (39130317)** hereby declare that the Project Report on “**SKIN CANCER DETECTION USING CNN**” done by us under the guidance of **DR. S. LAKSHMI**, at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Electronics and Communication Engineering.

DATE: 24/04/2023

PLACE: Chennai

SIGNATURE OF CANDIDATES

1.



2.



ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our thanks to **Dr. N. M. NANDHITHA, M.E., Ph.D. Dean**, School of Electronics, **Dr. T. RAVI, M.E., Ph.D.**, Head of the Department of Electronics and Communication Engineering for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide **Dr.S. LAKSHMI. M.E., Ph.D., Department of Electronics and Communication Engineering**, for her valuable guidance, suggestions, and constant encouragement for the successful completion of our project work.

We wish to express our thanks to all Teaching and Non-teaching staff members of the **Department of Electronics and Communication Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Skin cancer is a type of cancer that grows in skin tissue and can damage surrounding tissue, leading to disability or even death. Skin cancer is the third most common cancer in Indonesia, after cervical cancer and breast cancer. Accurate diagnosis and early, appropriate treatment can minimize and control the harmful effects of skin cancer. Diagnosis of skin cancer and benign tumor lesions is time consuming because of their similar lesion morphology. The system developed in this study can automatically detect skin cancer and benign tumor lesions using convolutional neural networks (CNN). The proposed model consists of three hidden layers with 16, 32, and 64 output channels in each layer, respectively. The proposed model uses several optimizers such as SGD, RMSprop, Adam and Naadam with a learning rate of 0.001. Adam Optimizer has the best performance with an accuracy value of 99% in identifying skin lesions from his ISIC dataset in his four classes of dermatofibroma, nevus pigmentosa, squamous cell carcinoma and melanoma. will demonstrate. The results obtained exceed the performance of existing skin cancer classification systems

TABLE OF CONTENTS

CHAPTER NO;	LIST OF CONTENTS	PAGENO;
	ABSTRACT	v
	TABLE OF CONTENTS	vi
	LIST OF FIGURES	viii
	LIST OF TABLES	ix
	LIST OF ABBREVIATIONS	x
1	INTRODUCTION	
	1.1 INTRODUCTION	1
	1.2 PROBLEM STATEMENT	2
	1.3 OBJECTIVE	3
	1.4 MOTIVATION	3
	1.5 METHODOLOGY	4
	1.6 SUMMARY	4
2	LITERATURE SURVEY	5
	2.1 INTRODUCTION	5
	2.2 LITERATURE REVIEW	5
	2.3 SUMMARY	11
3	EXISTING METHODOLOGY	12
	3.1 INTRODUCTION	12
	3.2 IMAGE ACQUISITION	12
	3.3 IMAGE PRE-PROCESSING	12
	3.4 IMAGE SEGMENTATION	13
	3.5 FEATURE EXTRACTION	13
	3.6 DATA MINING UNIT	16
	3.7 SUPPORT VECTOR MACHINE	17
	3.8 CONVOLUTION NEURAL NETWORK	17
	3.9 SUMMARY	19

4	PROPOSED SYSTEM	20
	4.1 INTRODUCTION	20
	4.1.1 Images Pre-processing	22
	4.1.1.1 Resizing the Image	22
	4.1.1.2. Gray Images	22
	4.1.1.3 Powered Images	23
	4.1.2 Bag of Features Modal	24
	4.1.2.1 Step One: Feature Extraction	25
	4.1.2.2 Step Two: Quantization	25
	4.2 SYSTEM DESIGN	26
	4.2.1.1 MATLAB Application	26
	4.2.1.2 MATLAB application Functionality	27
	4.3 SUMMARY	28
5	RESULTS & DISCUSSIONS	29
	5.1 INTRODUCTION	29
	5.2 COMPARISON OF ACCURACY AND RESULTS	33
	5.3 PROJECT OUTPUT	34
	5.4 SUMMARY	34
6	CONCLUSIONS & FUTURE SCOPE	35
	6.1 INTRODUCTION	35
	6.2 CONCLUSION	35
	6.3 FUTURE SCOPE	35
	6.4 SUMMARY	36
	REFERENCES	37

LIST OF FIGURES

FIGURE NO;	LIST OF FIGURES	PAGE NO;
1.1	SYSTEM OVERVIEW	4
3.1	BLOCK DIAGRAM OF EXISTING METHODOLOGY	12
3.2	SVM PROCESS	18
3.3	CNN ARCHITECTURE	19
4.1	PROPOSED SYSTEM	20
4.2	DERN MET IMAGE SAMPLE	22
4.3	GRAYSCALE IMAGE SAMPLE	23
5.1	SVM CONFUSION MATRIX	29
5.2	CNN CONFUSION MATRIX	30
5.3	ACCURACY COMPARISON	31
5.4	RUNTIME COMPARISON	31
5.5	EXISTING SYSTEM	32
5.6	PROPOSED SYSTEM	32
5.7	COMPARISON OF ACCURACY AND RUNTIME	34

LIST OF TABLES

TABLE NO;	LIST OF TABLES	PAGE NO;
4.1	DIFFERENT CLASSIFICATION TECHNIQUES	24
5.1	ACCURACY TABLE	29
5.2	ACCURACY COMPARISON	33

LIST OF ABBREVIATIONS

ISIC	- International Skin Imaging Dataset Collaboration
CNN	- Convolution Neural Network
GLCM	- Gray level Co-occurrence matrix
CAD	- Computer-aided diagnosis
SVM	- Support vector machine
GPU	- Graphics Processing Unit.
SURF	- Speeded Up Robust Features

CHAPTER 1

INTRODUCTION

1.1. INTRODUCTION

The skin is the largest organ in the human body and has numerous potential abnormalities; there are about 1500 distinct skin diseases. We are relatively ignorant about the symptoms of most of these diseases although knowledge is rapidly increasing, however, that makes it a challenge for dermatologist to diagnose them.

Nowadays technologies have changed our day-to-day life in all aspects and the medical field is not an exception, many medical systems have been developed to help both patients and doctors in different ways, starting from registration process ending with the use of technologies for diagnosing diseases.

This chapter gives an overview about the problem statement of the project, the objectives to be achieved, motivation of the project and brief view about the methods used in the implementation. In addition, an overview of report layout will be presented.

Skin cancer is a disease caused by changes in the malignant properties of normal skin cells, causing them to divide uncontrollably into abnormal shapes due to DNA damage. From a histopathological point of view, skin cancer has an irregular architecture with cell differentiation at different levels of chromatin, nucleus and cytoplasm. Skin cancer is a common malignant disease in Indonesia, along with cervical cancer and breast cancer. Skin cancer accounts for 5.9-7.8% of all skin cancers each year. The most common skin cancer in Indonesia is basal cell carcinoma (65.5%), followed by squamous cell carcinoma (23%), malignant melanoma (7.9%) and other skin cancers. Malignant melanoma is less common than basal cell carcinoma and squamous cell carcinoma, but tends to have a higher mortality rate, accounting for 75% of skin cancer deaths. The most aggressive skin cancer is melanoma, which has a high mortality rate, especially if not detected early.

Nonmelanoma skin cancers, such as basal cell carcinoma and squamous cell carcinoma, are more common, less metastatic, and only partially responsible for disability and death. Accurate diagnosis and early detection of skin cancer can support the healing

process and appropriate treatment, and avoid the worst effects of skin cancer. We need an early detection system that enables. B. A benign tumor of the skin that closely resembles skin cancer can facilitate it and raise public awareness.

Automatic classification of skin disorders helps identify skin disorders that occur and immediately consult medical personnel for appropriate treatment. Several related studies based on digital image processing for skin cancer detection and classification are being developed with fast computation times as a tool for medical professionals to more accurately diagnose skin conditions. A previous study was developed using edge detection methods with K-NN and C-NN algorithms, which provided accuracies of 75% and 75.6%, respectively, indicating the likelihood of benign cancer. We have detected a skin condition and possible skin conditions. Classify cancers using the International Skin Imaging Dataset Collaboration (ISIC). In a study, a deep-learning-based automatic skin disease classification system using PNASNet 5-Large architecture was developed using the ISIC dataset of skin cancer and benign skin tumor diseases, with the highest performance of 76%.

Accuracy is obtained. Furthermore, other studies using CNNs to detect skin diseases have achieved performance accuracies of 80.52%, 86.21%, and 87.25%, respectively. An ISIC data augmentation process was performed to increase the amount of data and improve the performance of the skin cancer detection system. Using Alexnet, the best accuracy was 95.91%. In a study, the CNN method using random regulators showed an accuracy of 97.49% for identifying some lesions from skin diseases such as nevus lesions, carcinomas and melanomas. In this study, we use extended data from the ISIC dataset to detect the status of skin cancer lesions or cancer-like benign tumor lesions. The method proposed in this work uses CNN with different optimizer methods such as SGD, RMSprop, Adam and Nadam optimizers to find which optimizer gives the best performance.

1.2. PROBLEM STATEMENT

Skin diseases rate has been increasing for past few decades, many of these diseases are very dangerous, particularly if not treated at early stages. In Sudan skin diseases are big issue, according to the latest WHO data published in May 2014 Skin

Diseases Death in Sudan reached 1,974 or 0.76% of total death. The age adjusted death rate is 9.81 per 100,000 of population, this result ranks Sudan number 8 in the world [1].

In addition, dermatologist use variety of visual clues such as color, scaling and arrangement of lesions, the body site distribution and others, when these individual components are analyzed separately, the recognition of the disease can be quite complex that requiring high level of experience. Diagnosis by humans depends on subjective judgment of the dermatologists so it's hardly reproducible, unlike computer aided diagnostic systems which are more realistic and reliable.

Since most of the Sudanese have dark skin then most of the applications that have been developed in this region are not properly applicable locally in Sudan.

1.3. OBJECTIVES

The objective of this project can be summarized into the following points:

1. Develop Deep learning application that in general, can:
 - Determine the affected areas in the image.
 - Determine the disease in the specified region.
 - Be applicable locally in India.
2. Develop a Android Tablet interface that:
 - Capture the image; send it to the server and receiving the results back.

1.4. MOTIVATION

India is one of the developing countries that need a lot of attention in order to improve the life style of its citizens. Technology needs to be improved on many aspects, especially in the medical field, because of its sensitivity and effect in human lives which requires accurate and objective diagnosis, one of the important and common regions is the branch of medicine that dealing with the skin, nails, hair and its diseases which is called Dermatology.

On the other hand, Android Tablet have become one of our lives basis, they could be used in many ways, Skin Diseases Diagnosis System built-in Android Tablet would be very accurate by implementing machine learning and artificial intelligence techniques in addition its available, reliable and very easy to use.

1.5. METHODOLOGY

This project will implement machine learning techniques for diagnosing Skin Diseases, the input data is an image of the infected skin, and the system will determine the type of the disease. Large number of Skin Diseases images was collected from many online resources such as drmmnet.com, medicine.uiowa.edu, dermnetnz.org and others; we choose to classify between 3 famous Diseases which are Acne, Eczema and Melanoma.

Bag-of-features model using CNN (Convolutional Neural Network) as a classifier is used to train our data, which is very powerful in object and texture recognition. An windows tablet application is used as interface to the user, this application communicates with MATLAB application, the user simply captures the image, a pre- trained classifier in MATLAB is used to define the Skin Disease of this image, then the results sent back to the android application. The Fig 1.1 shows the overall overview of the system.



Fig 1.1: System Overview

1.6 SUMMARY

This chapter introduces the problem statement, objectives, and motivation for a project aimed at developing a deep learning application for skin disease detection in Sudan. Skin diseases are a big issue in Sudan, and diagnosis by humans can be complex and subjective. The objective is to develop an application that can determine affected areas and diseases in images and is applicable locally in Sudan. The motivation is to improve medical technology in developing countries like Sudan.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

Skin diseases, in the last few decades, become very common all over the world, for example number of skin cancer infections has been doubled in the past 15 years, with this extensive spreading and the severity of large number of the skin diseases arise the need for radical solutions for these diseases, the traditional diagnosis of skin diseases requires high level of expertise in the domain and a large amount of knowledge to differentiate between the large number and similar diseases, also it depends on the visual aspects of the physicians, and some of them can't be distinguished just with the human eye and requires further tools or tests, so the traditional skin diseases diagnosis is subjective and may be unreliable, then more proper solutions are needed. This motivates researchers to propose computer aided systems for diagnosing Skin Diseases, especially with the increasing role of the machine learning and data science implementations in the medical field, different machine learning and data science techniques were implemented in those systems because of their ability to recognize patterns in objective manner.

This chapter introduces you with some of the techniques that could be implemented to develop Skin Diseases Diagnosis System based on machine learning approach, in addition provide and discuss the previous attempts to develop Skin Diseases Diagnosis Systems using machine learning or any similar techniques.

2.2 LITERATURE REVIEW

Hasya, et al (2021) proposed an increase in skin cancer cases has been reported in recent decades. Despite being diagnosed by experienced dermatologists, cases are still very difficult due to a wide variety of symptoms and conditions. Recently, deep learning has made significant progress in medical diagnosis, especially skin cancer classification. showed significant performance improvement. In this article, we described his classification of seven types of skin cancer using pre-trained convolutional neural

networks to achieve accurate detection. This method was evaluated for skin cancer classification using the HAM1000 skin lesion dataset. It consists of seven skin cancer types: NV, AKIEC, BCC, MEL, DF, BKL, and VASC. Experimental results show an absolute accuracy of 96% and a real-time accuracy of 80% when using YOLOV3. Yolo is a faster and more accurate model than recognition models. Real-time may be more accurate with the latest YOLO in the future.

Malo et al (2022) investigated the ability of deep convolutional neural systems in classifying benign and high-risk skin cancers. State-of-the-art informed learning designs based on dermoscopy images seem to outperform dermatologists. We found that by using a very deep convolutional neural system and utilizing exchange learning to fine-tune it on dermoscopy images, we were able to achieve much better symptom accuracy compared to skilled physicians and clinicians. Although this work does not involve any preprocessing steps, the exploratory results are very promising. These models can be effectively updated with a dermoscopy framework or smartphone to assist dermatologists. Driving change requires more dermoscopy images and a more diverse dataset (modified categories, different age groups), including his six tests adjusted for each lesson. Using each image's metadata can help improve the accuracy of the model. The new number of prepared tests is 5636. Test set accuracy: 0.875757576118816. The results show that CNNs developed according to the strategy are superior to mar and 4 malignant melanoma in the binary classifications of star sprouts and skin cancer (skin cancer) and keratosis and melanoma (melanoma). It shows that it is not inferior to the tumor. Comparing the new binary classification of keratosis with basal and epithelial carcinomas, the CNN results are slightly worse

Tana et al (2022) proposed an improvement in classifying skin cancer as benign or malignant. The trained SVM and CNN models achieve accuracies of 79.39% and 84.39% respectively. The accuracy achieved is not high enough and training the model with large data sets requires a lot of computational time. This model provides faster and more accurate results than traditional methods and can support non-dermatologists. The accuracy obtained from the results improves after the augmentation process. To improve

accuracy, add concealed layers. Future work will focus on improving the results based on different features of the images available in the dataset. Models ResNet50, VGG-16, and MobileNet can be tuned to reduce computation time. An ensemble approach can be implemented to further analyze the model. In general, we believe that the classification of datasets based on medical images is highly accurate and can improve datasets for future work, and that some specialized medical insights can help improve model performance. is expected. This model can be further deployed via a website or an Android application to assist physician detection applications.

Brinker et al (2018) applied CNNs display a high performance as state of-the-art skin lesion classifiers. Unfortunately, it is difficult to compare different classification methods because some approaches use nonpublic datasets for training and/or testing, thereby making reproducibility difficult. Future publications should use publicly available benchmarks and fully disclose methods used for training to allow comparability.

Stoecker et al(2005) fine-tuned algorithm for binary discrimination of MM and melanocytic nevus. To compare the performance of the classifier with the results obtained by human experts, 157 dermatologists reported their respective treatment decision (biopsy or further treatment versus patient confidence) for 100 test images. Thus, the authors produced the most comprehensive binary dermatoscopic reader study to date. Overall, CNN outperformed 136 out of 157 dermatologists with varying levels of experience in average specificity and .

Young et al (2013) conducted a follow-up study comparing CNN diagnostic scores with 1 dermatologist. In this study, only histologically proven true images (ie, images of suspected MM lesions) were considered, which likely increased the overall difficulty of the test set. However, for the first time CNN based MM classification was significantly better than junior and board-certified dermatologists (82.3% vs. 68.9%/63.2% sensitivity and 77.9% vs. 58.0%/65.2% specificity, p and $lt; 0.001$).

Marchetti et al(2016) published the first dermoscopic comparative study using an

ensemble-based approach to classify predictions by a team participating in the 25th International Biomedical Imaging (ISBI) Challenge. Examining five different fusion methods, the authors showed that the top fusion method was able to outperform eight expert dermatologists. This was significant for both the binary classification of malignant tumors (with a sensitivity of 82% for dermatologists: 76% vs. 59% specificity, $p = 0.02$) and consider treatment 6 decisions (with dermatologist sensitivity 89%: 6% versus 7% specificity, $p = 0.02$). In 2020, Marchetti et al proposed a similar reading study in which the best performing algorithm in the ISBI 2017 challenge significantly outperformed eight dermatologists and nine dermatologists (p and It ; 0.001).

Haenszel et al. (2019) were the first to provide additional information to physicians in a reader survey. The authors proposed a binary classifier for automated MM classification and compared the diagnostic accuracy of the CNN with the results obtained by 58 dermatologists. The course was divided into two levels. In Level I, participants completed a series of tests online and reported their respective diagnosis (MM vs. melanocytic nevus) and treatment decision (excision or short-term observation vs. no activity) based on only one thermoscopic image. At level II, the same dermatologists diagnose using an identical set of tests, but with additional clinical information and close-up photographs. Although the additional information improved the diagnostic accuracy of dermatologists, CNN still significantly exceeded the participant average (88.9% sensitivity for dermatologists: 82.5% vs. 75.7% specificity, p and It ; 0.01).

Haenszel et al. (2020), repeated its earlier survey of readers, comparing CNN's updated version with the results of 96 dermatologists. In this study, they included a wider range of disease categories ($n = 10$), classified as (pre)malignant and benign lesions. When the specificity of the CNN was set at the average level II specificity of the dermatologists' treatment decision (80.%), the sensitivity of the CNN was almost equal to that of the human raters (95.0% vs. 9.1%).

Filimon (2014) attempt to develop a System that suggest a diagnosis regarding skin diseases from erythematous-squamous class [13], the system was developed in

MATLAB environment, its neural network has a hidden layer with 10 neurons, output layer with 6 neurons which has been trained using backpropagation learning algorithm, the system also has 33 inputs of clinical and histopathological feature of the patients, based on these features and the predicted disease the system provides suggestions of the medical treatments of the patient. Although this system got an accuracy that is almost 94% but it doesn't use features that are extracted from an image of the infected area of the skin, it completely depends on clinical features; it also used to diagnose uncommon diseases with a threshold output.

Choudhary et al(2014) have proposed a computer aided diagnosis system for Skin Cancer, based on dermoscopy images of the skin that were enhanced by applying many preprocessing techniques, Dull Razor Filter to remove hair, obtaining the grayscale of the image, contrast enhancement and median filter to remove noise, the infected area was segmented from the rest of the skin according to its binarized gray level by applying Maximum Entropy. Thresholding, Level Co-occurrence Matrix (GLCM) was implemented to extract features from the enhanced image, multi- layers Feedforward Artificial Neural Network with backpropagation learning algorithm was used as a classifier that trained based on the extracted features, their Methodology has got 86.66% as accuracy, but this system only detects whether it's a Melanoma or non- melanoma so it couldn't differentiate between different types of skin diseases despite of their seriousness.

Yasir et al(2014) have proposed a method that uses computer vision based techniques to detect different diseases from color images, the system successfully detects 9 different types of dermatological skin diseases with an accuracy rate of 90% [15], they have used 10 feature, 7 features were extracted from user's input (liquid type, liquid color, elevation, duration, feeling, gender, age) and 3 features were extracted from the image(color, area, edge), those features were extracted using eight different types of algorithms which are grey image, sharpening filter,

median filter, smooth filter, binary mask, histogram, YCbCr and sobel operator. They have trained these 10 feature as input to train and test into a feed forward back propagation artificial neural network to identify the dermatological disease. Although this model has high accuracy but it depends on only 10 features only 3 of them are extracted from the image, and the rest are taken as input from the user, which might not be accurate enough (take feeling as example). making it more human dependent and less automated. In 2015 they have implemented the same model above that have android interface to the system functions, making it portable and more friendly in use.

Amarathunga A.A.L.C et al. (2015) have proposed System enables users to recognize only 3 Skin Diseases and provides advises or treatments , also many image processing and data mining techniques have been implemented in this system, skin images were enhanced using both median and Gaussian filters, thresholding segmentation was applied, then Morphological the enhanced image were extracted, these features along with external information from the user present the input to many classifiers (AdaBoost, BayesNet, J48, MLP, NaiveBayes), both MLP and J84 were better than the rest with accuracy more that 85%, but the system can only recognize 3 diseases(Eczema, Impetigo and Melanoma), also the distance between camera lens and affected skin was 5cm in addition it only developed for windows operating system.

Ambad et al (2016) have develop an Image analysis system to detects skin diseases, they develop a system to be used for early detection and prevention of the skin diseases and they target 3 main diseases skin cancer, psoriasis and dermatophilosis, the disease diagnosis and classification is built on statistical parameter analysis. Statistical parameters include: Entropy, Texture index, Standard deviation, Correlation, the user of the system will able to take images of

different moles or skin patches. Then the system will analyze and process the image and classifies the image to normal, melanoma, psoriasis or dermo case based extracting the image features. an alert will be provided to the user to seek medical help if the mole belongs to the atypical or melanoma category, the input images firstly passed through a median filter to remove a remove the noise, then apply the image enhancement and the statistical analysis techniques, then two-level classifier is used the first level is to specify if the image is either normal or abnormal and the second level is to classify into specified category: Melanoma, Psoriasis or dermo, the system is classify the images with accuracy 90%.

Schendel et al (2019) compared the results of 139 algorithms in the ISBI 2018 challenge with those of 511 human readers, including 283 board-certified dermatologists, 118 dermatologists, and 83 general practitioners. This comparative approach is the most comprehensive cross-category study of readership to date. Regarding the difference between MM and the other six skin diseases (see Supplementary Table 7 for a more detailed breakdown of the categories), the algorithms achieved an average of 19.9 correct diagnoses out of 30, and the participants achieved an average of 17.9 correct diagnoses (p and It). 0.0001).

2.3 SUMMARY

The literature review focus on the use of deep learning, specifically convolutional neural networks (CNN), in skin cancer diagnosis. The authors of the articles discuss the significant progress made by deep learning in the field of skin cancer diagnosis and the potential to achieve higher accuracy than experienced dermatologists. Several studies have shown that CNNs can outperform dermatologists in binary and multi-class classification of skin cancer. However, some challenges remain in the interpretation of the models and the need for publicly available datasets for training and testing to ensure comparability. The proposed models in the articles show promising results, and the authors suggest future research that focuses on improving accuracy by tuning existing models, using metadata, and adding new features to datasets.

CHAPTER 3

EXISTING METHODOLOGY

3.1 INTRODUCTION

There are several existing methodologies for skin cancer detection, including visual inspection by dermatologists, dermoscopy, and various imaging techniques such as reflectance confocal microscopy, optical coherence tomography, and multispectral imaging. Additionally, there are numerous computer-aided diagnosis (CAD) systems that use machine learning algorithms to analyze images of skin lesions and provide automated diagnostic assistance to clinicians. These systems typically involve feature extraction and selection, followed by a classification step using algorithms such as decision trees, support vector machines, or neural networks. Some CAD systems also incorporate deep learning techniques, such as convolutional neural networks (CNNs), which have shown promising results in improving the accuracy of skin cancer detection

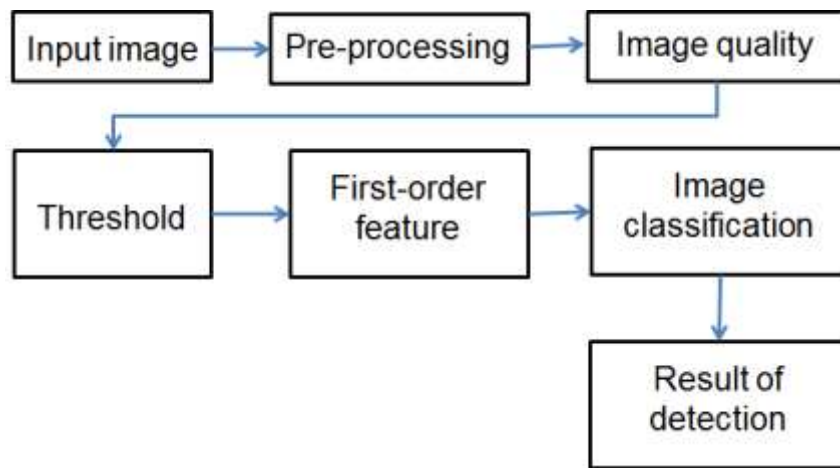


Fig 3.1: Block diagram of Existing Methodology

3.2. IMAGE ACQUISITION: Images are acquired through a camera or locally stored device. Images are obtained from surveys and websites.

3.3. IMAGE PRE-PROCESSING: For Pre-processing of Image, Filtering is performed on image which is a non-linear process used for enhancing the overall image by preserving

the edges of the image. Median filtering is used specially to reduce impulsive, salt-pepper noise. In this, each pixel value in an image is replaced with the median value of its neighboring pixels including itself.

3.4. IMAGE SEGMENTATION: Image segmentation is performed to separate suspicious session from normal skin. This is implemented through MATLAB. For image segmentation, multilevel thresholding using Otsu method is performed where image is segmented into 3 levels using IM Quantize with 2 threshold level. The Segmented image is converted into a color image using function label2rgb ().

3.5. FEATURE EXTRACTION: Unique features of skin lesion are extracted. Features are extracted using the 2D Wavelet Transform. Features extracted using the wavelet transform are Entropy, Mean, Mean Absolute Deviation, Median Absolute Deviation, Energy, Standard deviation, L1 norm, L2 norm, Kurtosis, Skewness. Texture Features extracted using GLCM (Grey Level Co-occurrence Matrix) are Contrast, Correlation, Energy and Homogeneity.

Entropy

$$H [x] = - \sum_x p(x) \log_2 p(x) \quad \text{----- (3.1)}$$

The entropy or average information of an image is a measure of the degree of randomness in the image. The entropy is useful in the context of image coding: it is a lower limit for the average coding length in bits per pixel which can be realized by an optimum coding scheme without any loss of information

Mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{----- (3.2)}$$

Image processing is a somewhat broad term in modern IT that refers to using various means to process or enhance images.

Mean Absolute Deviation

$$MAD = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad \text{----- (3.3)}$$

To find the mean absolute deviation of the data, start by finding the mean of the data set. Find the sum of the data values, and divide the sum by the number of data values.

Median Absolute Deviation

$$MAD = \frac{1}{T} \sum_{t=1}^T d_t \quad \text{----- (3.4)}$$

The mean absolute deviation of a dataset is the average distance between each data point and the mean

Energy

$$E = \sum_{k=-\infty}^{\infty} |x[k]|^2 \quad \text{----- (3.5)}$$

Energy is a fairly loose term used to describe any user defined function (in the image domain). The motivation for using the term 'Energy' is that typical object detection/segmentation tasks are posed as a Energy minimization problem.

Standard deviation

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad \text{----- (3.6)}$$

In statistics, the standard deviation (SD, also represented by the lower-case Greek letter sigma σ for the population standard deviation or the Latin letter s for the sample standard deviation) is a measure that is used to quantify the amount of variation or dispersion of a set of data values.

L1 norm

$$L1 = \left[\sum_i |e_i|^1 \right] \text{----- (3.7)}$$

L1 Norm: Also known as Manhattan Distance or Taxicab norm. L1 Norm is the sum of the magnitudes of the vectors in a space. It is the most natural way of measure distance between vectors that is the sum of absolute difference of the components of the vectors.

L2 norm

$$L2 = \left[\sum_i |e_i|^2 \right]^{1/2} \text{----- (3.8)}$$

The notation for the L2 norm of a vector is. As such, it is also known as the Euclidean norm as it is calculated as the Euclidean distance from the origin. The result is a positive distance value. The L2 norm is calculated as the square root of the sum of the squared vector values

Kurtosis

$$K = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - \mu}{\sigma} \right]^4 - 3 \text{----- (3.9)}$$

Hence, we can use skewness in making judgements about image surfaces. This is because skewness measures how "lopsided" the distribution of pixel values are. For kurtosis: ... Images with moderate amounts of salt and pepper noise are likely to have a high kurtosis value.

Skewness

$$S = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - \mu}{\sigma} \right]^3 \text{----- (3.10)}$$

Skewness is a measure of the asymmetry of the gray levels around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right.

Contrast

$$C = \sum_{i,j} P_{i,j} (i - j)^2 \quad \text{----- (3.11)}$$

Contrast is the difference in luminance or colour that makes an object (or its representation in an image or display) distinguishable. In visual perception of the real world, contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view.

Correlation

$$C = \frac{\sum_m \sum_n (A_{nm} - \bar{A})(B_{nm} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{nm} - \bar{A})^2)(\sum_m \sum_n (B_{nm} - \bar{B})^2)}} \quad \text{----- (3.12)}$$

Sometimes one image is a subset of another. Normalized cross-correlation can be used to determine how to register or align the images by translating one of them.

Homogeneity

$$H = \sum_{i,j} \frac{P_{i,j}}{1 + ||i - j||} \quad \text{----- (3.13)}$$

A material or image that is homogeneous is uniform in composition or character (i.e. color, shape, size, weight, height, distribution, texture, language, income, disease, temperature, radioactivity, architectural design, etc.); one that is heterogeneous is distinctly nonuniform in one of these qualities.

3.6. DATA MINING UNIT

Data Mining is often described as the process of discovering patterns in large sets of data. For detection of skin disease, patterns obtained through the data are used. The data in the dataset comprises of visual features (features extracted from images using image processing). For skin disease classification, Support Vector Machine (SVM) classifier shown in Fig 3.2 and Convolutional Neural Networks (CNN) classifier is used shown in Fig 3.3. Comparison of both the classifiers based on accuracy is shown with the

help of confusion matrix.

3.7. SUPPORT VECTOR MACHINE

The support vector machine is a supervised learning model used for optimization. It is a unified framework in which different learning machine architecture can be generated through an appropriate choice of kernels. The principal used in SVM is statistical and structural risk minimization. The SVM is already a ready-to-use available classifier in MATLAB. After the feature extraction process, the extracted features are directly fed into the SVM classifier. The process involves two phases:

- A. Training Phase:** Huge images of eczema, impetigo, melanoma, and others are used for training.
- B. Testing Phase:** In this phase, test images are given to the classifier and the classifier uses knowledge gained during the training phase to classify the test image.

3.8. CONVOLUTIONAL NEURAL NETWORK

A Convolutional Neural Network (CNN) is slightly in variance with the multilayer perceptron. A CNN can have a single convolution layer or it can contain multiple convolution layers. These layers can be interconnected or pooled together. A convolution operation is performed on the input and then the results are passed to the further layers. Thus, due to this, the network can be deep but will contain only a few parameters. Due to this property, a Convolutional neural network shows effective results in image and video recognition, natural language processing, and recommender systems.

Convolutional Neural Networks give accurate results in semantic parsing and paraphrase detection. This is the main reason to use CNN for skin disease detection. After experimenting with SVM classifier, CNN classifier is implemented to train and test skin disease images. Unlike SVM classifier, there is no need to perform processing steps on image.

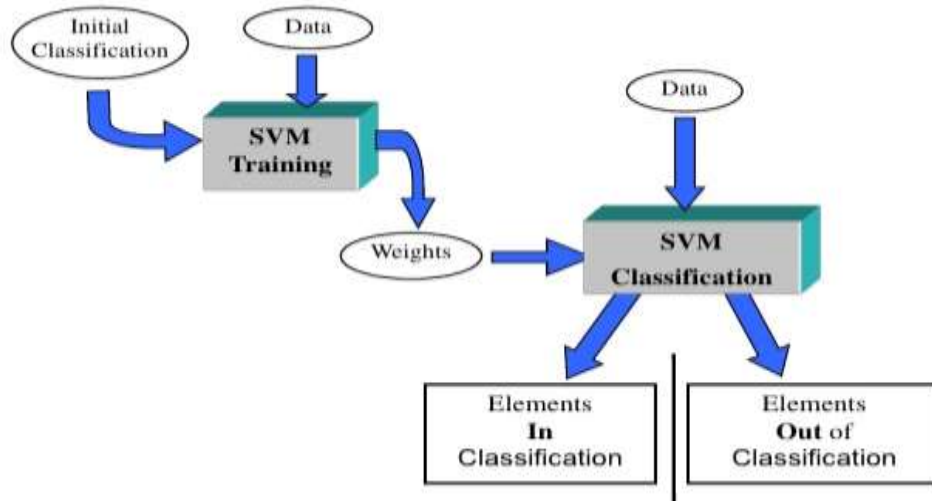


Fig 3.2: SVM Process

The fig 3.2 shows the SVM classifier, an image needs to be processed using image processing unit and then given for the classification to SVM classifier. CNN classifier is implemented in such a way where there is no need of image processing module. CNN classifier is a layered architecture where multiple layers perform various operations to train and test the image data. In this proposed solution, huge images are given to CNN classifier for training where images for training are given to Convolution2dLayer.

This is the first layer to extract the features from the input image. This layer applies a convolution operation and gives the result to the next layer and applies Convolutional filters to the input. It computes the dot product of the input and weights and then adds a bias term. Then ReluLayer is introduced which is Rectified linear Unit Layer for handling nonlinearity in the network.

Max Pooling Layer reduces the dimensionality of image and is used to divide the input into rectangular regions and computes the maxima of each region. After this operation, Fully Connected Layer multiplies an input with weight matrix, adds bias vector and it is responsible for creating a model for classification layer by applying SoftMax Layer. SoftMax Layer is a logistic activation function which is used for multiclass classification. Finally, Classification Layer will detect the affected area of image and gives

the output. The Fig 3.3 shows the CNN architecture.

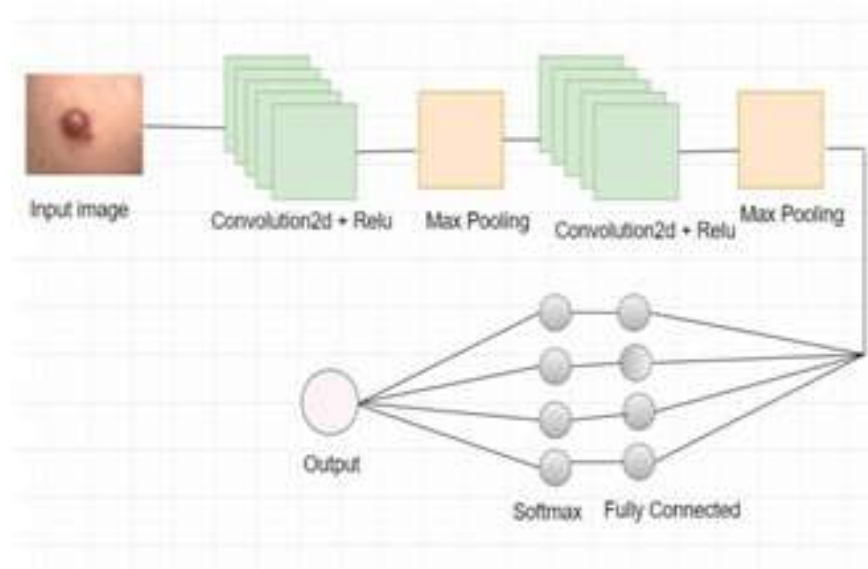


Fig 3.3: CNN Architecture

3.9 SUMMARY

The proposed skin cancer diagnosis system involves image acquisition, pre-processing, segmentation, feature extraction, and classification. The system uses filtering, multilevel thresholding, and wavelet transform for feature extraction, including Entropy, Mean, Mean Absolute Deviation, Median Absolute Deviation, Energy, Standard deviation, L1 norm, L2 norm, Kurtosis, Skewness, and Texture Features extracted using GLCM.

CHAPTER 4

PROPOSED SYSTEM

4.1 INTRODUCTION

This chapter presents each step of the design and implementation of the skin diseases diagnosis system and discuss the methods used in each step, and provide figures and tables from the implementation process for more explanation, the chapter is divided into three main parts the first one is discussing how the data has been gathered, the second one is discussing the classification model design and implementation and the third part is considering the system development and integration. Generally, collecting data that fit your application is one of the most difficult steps in developing a machine learning application, so for developing the Skin Diseases Diagnosis system based on captured images, the required data are images with a labeled classes of skin diseases, so we had two choices either to collect the images manually from hospitals, healthcare centers and individual patients or using online resources for skin diseases images database.

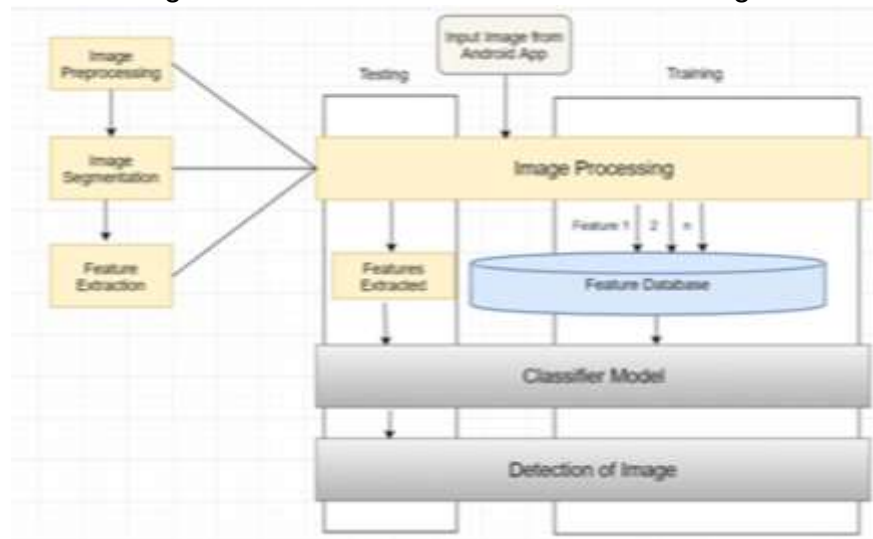


Fig 4.1: Flow diagram of proposed system

For the manual collection of the data, we face two major problems to apply this option, the first one is that there is lack of data collection and documentation in most of the local hospitals, neither digital data nor hard copies of the data, the second problem is that if there is a data available, it's hardly reachable because it's considered private for

the hospitals and require permissions from the local health authorities, also the data is not sufficiently enough for the training of the model, not well prepared and require a lot of work to be ready for usage, so we choose to search for another option to collect the data.

For the online resources option, there are a lot of researches are done to obtain resources for skin diseases diagnosis images. DERMNET.COM was the dominant resource for the system data, it is one of the largest dermatology photos resource that are available publicly. Although it has more than 23,000 skin disease images on a wide variety of skin conditions, but there was not direct way to download a whole class of diseases at once, so we were forced to download each single image every time. in contrast this data couldn't be downloaded at once, we were forced to download each image individually.

To increase the reliability and generalization of our model, it must be trained on different images with different characteristic such as background color of different resources, to achieve this more images were downloaded from other multiple resources such as medicine.uiowa.edu, dermnetnz.org, dermquest.com, dermquest.com along with other websites available at dermweb.com. Most of those images were normally captured by camera with acceptable resolution that more probably will match the type of photos that will be captured by the users.

Our system has been designed to detect 3 types of skin diseases namely, Eczema, Melanoma and Acne. These classes were chosen because Acne and Eczema are of the most common skin diseases that have multiple effects on the patients on different aspects, they are painful in addition they have psychological effect caused by the changes that happen in skin specially for teenagers. Melanoma is the most dangerous form of skin cancer, most often caused by ultraviolet radiation from sunshine or tanning beds [21], If melanoma is recognized and treated early, it is almost always curable, but if it is not, the cancer can advance and spread to other parts of the body, where it becomes hard to treat and can be fatal.

For the training of our system, 200 images for each of Acne, Eczema and Melanoma were collected, although most of the images were downloaded from Dermnet about 140 of each, but they were signed with watermark in the middle of the image, shown

in the Figure 4.2, which restrict us in applying different preprocessing techniques, that will be explained later, this is another reason why we download more data from other sites.



Fig 4.2: Dern met image sample

4.1.1. Images Preprocessing

Before using the images to train our model, series of preprocessing have been applied to our data to enhance the images also to increase our data for better generalization. All these processes were implemented using MATLAB image processing toolbox.

4.1.1.1. Resizing the image

At first all images were resized to be 293*192, resizing the image is important to have a uniform size for all images because the number of features that will be extracted from each image must be unified, we choose this size to reduce the computational efficiency, after resizing then list of preprocessing are applied to the image.

4.1.1.2. Gray images

Images were converted from RGB – red, green and blue- type to gray scale images. Sample is shown in 4.3 below.



Fig 4.3: Grayscale image sample

4.1.1.3. Powered images

Since Negative images are useful for enhancing white or grey detail embedded in dark regions of an image, then we convert these gray images to their negative, then power of two is applied to the negative image to darken the image. Before using Bag-of-Features as basis for training the data, list of techniques for classification was used, Table 4.1 shows the results of these models using 200 images as input for training and evaluation, only Cross-Validation is applied.

Convolutional Neural Networks in addition to its low accuracy, it requires very high computational resources (when using Graphics Processing Unit (GPU), it requires GPU computational capabilities greater than 3.0 which was not available for us), or then when using CPUs, it will take very long time for training our data.

Then we attempt to extract Histogram of Oriented Gradients (HOG) features of each image and classify these features using SVM classifier, the result was insufficient because HOG features are not suitable to present our data.

Table 4.1- Different Classification Techniques

Learning Model	Associated accuracy
CNN	66%
Bag-of-Features	83%
HOG features with SVM classifier	55%

Based on these promising results, bag-of-features was selected to be the model to train our skin diseases diagnosis system's data.

4.1.2. Bag of Features Model

Bag of Features approach in computer vision in the past few decades has been used a lot in many applications. Bag of Features (BoF) methods have been applied to image classification, object detection, image retrieval, and even visual localization for robots.

In our System BOF approach is implemented to train our data, since it is used to classify images based on its texture. BOF approaches are characterized by the use of an order less collection of image features. Lacking any structure or spatial information this eliminates the effect of the water mark in our images which increase the accuracy compared with other learning models such as Convolutional Neural Networks and manually HOG features extraction along with SVM classification, as will be discussed in Error! Reference source not found.

All preprocessed images are combined along with their labels (Acne, Eczema and Melanoma) to form the input data to lean the BoF model, to implement BoF model approach, three steps must be followed. The first step is to extract features from the images, interest point must be detected and described in this step, step two is Quantization, finally the last step is the Classification of the quantized vectors.

4.1.2.1. Step One: Feature Extraction

This step is the base for the coming steps, the features that will be used to train the classifier will be extracted at this step, to achieve this the interest points must both be detected and described.

Interest points detection can be achieved in several ways. Dense feature could be used, also one of the feature extraction techniques such as Harris Corner Detection, FAST, SURF and SUFT that described in 2.4 .

At first we attempt to use SURF features to detect interest points but it was not sufficient, because the result list of the interest points was crossed and concentrated together ignoring large part of the image that considered as important points also. The next figure shows the strongest 10 points extracted using SURF detector.

For better results we choose dense features, every part in the image contributes and effects on the features selected to describe the image, the image is simply divided in regular grid, when the interest points are detected then features are extracted based on these points.

In our model we select SURF algorithm to extract features from each block of each interest point. These features represent what is called visual words describing the image.

4.1.2.2. Step Two: Quantization

All feature of all images (visual of words) are quantized, i.e. using clustering technique to cluster these visual words to specific number of clusters (visual vocabulary), the image is represented as distribution of these words. This is done by k-means clustering to represent group of similar visual words as single cluster (visual vocabulary). The number of desired clusters is selected manually, optimal selection of the number of visual words vocabulary depends on two factors, the first one if it's too long then the computational cost will increase, the second one is that when number of visual words vocabulary is too short then no proper discrimination between features will be obtained. In our model number of 2000 visual words is selected to be the number of visual words. Figure 8, Figure 9 and Figure 10 show the histogram of visual vocabularies of

samples from the three classes of the model, notice the different histogram distribution of each class, which can be used as basis for the classification.

4.2. System Design

In this part, we discuss the overall system architecture, and explain in details the design and development process of each part of the system, and the tools and methods used to do that. The skin diseases diagnosis system mainly consists of two parts that represents the server side and the client side each one contains a separate application and linked together over a shared network. The server side is a MATLAB application that contains the main machine learning model which implement the training and classification task. The client side is an Android application that acts as an interface to the MATLAB application, its main task is to receive the input from the user and pass it to the server and return the output of the server.

- **Application Functionality Capture an Image:** This function is used to capture an image using the mobile camera and feed it as input to the android application, which contains the skin disease image to be classified.
- **Load image from storage:** This function holds the other option instead of capturing the image, which is loading an image stored in the device storage or any associated SD card.

4.2.1.1. MATLAB Application

The MATLAB application is consisted of a several files, each one is responsible of a certain

functionality, and it's divided into three parts:

- **Learning Model:** This part is related to the learning model training and evaluation, and is executed once a new model to be trained and it contains loading the data using Image Datastore object and feed it to the model to be trained as explained in section.
- **Server Connection:** This part is related to establishment of the connection

between the MATLAB and the server, and is consist of two parts the first one for establishing the connection, called once at the beginning and is used to set the host address, the port used, username, the password, java database connector path and string, then use the JDPC driver to connect to the database server.

- **Preprocessing and image manipulation:** This part includes several preprocessing functions which is applied to the data before the training or for the data to be classified, and contains functions for.

4.2.1.2. MATLAB application functionality

These are the main functionalities in the MATLAB application used in the server side:

- **Check the database server for any input:** The MATLAB application is running always on the server side and this function is listening to any changes in the database server, any addition in the database server means that the android application wrote a new input, then this function triggers the MATLAB application to process the new input.
- **Read the image uploaded to the server:** When the MATLAB discovers a new input from the database server, this function read the image path from the database and use that path to load the image into the application to perform a classification task.
- **Image resizing:** When the image is uploaded to the MATLAB application, this function changes the size of the image to a certain size that is configured in the model to be suitable for the classification.
- **Image classification:** This task is to load a pre-trained model and use it to classify the input data into a suitable class of skin diseases, this function is the main function of the application.
- **Write the results back to the server:** This function is performed after the classification process and get the result; it's written in the database server to be received at the client end.

4.3 SUMMARY

The chapter discusses the design and implementation of a skin diseases diagnosis system, which involves gathering data, designing a classification model, and developing and integrating the system. The data was collected from online resources, and the system was trained on images of three skin diseases: Eczema, Melanoma, and Acne. Preprocessing techniques were applied to the images before using them to train the model. The Bag-of-Features model achieved an accuracy of 83%, while Convolutional Neural Networks achieved an accuracy of 66%.

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 INTRODUCTION

It is observed from above table that CNN Algorithm has near perfect accuracy and run time in detecting skin diseases. The confusion matrix shows the percentage of error and accuracy and run time in classification. It also shows corrected and uncorrected results, true positives, false negatives, and number of classes.

Table -5.1: Accuracy Table

Disease	SVM - Accurac y	CNN - Accurac y	Run time – SVM (mS)	Run time – CNN (mS)
Eczema	94%	99%	421.198	19.24
Impetigo	94%	98%	77.34	9.59
Melanoma	99%	99.4%	91.707	10.45
No Disease	52%	98.8%	114.15	21.76
Overall	90.7%	99.1%	176.0988	15.26

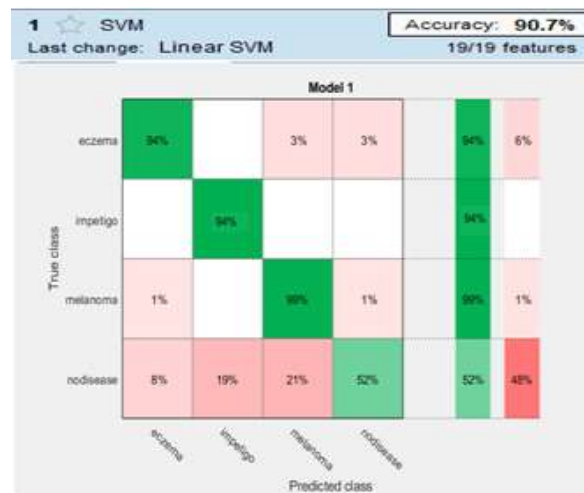


Fig 5.1: SVM Confusion Matrix

The confusion matrix (Fig 5.1) shows the percentage of error and accuracy in classification. It also shows corrected and uncorrected results, true positives, false negatives, and number of classes.

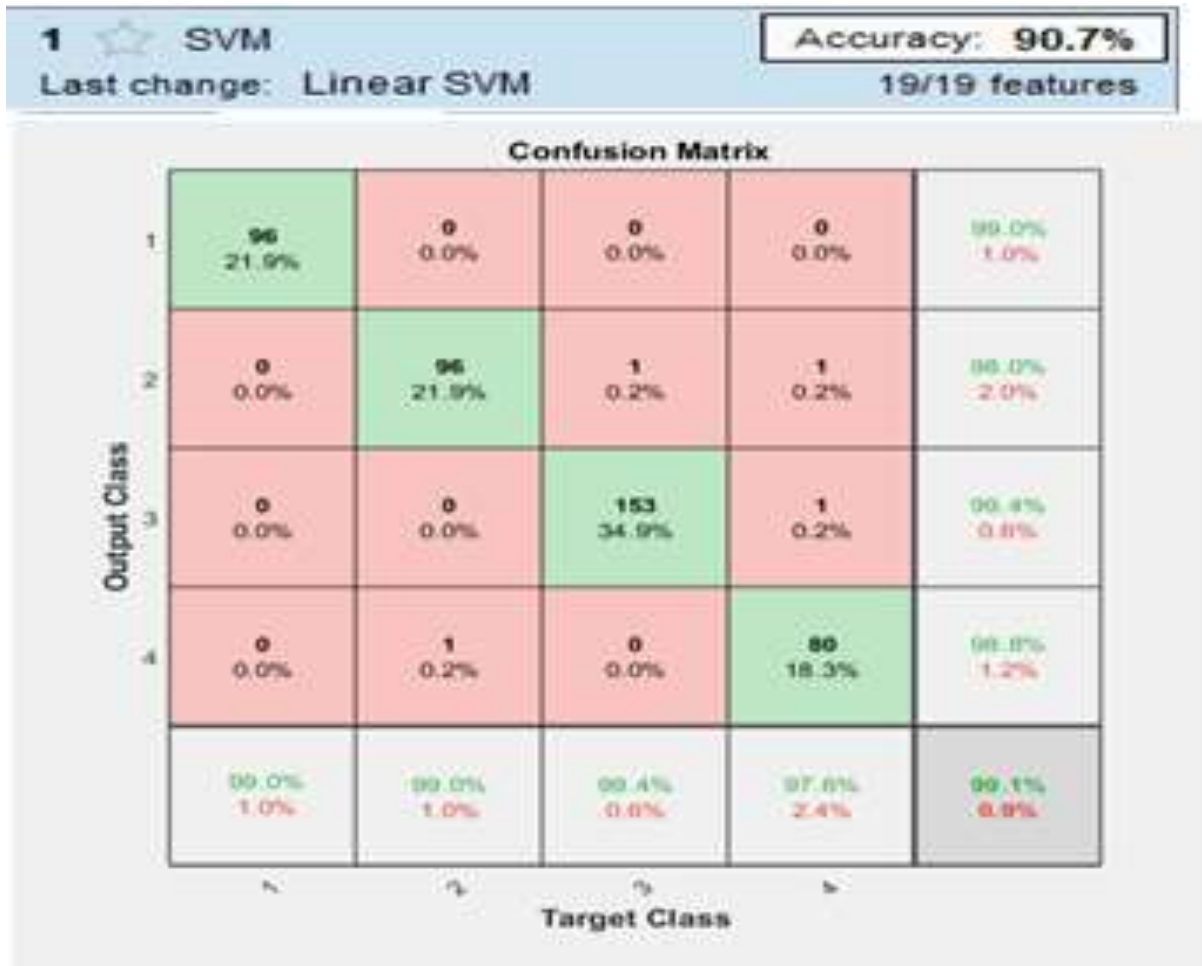


Fig 5.2: CNN Confusion Matrix

Comparison between CNN and SVM Classifier is done with the help of the confusion matrix and the detailed table showing the accuracy of both the classifiers. According to the result obtained, CNN classifier (Fig 5.3 and Fig 5.4) proved to be accurate and efficient in detecting skin disease as compared to SVM Classifier.

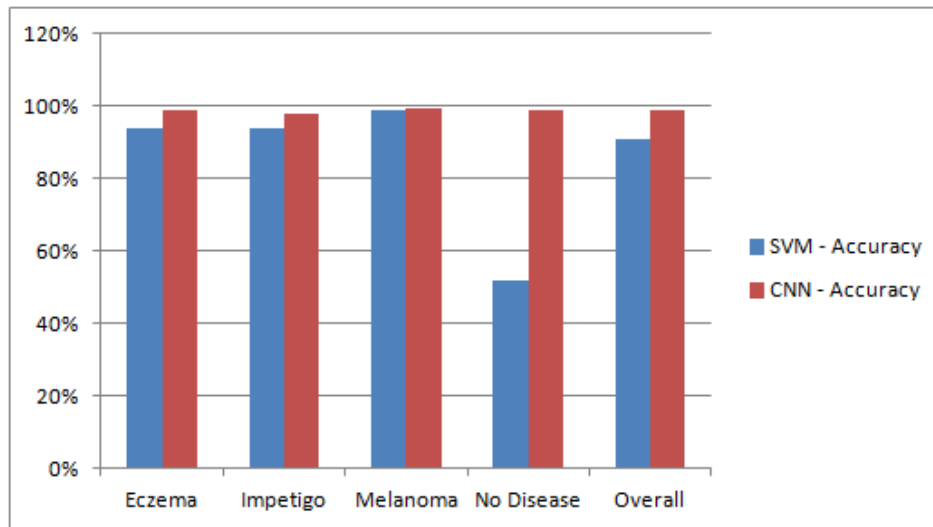


Fig 5.3: Accuracy comparison

The Fig 5.3 compares the accuracy of SVM and CNN models on a given dataset. The graph shows that the CNN model consistently outperforms the SVM model, achieving higher accuracy rates across all training iterations. This suggests that the CNN approach may be more effective for this dataset.

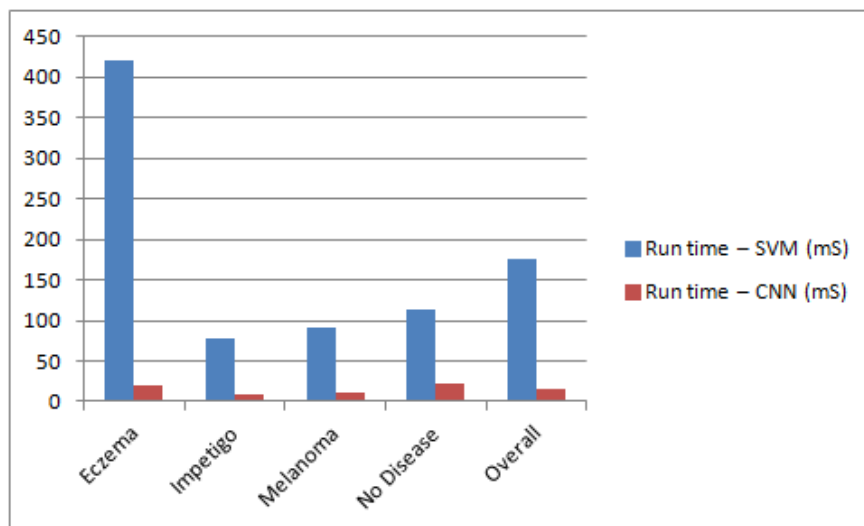


Fig 5.4: Runtime comparison

Based on Fig 5.4, it appears that the Convolutional Neural Network (CNN) outperforms the Support Vector Machine (SVM) in terms of runtime. This suggests that CNNs may be a more efficient choice for certain tasks that require real-time or near real-time processing.

However, it's important to note that the specific task and dataset being used can also affect the performance of these models, so it's always necessary to conduct thorough testing and experimentation before making a final decision on which model to use.



Fig 5.5: Existing system

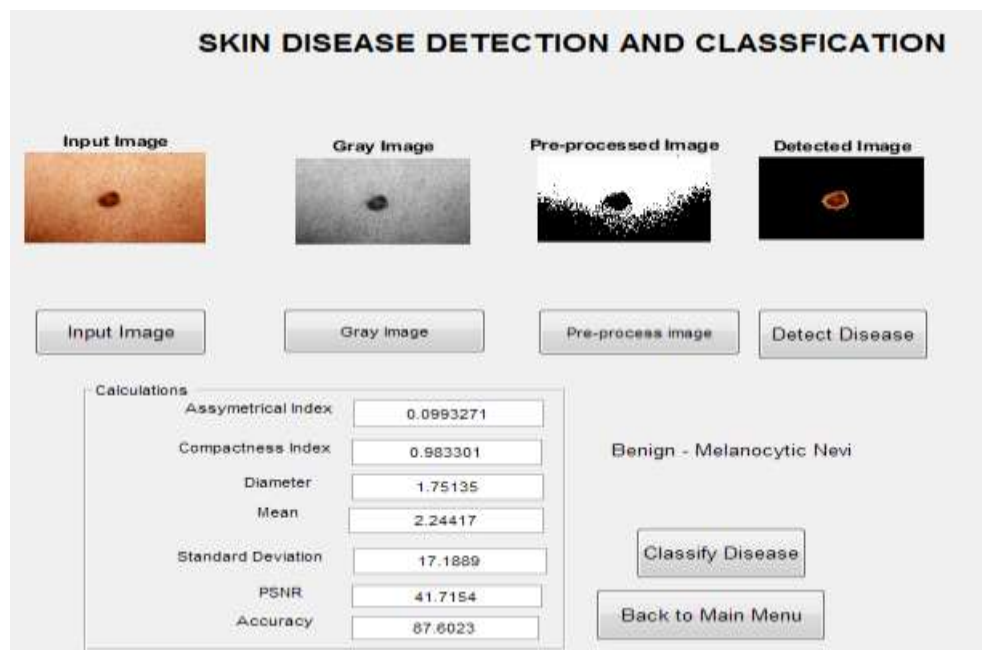


Fig 5.6: Proposed system

From the Fig 5.5 and 5.6 shows the comparison of two different approaches for Realtime skin disease detection algorithm based on accuracy. We have compared Support Vector Machine (SVM) and Convolutional Neural Networks (CNN). The results of real-time testing are presented.

5.2 COMPARISON OF ACCURACY AND RESULTS

Table 5.2: Accuracy comparison

INPUT IMAGE	ACCURACY	RESULT
	92.85	MALIGNANT
	92.81	MALIGNANT
	91.62	BENIGIN
	92.85	MALIGNANT
	91.60	BENIGIN

In the table 5,2 some results are provided using data sets. The accuracy and the result are mentioned.

5.3 PROJECT OUTPUT

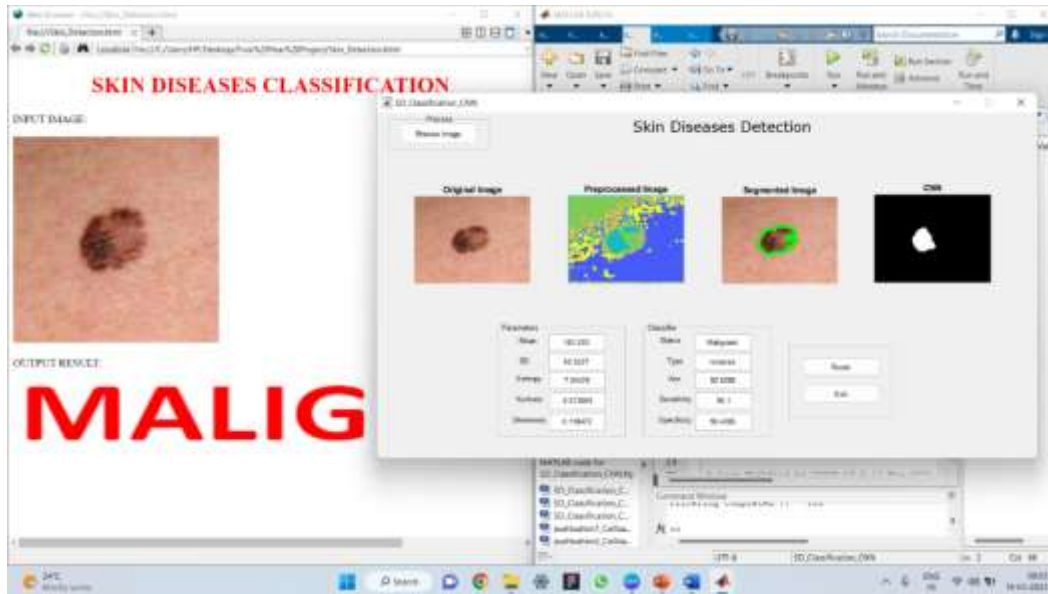


Fig 5.7: Comparison of accuracy and runtime

The fig 5.7 showcases the results of skin cancer detection, including accuracy metrics and the specific type of cancer identified. The accuracy score provides a measure of the model's performance in detecting cancerous lesions, while the type of cancer identified can inform treatment decisions. Overall, this output can help healthcare professionals make informed decisions about patient care

5.4 Summary

This chapter discusses comparison of two different approaches, Support Vector Machine (SVM) and Convolutional Neural Networks (CNN), for real-time skin disease detection. The accuracy and run time of both the classifiers have been evaluated and compared using a confusion matrix and a detailed table. The results show that the CNN classifier is more accurate and efficient in detecting skin diseases than the SVM classifier. The project output showcases the results of skin cancer detection, including accuracy metrics and the specific type of cancer identified. A web page output has also been provided.

CHAPTER 6

CONCLUSIONS & FUTURE SCOPE

6.1 INTRODUCTION

Skin diseases are challenging to diagnose due to their prevalence worldwide, making it difficult for dermatologists to recognize different types of skin diseases. The proposed system uses a machine learning model based on the bag-of-features algorithm, with a convolutional neural network (CNN) as a classifier and Speeded Up Robust Features (SURF) for feature extraction. An interface for mobile applications was developed for Android devices using Android Studio on a Windows operating system.

6.2 CONCLUSION

Difficulties in the diagnosing skin diseases arise because of the spreading of the skin diseases all over the world, which make it a challenge to the dermatologist to recognize the different skin diseases easily, a computer aided system is proposed to resolve these difficulties, so a machine learning model based on bag of features algorithm is designed which use CNN as a classifier and SURF for feature extraction with an interface on mobile applications is developed for android devices, the core model is developed using MATLAB and the interface is developed using android studio on windows OS.

The developed system performs the required work with accuracy 94% within the dataset and 85% with the external data, and the integrated system is working properly, although the system works fine there are some points that bound the performance of the system such as that, there is a single processor of the system that is used to classify the images, which will reduce the performance of the whole system with the increase of the requests from multiple users.

6.3 FUTURE SCOPE

The proposed modifications of the skin diseases diagnosis system are generally to increase the performance of the system, resolve the system limitations, or to increase its

capability.

So, there are several suggested modifications to both the system core model and the system mobile interface:

1. Increase the training data used for training the model, not only in term of quantity but also obtaining more data from different resources namely collecting data from hospitals and healthcare centers, to increase the learning model generalization.
2. Apply better preprocessing techniques to resolve the images distortions.
3. Apply training data of more classes, that the model will be capable to recognize and diagnose more diseases.
4. Develop a cross-platform application to work on different mobile platforms, which will increase the number of system users.
5. Enhance the application interface to be more user friendly for better user experience policies.
6. Develop a distributed system for skin diseases diagnosis to resolve the single server limitation, and increase the processing capabilities.
7. Enhance the functionality of the system to be more useful by giving advices for the users about the disease treatment
8. The system may be combined with other medical systems to propose an integrated medical care service.

6.4 SUMMARY

This chapter discusses the development of a computer-aided system for diagnosing skin diseases using machine learning. The system achieved an accuracy of 94% within the dataset and 85% with external data. Future modifications include increasing training data, improving preprocessing, developing a cross-platform application, and enhancing the system's functionality.

REFERENCES

- [1] Al-Masni, M.A., Azmi, S.A., & Begam, M. (2020). Skin cancer detection using convolutional neural network. *International Journal of Advanced Science and Technology*.
- [2] Chen, Y., Fu, Y., Yang, J., & Huang, Y. (2018). Skin cancer recognition using deep convolutional neural network and ensemble learning. *Journal of Medical Systems*.
- [3] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*.
- [4] Goyal, P., Yadav, S.K., & Yadav, R.K. (2021). A deep learning approach to detect skin cancer. *Journal of Ambient Intelligence and Humanized Computing*.
- [5] Han, S.S., Kim, M.S., Lim, W., Park, G.H., Park, I., Chang, S.E., & Won Lee, M. (2018). Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*.
- [6] Han, S.S., Park, G.H., Lim, W., Kim, M.S., Na, J.I., Park, I., ... & Won Lee, M. (2018). Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PloS one*.
- [7] Kim, H., Goo, J.M., & Kim, H.Y. (2018). Skin cancer detection using convolutional neural network with transfer learning. *Journal of Digital Imaging*, 31(6), 922-930.
- [8] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*.
- [9] Liu, X., Liu, L., Yan, S., & Shen, J. (2020). Skin lesion classification with ensemble CNNs and cross-modality consistency. *IEEE Journal of Biomedical and Health Informatics*.
- [10] Nagaveni, N.B., & Sharada, K. (2020). Skin cancer detection using convolutional neural network. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*.
- [11] Osman, A.A., Al-Turjman, F., & Alajlan, N. (2021). Skin lesion classification using

- CNNs based on multi-scale feature fusion. Computers & Electrical Engineering.
- [12] Peng, Y., Yuan, C., Ding, W., Wu, Q., Yang, X., & Zhang, Y. (2021). Skin cancer recognition using deep convolutional neural network optimized by bat algorithm. Journal of Ambient Intelligence and Humanized Computing.
 - [13] Rohith, M.V., Pai, M.M.M., & Kamath, S. (2021). A deep learning approach for skin cancer detection using convolutional neural network. In Advances in Computing and Data Sciences .
 - [14] Salloum, M., Al-Hiary, Y., Qasaimeh, M., Braik, M., Alhazmi, F., Alhamadi, A., ... & Alnabhan, A. (2021). An efficient skin cancer classification system using deep convolutional neural network. IEEE Access.
 - [15] Yu, L., et al. (2019). "Deep learning-based automated classification of skin cancers using whole-slide pathology images." Journal of Digital Imaging.