# WhatsApp Bot For Disease Classification

**A Project Work Synopsis**

*Submitted in the partial fulfillment for the award of the degree of*

**BACHELOR OF ENGINEERING**

**COMPUTER SCIENCE AND ENGINEERING**

**IN**

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

**Submitted by:**


**Puranjay Kwatra**
**Rajeswar Sharma**
**Vishal Choudhary**
**Aman Sharma**
**Kritesh Sharma**

**19BCS6061**

**19BCS6051**

**19BCS6073**

**19BCS6043**
**19BCS6057**

**Under the Supervision of:**

**DR. MONIKA SINGH**

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413, PUNJAB**


**MARCH, 2022**

# Table of Contents

# ***List of Tables***

# 1. INTRODUCTION

## 1.1 Problem Definition:

In the past 10-year period, from 2008 to 2018, the annual number of melanoma cases has increased by 53%, partly due to increased UV exposure . Although melanoma is one of the most lethal types of skin cancer, a fast diagnosis can lead to a very high chance of survival. The first step in the diagnosis of a malignant lesion by a dermatologist is visual examination of the suspicious skin area. A correct diagnosis is important because of the similarities of some lesion types; moreover, the diagnostic accuracy correlates strongly with the professional experience of the physician. Without additional technical support, dermatologists have a 65%-80% accuracy rate in melanoma diagnosis . In suspicious cases, the visual inspection is supplemented with dermatoscopic images taken with a special high-resolution and magnifying camera. During the recording, the lighting is controlled and a filter is used to reduce reflections on the skin, thereby making deeper skin layers visible. With this technical support, the accuracy of skin lesion diagnosis can be increased by a further 49%. The combination of visual inspection and dermatoscopic images ultimately results in an absolute melanoma detection accuracy of 75%-84% by dermatologists.

## 1.2 Project Overview:

This paper presents the first systematic review of the state-of-the-art research on classifying skin lesions using CNNs. The presented methods are categorized by whether a CNN is used exclusively as a feature extractor or whether it is applied for end-to-end-learning. The conclusion of this paper discusses why the comparability of the presented techniques is very difficult and which challenges must be addressed in the future.

## 1.3 Software Specifications:

The following software specifications are required for the successful completion of the proposed project:

- Python 3.7 or higher
- Jupyter/Colab Notebook
- CNN
- VS Code
- Whatsapp API

## 1.4 Hardware Specification:

The hardware specifications required for the proposed project includes:
and GPU:

- A minimum of 8GB RAM and 2GB of GPU Memory
- A minimum of 20GB hard drive space
- Internet connection for online services

## 2. LITERATURE REVIEW

For some time, the problem of classifying skin lesions has also moved into the focus of the machine learning community. Automated lesion classification can both support physicians in their daily clinical routine and enable fast and cheap access to lifesaving diagnoses, even outside the hospital, through installation of apps on mobile devices . Before 2016, research mostly followed the classical workflow of machine learning: preprocessing, segmentation, feature extraction, and classification .

However, a high level of application-specific expertise is required, particularly for feature extraction, and the selection of adequate features is very time-consuming. In addition, errors and the loss of information in the first processing steps have a very strong influence on the classification quality. For example, a poor segmentation result often leads to poor results in feature extraction and, consequently, low classification accuracy.

In 2016, a change occurred regarding the research of lesion classification techniques. An indication of this change can be found in the methods submitted to the 2016 International Symposium on Biomedical Imaging (ISBI) . The 25 participating teams did not apply traditional standard machine learning methods; instead, they all employed a deep learning technique: convolutional neural networks (CNNs) .

This paper presents the first systematic review of the state-of-the-art research on classifying skin lesions using CNNs. The presented methods are categorized by whether a CNN is used exclusively as a feature extractor or whether it is applied for end-to-end-learning. The conclusion of this paper discusses why the comparability of the presented techniques is very difficult and which challenges must be addressed in the future.

We limited our review to skin lesion classification methods. In particular, methods that apply a CNN only for lesion segmentation or for the classification of dermatoscopic patterns as in Demyanov et al are not considered in this paper. Furthermore, only papers that show a sufficient scientific proceeding are included in this review. This latter criterion includes presenting the approaches in an understandable manner and discussing the results sufficiently. Works in which the origin of the performance was not plausible are not considered in this work, for example, in Carcagnì et al or Dorj et al .

### 2.1 Literature Review Summary

| Title | Author Name | Date | Reference |
|---|---|---|---|
| *Expert Rev Dermatol.* 2014 Jan 10;7(1):1–3. doi:10.1586/edm.11.79. [CrossRef] [Google Scholar] | Nami N | Jan 10, 2014 | 1 |
| *Cancers (Basel)* 2010 Nov 24;2(4):1980–1989. doi: 10.3390/cancers2041980. http://www.mdpi.com/resolver?pii=cancers2041980. [PMC free article] [PubMed] [CrossRef] [Google Scholar] | Fabbrocini G | Nov 24, 2010 | 2 |

# 3 PROBLEM FORMULATION

For some time, the problem of classifying skin lesions has also moved into the focus of the machine learning community. Automated lesion classification can both support physicians in their daily clinical routine and enable fast and cheap access to lifesaving diagnoses, even outside the hospital, through installation of apps on mobile devices . Before 2016, research mostly followed the classical workflow of machine learning: preprocessing, segmentation, feature extraction, and classification . However, a high level of application-specific expertise is required, particularly for feature extraction, and the selection of adequate features is very time-consuming. In addition, errors and the loss of information in the first processing steps have a very strong influence on the classification quality. For example, a poor segmentation result often leads to poor results in feature extraction and, consequently, low classification accuracy.
A basic requirement for the successful training of deep CNN models is that sufficient training data labeled with the classes are available. Otherwise, there is a risk of overfitting the neural network and, as a consequence, an inadequate generalization property of the network for unknown input data. There is a very limited amount of data publicly available for the classification of skin lesions. Almost all published methods use datasets that contain far less than 1000 training data points per training class. In comparison, well-known CNN models for image classification, such as AlexNet , VGG , GoogLeNet , or ResNet , are trained via the large image database ImageNet and have over 1000 training images for each training class.

However, through the use of a specific training procedure called transfer learning, powerful CNN models with several million free parameters can also be employed for classification, even if only a small amount of data are available for training. In this case, the CNN is pretrained using a very large dataset, such as ImageNet; it is then used as an initialization of the CNN for the respective task. In particular, the last fully connected layer of the pretrained CNN model is modified according to the number of training classes in the actual classification task. There are then two options for the weights of the pretrained CNN: to fine-tune all layers of the CNN or to freeze some of the front layers because of overfitting problems and to fine-tune only some back layers of the network. The idea behind this technique is that that the front layers of a CNN contain more generic features (eg, edge or color-blob detectors) that are useful for many tasks, but the back layers of the CNN become increasingly specific to the details of the classes contained in the original dataset.

# 4 OBJECTIVES

The proposed work is aimed to provide people easy and smart solution to diagnose their skin condition(diseases) on whatsapp without any hurdle:

1. Data Collection and analysis.
2. Data Wrangling.
3. Model Selection and Training.
4. Model Deployment.
5. Model Integration using WhatsApp API

# 5 METHODOLOGY

The following methodology will be followed to achieve the objectives defined for proposed project:

1. Installation and hand on experience on existing approaches of image acquisition be done. Relative pros and cons will be identified.

2. Various parameters will be identified to evaluate the proposed system.

3. Comparison of new implemented approaches with existing approaches will be done.

4. The dataset is to be created and cleaning is to be done.

5. Model to integrated using WhatsApp APIs.

# 6 TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK

**CHAPTER 1: INTRODUCTION**

This article provides advanced assistance to dermatologists through deep learning. The essence of the approach is that computers are trained to analyze images of skin cancer to identify problems. The novelty of the presentation lies in the fact that computer models can be developed without programming knowledge. The average diagnostic accuracy using this model is about 98.89%, with the highest being 100%. The mechanical diagnosis presented here overcomes the problems of delay, accuracy, and shortage of dermatologists in public health. According to studies, there are many scientific papers in the field of skin cancer diagnosis and image classification. A detailed review of these methods is available in the work. Each of these articles used the latest technology available at the time and claimed performance improvements. Popular methods used for image classification range from application of decision tree algorithms, Bayesian classifiers supporting vector machines, to various approaches based on artificial intelligence. However, what all these articles have in common is that they are presented as the work of experts in the fields of computer and software development. Building these diagnostic models requires a basic level of programming knowledge in computer languages such as Java, R, and Python. This article reviews how to develop a deep learning-based image classification model for skin cancer detection without prior programming knowledge.

**CHAPTER 2: LITERATURE REVIEW**

Skin cancer is a dangerous and widespread disease. There are approximately 5.4 million new cases of skin cancer each year in the United States alone. The global statistics are not so surprising. According to a recent report, 53% of new melanoma cases are diagnosed annually between 2008 and 2018. Mortality rates from this disease are expected to increase over the next decade. The survival rate is less than 14% when diagnosed at a later stage. However, if skin cancer is detected early, the survival rate is nearly 97%. This requires early detection of skin cancer. This article discusses the problem of early diagnosis with increased accuracy. It has been shown that experienced dermatologists follow a series of steps, usually starting with gross examination of the suspected lesion, followed by dermatoscopy (microscopic enlargement of the lesion) and subsequent biopsy. This takes time and the patient can progress to a later stage. Also, the exact diagnosis is subjective and depends on the qualifications of the doctor. The best dermatologists have been found to misdiagnose skin cancer with less than 80% accuracy. Compounding these challenges is the lack of qualified dermatologists in the global public health system. Extensive research has been conducted to diagnose skin cancer at an early stage and to develop a computerized image analysis algorithm to solve some of the above problems. Most of these algorithmic solutions were parametric, requiring a normal data distribution. These methods cannot accurately diagnose diseases because you cannot control the nature of the data. However, non-parametric solutions do not depend on the constraint that the data are normally distributed.

**CHAPTER 3: BACKGROUND OF PROPOSED METHOD**

Skin cancer detection research based on image analysis has advanced significantly in recent years. Many different techniques have been tried. The 2018 International Skin Imaging Collaboration (ISIC) event is a challenge that has become the de facto benchmark for skin cancer diagnosis. There are also reports that a mobile application can be used to detect skin cancer. In all these efforts, researchers have attempted to increase the accuracy of the diagnosis by using a variety of classification algorithms and methods. Image classification took it to a new level when the Convolutional Neural Network (CNN) framework was introduced by Fukushima (1988) and later LeKun (1990). They used CNNs for image classification. CNNs basically mimic the human visual perception system and are considered the most modern image classification methods. Although much literature on image classification is available, we limit our literature review to deep learning methods on skin cancer images. The first breakthrough in skin cancer classification using the pretrained GoogLeNet Inception V3 CNN model is Esteva et al. They used 129,450 clinical skin cancer images including 3,374 dermatoscopic images. The reported accuracy of classification is 72.1 ± 0.9. In 2016, Yu et al. developed a CNN with over 50 layers on the ISBI 2016 challenge dataset for the classification of malignant melanoma cancer. The best classification accuracy reported in this challenge was 85.5%. In 2018, Haenssle et al. We used a deep convolutional neural network to classify binary diagnostic categories of dermoscopic melanocytic images and reported 86.Sensitivity and specificity for classification 6%. Multiclass classification using ECOC SVM and CNN deep learning is described by Dorj et al. FYI, the approach was to use ECOC SVM with pre-trained AlexNet Deep Learning CNN and multi-class data classification. This article reports an average accuracy of 95.1%. Link Khan et al used deep convolutional neural networks to classify clinical images of 12 skin diseases. With the highest classification accuracy, the claimed copies vary within 96.0% ± 1%. A detailed review of classifiers is beyond the scope of this article. However, a systematic review of deep learning classifiers can be found in Ref.

## CHAPTER 4: METHODOLOGY

The DLS, Model Driven Architecture Tool provides Neural Network Modeling components as a stack of drag-drop & develop art. The sequence of important general steps involved in a research methodology in this paper is as follows:

1. Data Preparation

2 .Creating a project and loading the dataset

3. Building a deep learning classifier

4. Tuning the model

5. Checking result

6. Drawing Inferences

7. Code Access

8. Model Intergration with WhatsApp.

Data preparation is a preprocessing phase in which the raw data is processed by cleaning and formatting. The DLS accepts the data in the form of comma-separated variables (.csv) files. We have used the app (Irfan view) for bulk naming and customizing image features. Creating a project and the rest of the steps follow a brief introduction to the DLS dashboard. Deploying the model as a REST API is reserved for the future. In our study, we tested the following set of models.

## CHAPTER 5: EXPERIMENTAL SETUP

Software Requirements:

• Any Python IDE for creating model

• WhatsApp

• Technologies:

• RNN and LSTM

• GitHub - for tracking project progress

Hardware Requirements:

• A descent Desktop or Laptop

• 8 GB RAM Hardware Required:

• Intel core i3

• 4GB

Software Required:

Jupyter Notebook:

• JupyterLab is a web-based interactive development environment for Jupyter notebooks, code, and data. JupyterLab is flexible: configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning. JupyterLab is extensible and modular: write plugins that add new components and integrate with existing ones. Visual Studio:

## CHAPTER 6: RESULTS AND DISCUSSION

## CHAPTER 7: CONCLUSION AND FUTURE SCOPE

There is a need to incorporate clinicians in the design process for these algorithms and a need to build intrinsically explainable deep learning algorithms for cancer detection. It is important to evaluate proposed methods alongside clinicians to determine the clinical strengths and weaknesses of the methods As per the Covid rules and guidelines there is need of these new methods to be introduced.

This article used Deep Learning Studio and Model-Driven architecture for deep learning. This article introduces the capabilities of the DLS tool and shows the process of building a deep learning model using it. This paper describes the data preparation process using dermal cell imaging and the application of the test in a DLS model for cancer cell detection. The DLS model achieved an AUC of 99.77% when detecting cancer cells in cancer cell images. This document represents a possibility for programmers to obtain the model's program code for further study. Loading trained models and enabling the ability to develop enterprise-grade applications is the best baseline exploration for future work in this article. Finally, the article achieved the goal of the introductory part.

Future Scope:

- Increasing accuracy of model
- Loading more dataset
- Recognising more types of diseases
- Web deployment

**PUBLICATIONS (Optional)**

# 7 REFERENCES

[1]A. Masood, A.A. Al-Jumaily Review article computer aided diagnostic support system for skin cancer: a review of techniques and algorithms

[2] R. Siegel, D. Naishadham, A. Jamal Cancer statistics, 2012 Cancer J. Clin. (2012), pp. 10-29, 10.3322/caac.20138

[3] R. Siegel, K.D. Miller, A. Jamal Cancer statistics, 2018 Cancer J. Clin. (2018), pp. 7-30, 10.3322/caac.21442

[4] A. Esteva1, B. Kuprel1, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun Dermatologist-level classification of skin cancer with deep neural networks Nature (2017), pp. 115-126, 10.1038/nature21056

[5] H. Kittler, H. Pehamberger, K. Wolf, M.J.T.I.O. Binder Diagnostic accuracy of dermoscopy Clin Dermatol (2002), pp. 159-165, 10.1046/j.1365-2230.2000.00693.x

[6] M.M.M. Zorman, S.P. Kokol, I. Malcic The limitations of decision trees and automatic learning in real world medical decision making J Med Syst (1997), pp. 403-415

[7] D Ruiz, V Berenguer, A Soriano, B Sanchez A decision support system for the diagnosis of melanoma: a comparative approach Expert Syst Appl, 38 (12) (Nov-Dec 2011), pp. 15217-15223