Web-based Neural Expert System to detect and classify Skin Cancer

HIGHER DIPLOMA IN SOFTWARE ENGINEERING

<u>Supervised By</u>

Mr. Thilina Rajapaksha

<u>Submitted By</u>

GAHDSE231F-002	Dimuth Nilanjana
GAHDSE231F-004	Hiru Amarajeewa
GAHDSE231F-006	Chandures Priyawindya
GAHDSE231F-013	Thilina Shalom

ACKNOWLEDGEMENT

This note to acknowledge was created for expressing our gratitude and appreciation to everyone who supported us and advised us in various ways throughout the course of this project. First and foremost, we would like to express our gratitude to Mr. Thilina for his invaluable advice and constant encouragement throughout the project.

We would like to show gratitude to our friends and all those who directly and indirectly contributed to the successful completion of this project. Finally, we would want to express our gratitude to our classmates and those who assisted us in finishing this project.

Date………….

Reg. No………

Table of Content

Overview

Skin cancer is a dangerous and widespread disease. It is one of the most active types of cancer in the present decade. As the skin is the body's largest organ, the point of considering skin cancer as the most common type of cancer among humans is understandable. Each year there are approximately 5.4 million new cases of skin cancer recorded in the USA alone. The global statistics are equally alarming. It is generally classified into two major categories: melanoma and nonmelanoma skin cancer. Melanoma is a hazardous, rare, and deadly type of skin cancer.

Skin cancer is found in various types such as Melanoma, Basal and Squamous cell Carcinoma among which Melanoma is the most unpredictable. Laennec made the first description of melanoma in 1804, Jacob of basal cell carcinoma in 1827, and Bowen of squamous cell carcinoma in situ in 1912. Recent reports show that from 2008 to 2018, there has been a 53% increase in new melanoma cases diagnosed annually. The mortality rate of this disease is expected to rise in the next decade. The survival rate is less than 14% if diagnosed in later stages. However, if skin cancer is detected at early stages then the survival rate is nearly 97%. This demands the early detection of skin cancer. This paper addresses the issue of early diagnosis, with improved accuracy. Skin cancer is by far the most common type of cancer. If you have skin cancer, it is important to know which type you have because it affects your treatment options and your outlook (prognosis).

The diagnosis of skin cancer is done by dermatologists where they can access the images of cancer patients and analyze the result of whether the patient has cancerous cells or not. Because of having cancerous cells, dermatologists suggest it as malignant melanoma254 and benign on vice versa. The issue with this framework is, it sets aside a lot of time to process a ton of patients and it takes a great deal of labour to expand the rate of recognition which makes the cost go up. The developing computerized system can automate this skin cancer detection process that will assist the dermatologists, and makes their works easier and faster. Different methods or techniques have been developed for years to make skin cancer diagnoses. A closed elastic curve technique along with an intensity threshold method is proposed to detect the skin lesion boundary accurately. There have been many proposed systems and algorithms such as the seven-point checklist, ABCD rule, and the Menzies method to improve the diagnostics of skin cancer.

At the current state, there are many new technologies/ techniques which are having the potential by which we can analyze this skin cancer at the very early stage with bare minimum needs. This particular project is also using some advanced expert systems and some deep learning techniques which are having some great modern-day support which help this project to detect skin cancer in a very convenient and cost-effective way which is very time efficient and the use of advanced expert systems also make the output reliability percentage increase by a great amount.

Problem Statement

In skin health, diagnosis or diagnostics is the process of identifying a skin texture or problem by its signs, symptoms and the result of various diagnostic procedures. The conclusion reached through this process is called a diagnosis. The diagnosis system is a system that can be used to analyze any problem by answering some questions that lead to a solution to the problem.

The current problem that always happens is that people do not know several things about their skincare. The people only know their problem from their naked eye but what happens in their skin is more serious from that.

The doctor's diagnosis is reliable, but this procedure takes lots of time, effort. These routines can be automated. It could save lots of doctors' time and could help to diagnose more accurately. Besides using computerized means there are good opportunities to store information with diagnostic information to use for further investigations or the creation of new methods of diagnosis. It is only a few minutes that the patients can wait without doing anything until images and other patient's information are all inputted at the store and the analysis results are outputted. Investigations show that early diagnosis is more than 90% curable and late is less than 50%.

We have searched for and implemented the best techniques available for this domain problem and have also integrated several other methods which will further contribute to better and efficient results during the duration of building the project.

The objective of the Research

General Objective

The general objective of our project is to provide a strong and reliable platform for severe diseases detection and a deep analysis of the actual character of the diseases. We are striving to provide useful information and meaningful results to Users mostly if they don't have much knowledge about it. It will also act as a helpful online tool that can tell doctors and lab technologists the highest probability diagnoses for a given skin lesion. This will help them quickly identify high priority patients and speed up their workflow. The app should produce a result in less than 3 seconds. To ensure privacy the images must be pre-processed and analysed locally and never be uploaded to an external server.

Specific objective

- To provide solutions for skin cancer diagnostic problems with state-of-the-art methods and technologies.
- A Reliable source with no misleading information.
- To reduce the complexity of the process by which any person can take advantage of the technology with ease.
- To create a system that handles the evolving nature and which always has room for improvement.
- To deliver the results very fast, within three seconds.
- To ensure privacy we don't upload the image to an external server and do all processing on the go.
- To develop a framework to analyze and assess the risk of melanoma using dermatological photographs taken with a standard consumer-grade camera.

Methodology

We followed the recommended methodology of dividing our Expert System into two parts - backend and frontend.

The backend consists of the following two modules:

- sc_detector: It contains our analysis and findings when dealing with the Skin Cancer Data as well as the Trained Model artefacts. It is the Knowledge Base of our Expert System.
- javascript: It contains the processing codes for utilizing the APIs to deliver results to The frontend from the input image.

The frontend webpage design is built using simple HTML and CSS and the functionalities are powered by JavaScript.

We used GitHub Pages to use the GitHub server for deploying our Website on its domain and also use it as cloud storage to hold our code files, artefacts and the Data that we used for training the model.

Tools used for Training the model or knowledge base

- Python Programming language running on a Jupyter Notebook kernel
- Pandas and NumPy for preprocessing the CSV data
- TensorFlow for Image Loading, Augmentations, Model Building and Training

Frameworks used in JavaScript Backend predictions

- tensorflow.js for preprocessing images and making predictions using the trained model - It is a new library that allows machine learning and deep learning models to run in the browser without having to download or install any additional software. Because the model is running locally, any data that a user submits never leaves his or her pc or mobile phone. Since privacy is especially important when it comes to medical data, we didn't use a server to upload the data and then work on it.
- chart.js for displaying an interactive graph depicting the distributions of the probabilities of each class of cancer

Data used

We used the MNIST: HAM10000 Dataset for training the knowledge base.

Excerpts from the paper:

"The HAM10000 Dataset: A Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions[1]

NV

Melanocytic nevi are benign neoplasms of melanocytes and appear in a myriad of variants, which all are included in our series. The variants may differ significantly from a dermatoscopic point of view.

Mel

Melanoma is a malignant neoplasm derived from melanocytes that may appear in different variants. If excised in an early stage it can be cured by simple surgical excision. Melanomas can be invasive or non-invasive (in situ). We included all variants of melanoma including melanoma in situ but did exclude non-pigmented, subungual, ocular or mucosal melanoma.

BKL

"Benign keratosis" is a generic class that includes seborrheic keratoses ("senile wart"), solar lentigo - which can be regarded as a flat variant of seborrheic keratosis - and lichen-planus like keratoses (LPLK), which corresponds to a seborrheic keratosis or a solar lentigo with inflammation and regression. The three subgroups may look different dermatoscopyically, but we grouped them because they are similar biologically and often reported under the same generic term histopathologically. From a dermatoscopic view, lichen planus-like keratoses are especially challenging because they can show morphologic features mimicking melanoma and are often biopsied or excised for diagnostic reasons.

Bcc

Basal cell carcinoma is a common variant of epithelial skin cancer that rarely metastasizes but grows destructively if untreated. It appears in different morphologic variants (flat, nodular, pigmented, cystic, etc), which are all included in this set.

Apiece

Actinic Keratoses (Solar Keratoses) and intraepithelial Carcinoma (Bowen's disease) are common non-invasive, variants of squamous cell carcinoma that can be treated locally without surgery. Some authors regard them as precursors of squamous cell carcinomas and not as actual carci- nomas. There is, however, an agreement that these lesions may progress to invasive squamous cell carcinoma - which is usually not pigmented. Both neoplasms commonly show surface scaling and commonly are devoid of pigment.

Actinic keratoses are more common on the face and Bowen's disease is more common on other body sites. Because both types are in- induced by UV-light the surrounding skin is usually typified by severe sun damage except in cases of Bowen's disease that are caused by human papillomavirus infection and not by UV. Pigmented variants exist for Bowen's disease and actinic keratoses. Both are included in this set.

Vasc

Vascular skin lesions in the dataset range from cherry angiomas to angiokeratomas and pyogenic granulomas. Haemorrhage is also included in this category.

Df

Dermatofibroma is a benign skin lesion regarded as either a benign proliferation or an inflammatory reaction to minimal trauma. It is brown often showing a central zone of fibrosis dermatoscopically."

Model Architecture used to construct the Knowledge Base

We used the MobileNet pre-trained model for transfer learning using TensorFlow and trained it using the Data to generate the model artefacts. The main we used it was because MobileNet's small size and speed make it ideal for web deployment. We used these artefacts to make predictions using JavaScript through the frontend.

Techniques used during Training the Model

Image Processing

Image processing techniques are allowing earlier detection of abnormalities and treatment monitoring. Because time is a very important factor in cancer treatment, especially in cancers such as the lung and breast, imaging techniques are used to accelerate diagnosis more than with other cancers. It is also very important because users might input the images in different shapes and sizes while the knowledge base can work only on data of a specific shape.

Image Augmentation

Image augmentation is a technique of altering the existing data to create some more data for the model training process. This increased training images and thus helped us in further generalizing the model for predicting from any kind of skin lesion images.

Transfer Learning

Transfer learning plays a major role in medical image analysis. Transfer learning techniques are frequently used in ultrasound skin cancer image analyses. In this method, a pre-trained deep learning network and transfer learning are utilized. In addition to fine-tuning and data augmentation, transfer learning is applied using MobileNet by excluding the last 5 layers of the model and creating a new Dense layer for predicting the 7 corresponding classes of cancer using the softmax activation function.

Literature review

Noninvasive Real-Time Automated Skin Lesion Analysis System for Melanoma Early Detection and Prevention

In the diagnosis of skin melanoma by analyzing histopathological images, the detection of the melanocytes in the epidermis area is an important step. However, the detection of melanocytes in the epidermis area is difficult because other keratinocytes that are very similar to the melanocytes are also present. This paper 1 proposes a novel computer-aided technique for segmentation of the melanocytes in the skin histopathological images. To reduce the local intensity variant, a mean-shift algorithm is applied for the initial segmentation of the image. A local region recursive segmentation algorithm is then proposed to filter out the candidate nuclei regions based on the domain prior knowledge. To distinguish the melanocytes from other keratinocytes in the epidermis area, a novel descriptor, named local double ellipse descriptor (LDED), is proposed to measure the local features of the candidate regions. The LDED uses two parameters: region ellipticity and local pattern characteristics to distinguish the melanocytes from the candidate nuclei regions.

Classification of malignant melanoma and benign skin lesions implementation of automatic ABCD

Melanoma is the deadliest form of skin cancer. Incidence rates of melanoma have been increasing, especially among non-Hispanic white males and females, but survival rates are high if detected early. Due to the costs for dermatologists to screen every patient, there is a need for an automated system to assess a patient's risk of melanoma using images of their skin lesions captured using a standard digital camera. One challenge in implementing such a system is locating the skin lesion in the digital image. A novel texture-based skin lesion segmentation algorithm is proposed. A set of representative texture distributions are learned from an illumination-corrected photograph and texture distinctiveness metric is calculated for each distribution. Next, regions in the image are classified as normal skin or lesions based on the occurrence of representative texture distributions. The proposed segmentation framework is tested by comparing lesion segmentation results and melanoma classification results to results using other state-of-art algorithms. The proposed framework has higher segmentation accuracy compared to all other tested algorithms.

Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance

Melanoma spreads through metastasis, and therefore, it has been proved to be very fatal. Statistical evidence has revealed that the majority of deaths resulting from skin cancer are as a result of melanoma. Further investigations have shown that the survival rates in patients depend on the stage of cancer; early detection and intervention of melanoma implicate higher chances of cure. Clinical diagnosis and prognosis of melanoma are challenging since the processes are prone to misdiagnosis and inaccuracies due to doctors' subjectivity. Malignant melanomas are asymmetrical, have irregular borders, notched edges, and colour variations, so analyzing the shape, colour, and texture of the skin lesion is important for the early detection and prevention of melanoma. This paper proposes the two major components of a noninvasive real-time automated skin lesion analysis system for the early detection and prevention of melanoma. The first component is a real-time alert to help users prevent skin burn caused by sunlight; a novel equation to compute the time for skin to burn is thereby introduced. The second component is an automated image analysis module, which contains image acquisition, hair detection and exclusion, lesion segmentation, feature extraction, and classification. The proposed system uses the PH2 Dermoscopy image database from Pedro Hispano Hospital for development and testing purposes. The image database contains a total of 200 dermoscopy images of lesions, including benign, atypical, and melanoma cases. The experimental results show that the proposed system is efficient, achieving classification of the benign, atypical, and melanoma images with an accuracy of 96.3%, 95.7%, and 97.5%, respectively.

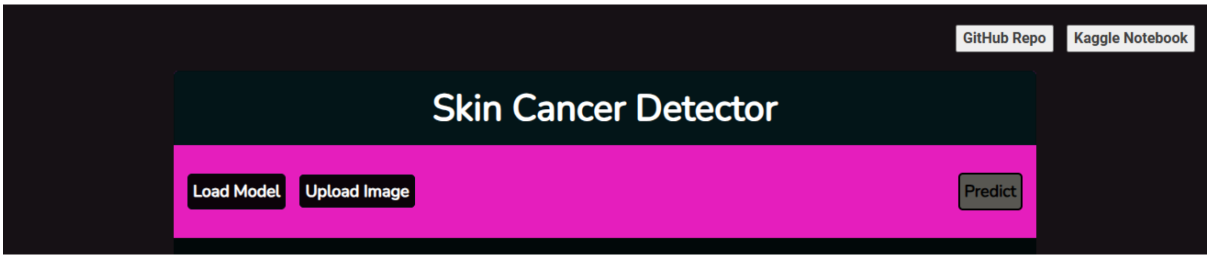
Scope and Limitation of the Project

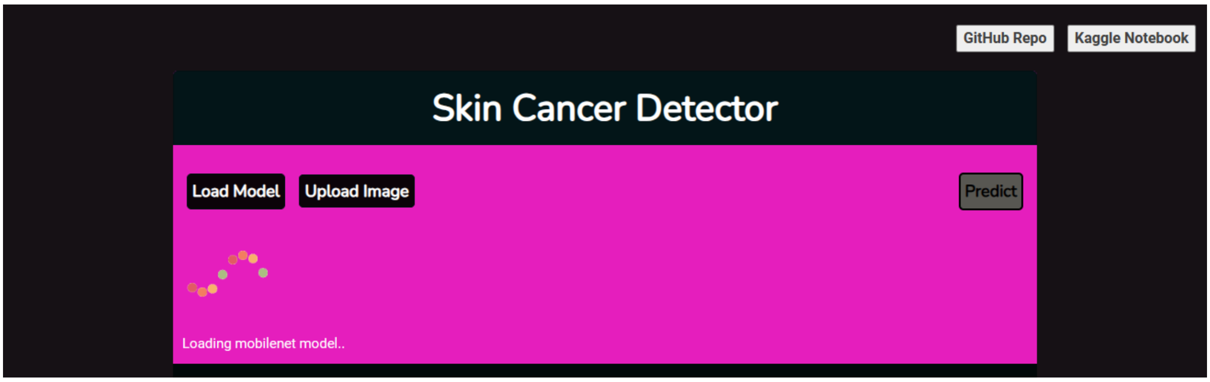
This project is built on the foundation which allows it to become a mainstream product in the field of medical/health care using an advanced expert system which is going to help a significant number of people by providing quick results with certainly reliable information (in the area of skin cancer detection). This project is having a potential environment that will provide a baseline segment for all the related/similar problems in a very advanced way, which will help these products to get upgraded and enhanced to meet the requirement of the constantly evolving world. The universal laws of technology always tell us that every technology is having its limitations so this project is also having its limitations, this project is having a major section completely based on transfer learning so every limitation transfer learning is having will be carried to this project also. One of the most challenging things which this project is going face is that any new addition to this project should be very well optimized and bug-free as we have to prevent misinformation and due to that the complexity will increase with a certain degree.

Significance of the project

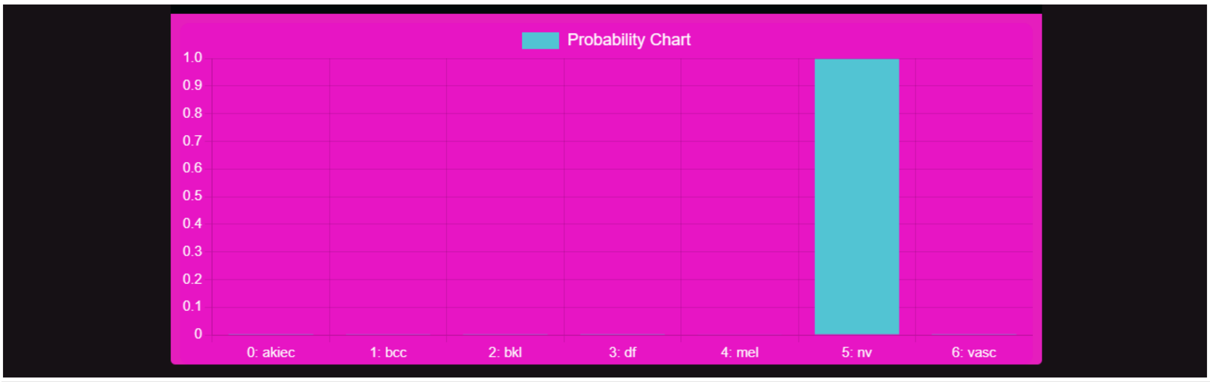
This project is being able to utilize the very advanced expert system and part of deep learning and image processing which create new finds in these fields. It will allow many people to deploy some new research or create processes that can create some new extensions which are more specific to advance health care analyzing and predicting systems at very high accuracy without needing too much optimization and fine-tuning again and again which can lead to many developers to create more dimensional and productive products.

Interfaces


GitHub Repo
Kaggle Notebook
Skin Cancer Detector
Load Model
Upload Image
Predict


GitHub Repo
Kaggle Notebook
Skin Cancer Detector
Load Model
Upload Image
Predict
Loading mobilenet model..


GitHub Repo
Kaggle Notebook
Skin Cancer Detector
Loaded Model
Upload Image
Predict
The predicted type of Skin Cancer is:
nv, Melanocytic Nevi
Top-3 Predictions
nv, Melanocytic Nevi(5) 0.997561
mel, Melanoma(4) 0.001392
bcc, Basal Cell Carcinoma(1) 0.000627


Probability Chart
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0: akiec
1: bcc
2: bkl
3: df
4: mel
5: nv
6: vasc

References

Tschandl, P., Rosendahl, C. and Kittler, H. (2018) *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*, *arXiv.org*. Available at: https://arxiv.org/abs/1803.10417 (Accessed: 16 March 2024).

Tirth27 (no date) *Tirth27/Skin-cancer-classification-using-deep-learning: Classify skin cancer from the skin lesion images using image classification. the dataset for the project is obtained from the Kaggle Siim-ISIC-melanoma-classification competition.*, *GitHub*. Available at: https://github.com/Tirth27/Skin-Cancer-Classification-using-Deep-Learning (Accessed: 16 March 2024).

Vbookshelf (no date) *VBOOKSHELF/skin-lesion-analyzer: AI powered web app that can analyze a picture of a skin lesion and instantly classify it into one of 7 types - including cancerous lesions like melanoma.*, *GitHub*. Available at: https://github.com/vbookshelf/Skin-Lesion-Analyzer (Accessed: 16 March 2024).

Abuzaghleh, O., Barkana, B.D. and Faezipour, M. (2015) *Noninvasive real-time automated skin lesion analysis system for melanoma early detection and prevention – doaj*, *IEEE Journal of Translational Engineering in Health and Medicine*. Available at: https://doaj.org/article/bd0f904a5ecf449586f55740c16e2022 (Accessed: 16 March 2024).

(No date) *Classification of malignant melanoma and benign skin lesions: ...* Available at: https://www.researchgate.net/publication/295074683_Classification_of_malignant_melanoma_and_benign_skin_lesions_Implementation_of_automatic_ABCD_rule (Accessed: 16 March 2024).

YC;, Y.Y.M. (no date) *Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance*, *IEEE transactions on medical imaging*. Available at: https://pubmed.ncbi.nlm.nih.gov/28436853/ (Accessed: 16 March 2024).

Tensorflow.js - introducing deep learning with client-side Neural Networks (2018) *YouTube*. Available at: https://www.youtube.com/watch?v=HEQDRWMK6yY (Accessed: 16 March 2024).

Performance measure on multiclass classification [accuracy, F1 score, precision, recall] (2017) *YouTube*. Available at: https://www.youtube.com/watch?v=HBi-P5j0Kec (Accessed: 16 March 2024).

How to evaluate a classifier in scikit-learn (2015) *YouTube*. Available at: https://www.youtube.com/watch?v=85dtiMz9tSo (Accessed: 16 March 2024).

Gallery (no date) *GitHub*. Available at: https://github.com/tensorflow/tfjs/blob/master/GALLERY.md (Accessed: 16 March 2024).