International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Deep Learning Based Skin Cancer Prediction

1st Nutan Chaudhari
Department of Computer Science
MCOERC,Nashik
nutanchaudhari001@gmail.com

2nd Virendra Gaikwad
Department of Computer Science
MCOERC,Nashik
veergaikwad777@gmail.com

3rd Sakshi More
Department of Computer Science
MCOERC,Nashik
moresakshinandu123@gmail.com

4th AmrutaSaraf
Department of Computer Science
MCOERC,Nashik
sarafamruta981@gmail.com

ABSTRACT

In cancer, there are over 200 different forms. Out of 200, melanoma is the deadliestform of skin cancer. The diagnostic procedure for melanoma starts with clinical screening, followed by dermoscopic analysis and histopathological examination. Melanoma skin cancer is highly curable if it gets identified at the early stages. Thefirst step of Melanoma skin cancer diagnosis is to conduct a visual examination of the skin's affected area. Dermatologists take the dermatoscopic images of the skinlesions by the high-speed camera, which have an accuracy of 65-80% in the melanoma diagnosis without any additional technical support. With further visual examination by cancer treatment specialists and dermatoscopic images, the overallprediction rate of melanoma diagnosis raised to 75-84% accuracy. The project aimsto build an automated classification system based on image processing techniques to classify skin cancer using skin lesions images.

Keywords:

Deep learning, AI Model driven architecture, Deep cognition studio, CNN, Cancer ,Image classification.

Introduction

Among all the skin cancer type, melanoma is the least common skin cancer, but itis responsible for **75%** of death SIIM-ISIC Melanoma Classification, 2020. Being a lesscommon skin cancer type but is spread very quickly to other body parts if not diagnosed early. The **International Skin Imaging Collaboration (ISIC)** is facilitatingskin images to reduce melanoma mortality. Melanoma can be cured if diagnosed and treated in the early stages. Digital skin lesion images can be used to make a teledermatology automated diagnosis system that can support clinical decision.

Currently, deep learning has revolutionised the future as it can solve complex problems. The motivation is to develop a solution that can help dermatologists better support their diagnostic accuracy by ensembling contextual images and patient-level information, reducing the variance of predictions from the model.tr

Motivation

The overarching goal is to support the efforts to reduce the death caused by skin cancer. The primary motivation that drives the project is to use the advanced image classification technology for the well-being of the people. Computer vision has made good progress in machine learning and deep learningthat are scalable across domains. With the help of this project, we want to reduce the gap between diagnosing and treatment. Successful completion of the project with higher precision on the dataset could better support the dermatological clinic work. The improved accuracy and efficiency of the modelcan aid to detect melanoma in the early stages and can help to reduce unnecessary biopsies

Application

We aim to make it accessible for everyone and leverage the existing model and improve the current system. To make it accessible to the public, we buildan easy-to-use website. The user or dermatologist can upload the patient demographic information with the skin lesion image. With the image and patient demographic as input, the model will analyse the data and return theresults within a split second. Keeping the broader demographic of people in the vision, we have also tried to develop the basic infographic page, which provides a generalised overview about melanoma and steps to use the online tool to get the results.

Literature Survey

1. **Paper Name:**Using Some Data Mining Techniques for Early Diagnosis ofskin Cancer

Author Name: ZakariaSulimanZubi and RemaAsheibaniSaad **Description:**skin cancer is a disease of uncontrolled cell growth in tissues ofthe skin, skin cancer isone of the most common and deadly diseases in the world. Detection of skin cancer in its early stageis the key of its cure. In gen-eral, a measure for early stage skin cancer diagnosis mainly includes thoseu-tilizing, CT, MRI, etc. Medical images mining is a promising area ofcom-putational intelligence applied to automatically analyzing patient's recordsaiming at thediscovery of new knowledge potentially useful for medical de-cision making. Firstly we will use someprocesses are essential to the taskof medical image mining, Data Preprocessing, Feature Extractionand RuleGeneration. The methods used in this paper work states, to classify thedigital X-ray chestfilms into two categories: normal and abnormal. The nor-mal state is the one that characterize a healthypatient. The abnormal stateincluding the types of skin cancer; will be used as a commonclassificationmethod indicating a machine learning method known as neural networks. In addition,we will investigate the use of association rules in the problem of x-ray chest films categorization.The digital x-ray chest films are storied inhuge multimedia databases for a medical purpose.

• Methodology

Our methodology consists of using Convolutional Neural Network (CNN) to identify and diagnose the skin cancer using the IS IC dataset containing 2637 images. The proposed model gives an accuracy of 88% for classifying the training dataset as either benign or malignant.

• Dataset

The project dataset is openly available on Kaggle (SIIM-ISIC Melanoma Classification, 2020). It consists of around forty-four thousand images from thesame patient sampled over different weeks and stages. The dataset consists of images in various file format. The raw images are in **DICOM (Digital Imaging and COmmunications in Medicine)**, containing patient metadata and skin lesion images. DICOM is a commonly used file format in medical imaging. Additionally, the dataset also includes images in **TFRECORDS (TensorFlow Records)** and JPEG format**.**

Furthermore, thirty-three thousand are in training set among the forty-four thousand images and around eleven thousand in the test set. However, our quick analysis found a significant class imbalance in the training dataset. Thirty-two thousand are labelled as **benign (Not Cancerous)** and only five hundred marked as **malignant (Cancerous)**. That is, the training set contains only ±1.76% of malignant images (Figure 1). Along with the patient's images, the dataset also has a CSV file containing a detail about patient-level contextual information, which includes patient id, gender, patient age, location of benign/malignant site, and indicator of malignancy for the imagedlesion.

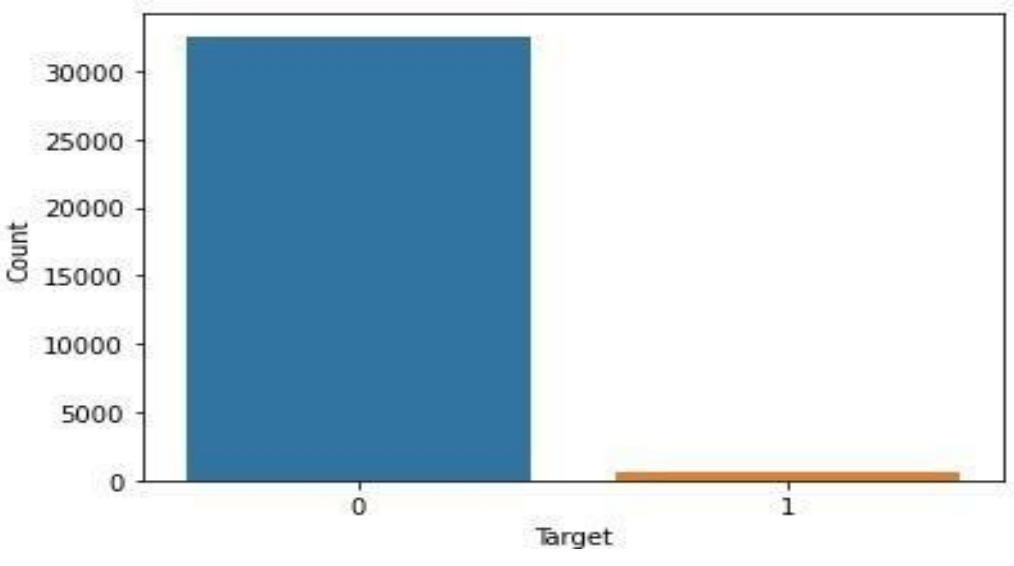


Figure 1: Class imbalance

Overview of the Architecture

The project contains two flow diagrams.

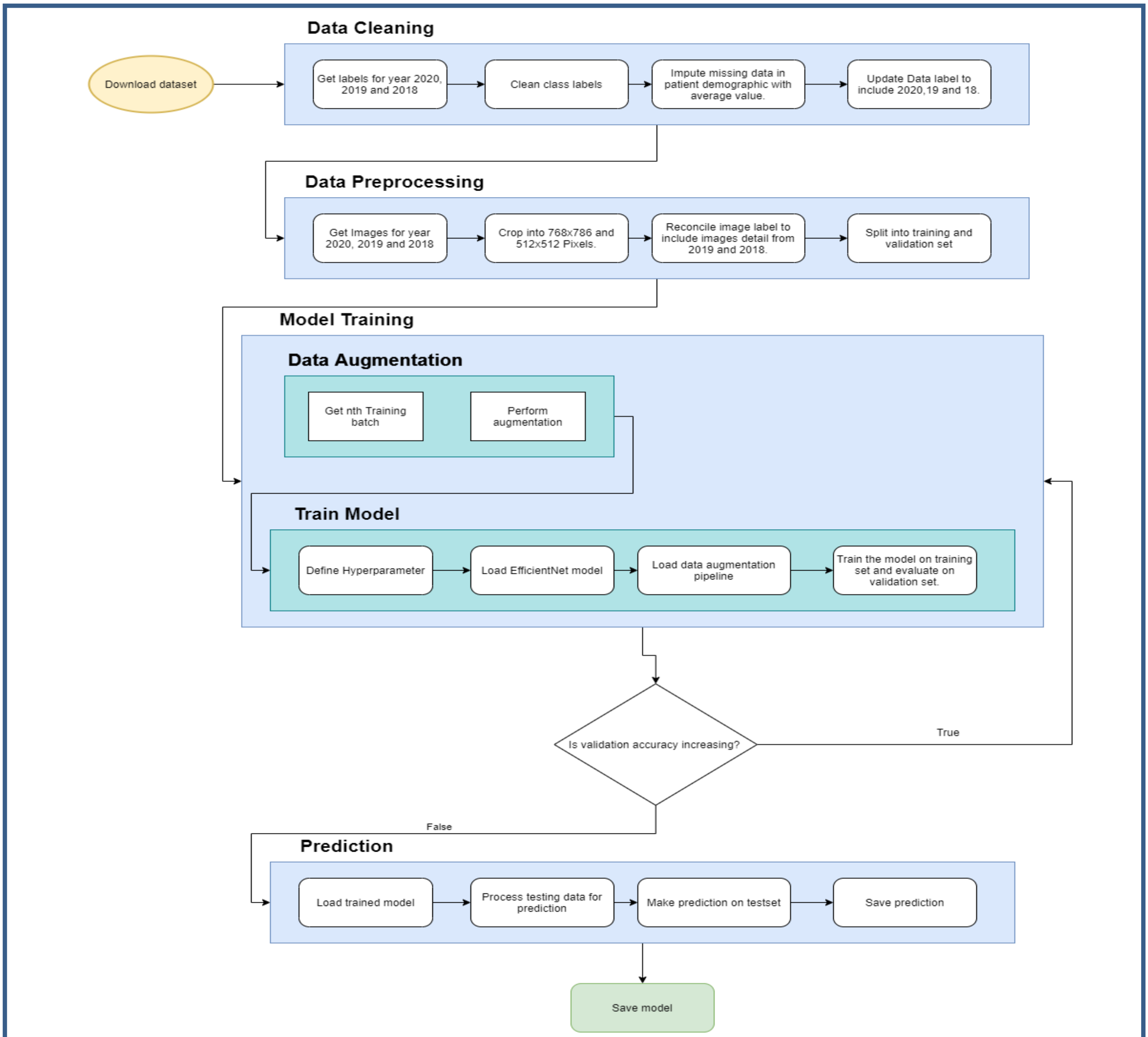


Figure 2: Model training flow diagram

The web UI contains five pages, of which four of them are used to explain the project andhow to use the proposed CAD system (Figure 3). The inference page named **"Our Solution"** is where the inference is made on the skin lesion images. All the validation is performed on the client-side to reduce the server overload. If the inserted information is not correct, then an error notification popup is shown; any user can easily understand that.Validated data is passed onto the server, where inference is performed by Onnx network, and response is return in the JSON format.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

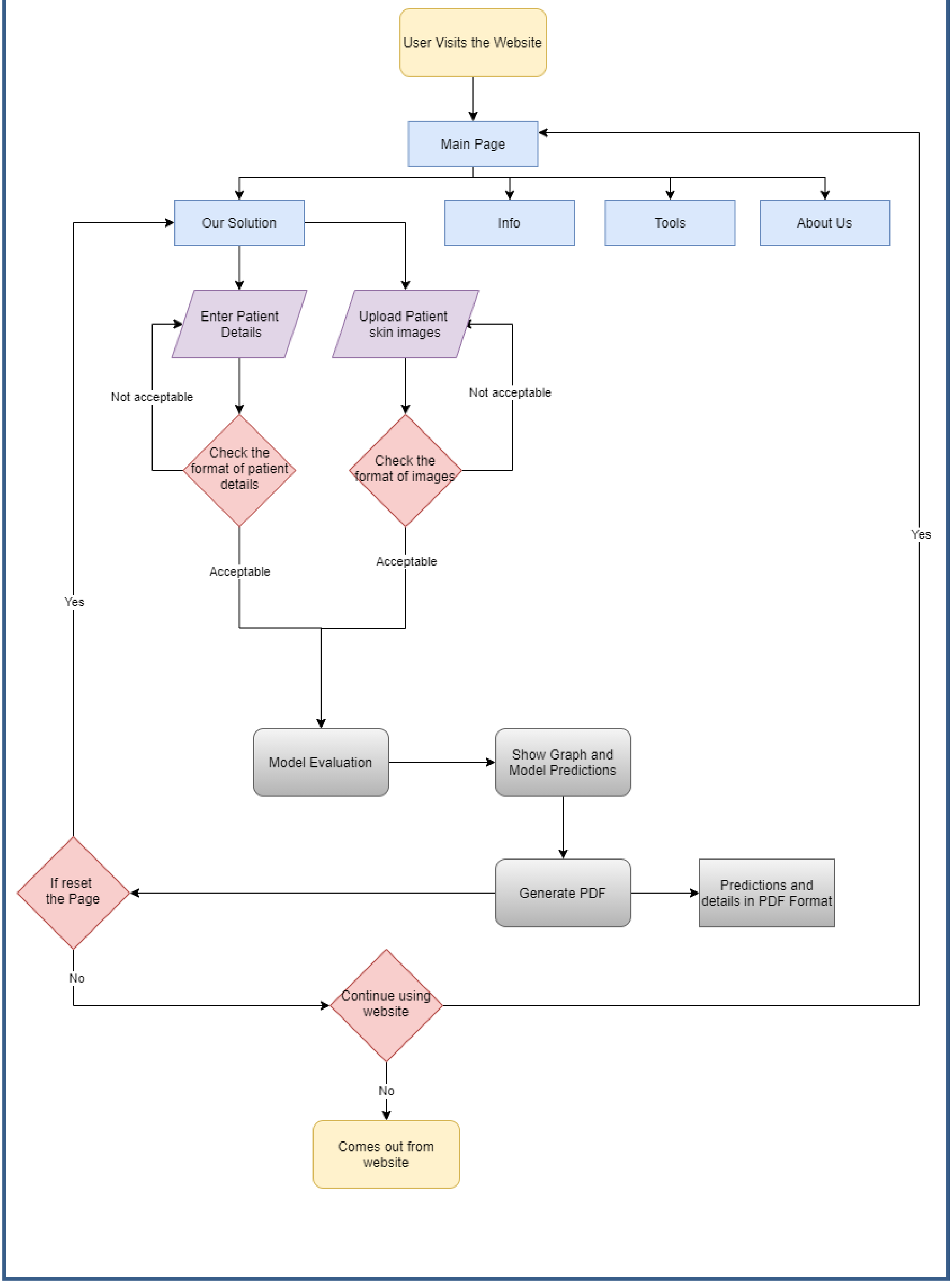
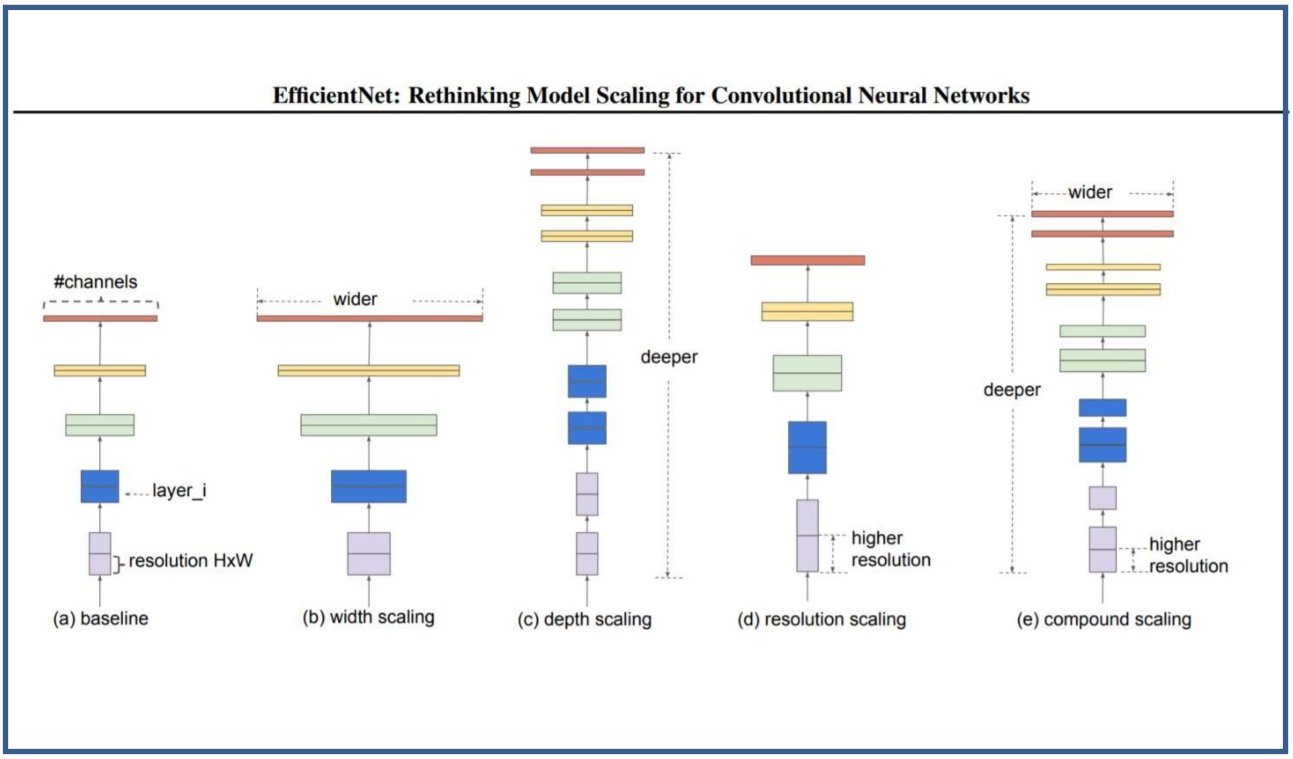


Figure 3: Web UI flow diagram

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

CNN Architecture Design:

The project aims to classify skin cancer using skin lesions images. To achieve higher accuracy and results on the classification task, we have used various Efficient Net models.Transfer learning is applied to the Efficien Net models. **We have unfrozen all the layer except Batch Normalization to stop the Batch Normalization layer from updating its means and variance statistics. If we train the Batch Normalization layer, it will destroywhat the model has leared, and accuracy will significantly reduce.**



GUI Design

To tackle the challenge of identifying skin cancer from skin lesions, we have to build a predictive model for **Computer-Aided Diagnosis (CAD)**. Taking the skin lesions image andpatient demographic information as input, we have developed a prototype web application that can help dermatologists interpret skin lesion images.

The web GUI consists of five main pages, of which four of them are used to explain the benefit of using the tool and way to reduce the death caused by skin cancer. The inference page named **"Our Solution"** is where the inference is performed using ensemblemethodology.

Conclusion

- One of the deadliest cancer forms is melanoma, and the proportion of people getting affected by melanoma is increasing rapidly. To make the solution availableto the public and dermatologists, we have successfully integrated the optimised model with our CAD system.

- The EfficientNet model is proved to be a better network for the skin cancer dataset. The network can generalise well on the dataset and have higher validationaccuracy (Table 3 and Figure 22). Plus, the ensemble of the model helps to reduce model prediction error and biases. The model prediction error can be further reduced if the ensemble is more significant with varied configuration, as proposed in Table 4.

- Along with optimising the training process, an equal amount of time is spent optimising the predictions. Based on the three core pillars of model serving, we have tick two of them: model size and latency. The last pillar (Prediction throughput) comes into account when the predictions are performed online over the internet. The prediction throughput measures how many predictions the system can perform in a given timeframe. The prediction throughput is beyond theproject's scope but should be considered when deploying the model on the web.

References:

SIC. (2018). Skin Lesion Analysis Towards Melanoma Detection. Retrieved March 20, 2021, from https://challenge2018.isic-archive.com/

ISIC. (2019). Skin Lesion Analysis Towards Melanoma Detection. Retrieved March 20, 2021, from https://challenge2019.isic-archive.com/

Mingxing, T., & Quoc, L. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. https://arxiv.org/abs/1905.11946

SIIM-ISIC Melanoma Classification. (2020). Identify melanoma in lesion images. RetrievedMarch 20, 2021, from https://www.kaggle.com/c/siim-isic-melanoma-classification

Tong, H., Zhi, Z., Hang, Z., Zhongyue, Z., Junyuan, X., Mu, L. (2018). Bag of Tricks forImage Classification with Convolutional Neural Networks. https://arxiv.org/abs/1812.01187

Qishen, H., Bo, L., Fuxu L. (2020). Identifying Melanoma Images using EfficientNetEnsemble: Winning Solution to the SIIM-ISIC Melanoma Classification Challenge. https://arxiv.org/abs/2010.05351