Classification of Melanoma Skin Cancer from Melanocyte Cell Images using Transfer Learning Techniques

MSc Research Project
Data Analytics

Soumya Nayak
Student ID: x21121427

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Soumya Nayak
Student ID:	x21121427
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Vladimir Milosavljevic
Submission Due Date:	01/02/2023
Project Title:	Classification of Melanoma Skin Cancer from Melanocyte Cell Images using Transfer Learning Techniques
Word Count:	6602
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Soumya Nayak
Date:	1st February 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Classification of Melanoma Skin Cancer from Melanocyte Cell Images using Transfer Learning Techniques

Soumya Nayak
x21121427

Abstract

Melanoma skin cancer, one of the most extreme skin diseases plaguing the world and a reason for large fatality rate has been the subject of broad research for many years. In order to identify Melanoma skin cancer, many Deep Learning approaches have given significant predictions. Since all the classes of skin lesion images have almost identical symmetry and optics, it is highly challenging to obtain accurate results and predictions, and this demands for a significant improvement in existing models. With the idea to enhance these results overlying in existing methods, this research focuses on implementing pre-trained transfer learning models for Melanoma skin cancer classification. Three different transfer learning models MobileNetV2, InceptionV3 and DenseNet201 were implemented by fine tuning the models to enhance the overall performance and make true predictions. The performance of all three models were compared and results shows that DenseNet201 outperformed MobileNetV2 and InceptionV3 models by achieving test accuracy and sensitivity of 68.30%.

1 Introduction

1.1 Background and Motivation

Till date, cancer possess the biggest challenge to medical science all over the world and has highest ratio of death cases among all other diseases. As per report provided by World Health Organization in 2020, almost 10 millions deaths were caused by cancer. As per (Younis et al.; 2019), genetic abnormalities turn out to be the most significant factor while environmental conditions and individual lifestyle also add to the development of this disease. Among all types of cancer, skin cancer is the most common to occur and also the easiest to ignore the symptoms. Among many forms of cancer, a special type of cancer that is skin cancer caused by turbulent growth of melanocyte cell images on different parts of body skin is the most major concern to deal with. The major cause of this is exposure to harmful ultraviolet radiations.
According to Global statistics stated in (Fikile Gasa et al.; 2020), records shows that every year around 8.2 million deaths and 14.1 million new cases across worldwide are caused due to skin cancer diseases. 95% of the cases can be cured by providing proper treatment if the skin cancer and its related symptoms are identified and diagnosed at early stages (Adegun and Viriri; 2020). In addition to requirement of bright and clear skin images

to identify the lesions, the manual technique of melanoma diagnosis involves procedures that are time-consuming and less accurate due to the lesions irregular forms and varied appearances. The basic methodology for melanoma identification and its classification to identify its class uses the ABCDE technique, however this approach is still prone to human error. This results in false or inaccurate predictions, which are then followed by incorrect or no therapy.

Various artificial intelligence and machine learning techniques have been used in the past to detect and categorize melanoma skin cancer from the images of melanocyte cells in an effort to overcome the difficulties and challenges associated with the manual approach. Several methods like Support Vector Machine, neural networks, multilayer perceptron, Dilated Residual Networks and many others were implemented in the past, although it was difficult to achieve 100% accuracy with true predictions using any of the approaches. Since medical science is very vast but sensitive area, therefore getting accurate results and true predictions especially in case of cancer is the major concern and more research needs to be done in this area to improve the performance of the models. With the aim to improve the accuracy and efficiency of prediction, this research work focuses on implementing various pre-trained transfer learning models like MobileNetV2, InceptionV3 and DenseNet201 to classify melanoma skin cancer from melanocyte cell images.

1.2 Problem Statement and Research Question

Manual approach, machine learning and deep learning techniques have provided good performance up to certain extent but still there is some room for improvement due to the reason that these approaches are vulnerable to false predictions. Therefore, there is need to implement such model that can prevent the prediction of false positive and false negative cases.

This research work focuses to answer and analyze the following research question:

How well pre-trained transfer learning models over deep learning algorithms can improve the overall efficiency and performance for the classification of Melanoma skin cancer from Melanocyte cell images?

1.3 Objective

The goal of this research work is to use Transfer Learning techniques to decrease the frequency of False Negative (FN) and False Positive (FP) outputs and simultaneously increasing the model's accuracy and sensitivity by accurate predictions of true cases.

1.4 Structure of the Report

The structure of this research project follows the following sequence: A brief introduction about the research topic, the challenges faced in manual approaches, research question and the objective for carrying out this work is described initially. Section 2 describes the related work done in the past using similar approaches for different kind of classification problems. Section 3 discusses about the research methodology followed to accomplish the research objective. Section 4 and Section 5 describes the design specification and Implementation. Section 6 describes the evaluation methods and metrics used to evaluate model performance and the last Section 7 deals with Conclusion and Future work associated with the research work.

2 Related Work

Medical science is a broad area where many researches have been done in the past to enhance the results obtained using manual approach and identifying automated systems that can produce results with higher accuracy and precision. One of the main issue in this field is medical image classification where a model is trained in such a way that significant results can be obtained and classify the image correctly. Skin cancer is one of the serious issue where classification of melanoma skin cancer from different categories of skin lesion images is a challenging task. On the basis of literature review, few papers related to melanoma classification and other image classification problems using different transfer learning techniques are analysed and considered the base for carrying out this research work.

2.1 Traditional Way of Skin Cancer Classification

Treatment and diagnosis of skin diseases is one of the most major concern. To carry out this a diagnosis system to identify skin related diseases was developed by (Amarathunga et al.; 2015) by employing the use of image processing technologies and data mining techniques. For this, the user needs to upload the image of the skin in the proposed system and needs to answer some basis questions related to their skin problem and symptoms experienced. The image will go through different preprocessing steps to generate an enhanced image by eliminating any noise associated with it. The system was able to identify the class of disease with 85%, 95% and 85% of accuracy belonging to Eczema, Impetigo and Melanoma class respectively. In addition to this, the system was also able to suggest some possible treatments associated with the class.

A digital system for diagnosing skin cancer based on ABCD protocol was presented by (Zapirain et al.; 2009). ABCD is a medical protocol that assess the characteristics like Asymmetry, Border Irregularity, Color and Diameter of infected area. For digital processing of image, an automatic algorithm called Isodata algorithm was developed to identify and segment the images. However traditional ways for skin cancer classification and segmentation are still prone to false positive and false negative cases which results in either wrong treatment or death, but this system was able to identify the correct and accurate results for the dataset containing 64 images.

Broadly there are mainly 3 types of skin cancer - Squamous cell carcinoma, Basal cell carcinoma and Melonoma among which melanoma is the most serious type of cancer which should be diagnosed at earliest stage for proper treatment. (Shalu and Kamboj; 2018) proposed an automated system for melanoma detection and classification using MED-NODE dataset consisting of 170 digital images with 2 classes benign and melanoma. The whole methodology includes 4 steps starting from image pre-processing followed by segmentation and feather extraction and then classification. For image pre-processing, thresholding algorithm and Dull-Razor technique was used. After carrying out first 3 steps, the performance of system was tested and evaluated using three different classifiers namely Decision Tree, K-Nearest Neighbour and Naïve Bayes. The classification accuracy was considered as the evaluation metrics and results shows that Decision Tree achieved the highest accuracy as 82.35% as compared to Naïve Bayes with 75.88% and KNN with 75.88%.

2.2 Transfer Learning Approaches for Image Classification

MobileNetV2

Convolutional neural network needs huge memory storage and a very high computation capability to detect skin cancer using smartphone cameras. To overcome this challenge (Hartanto and Wibowo; 2020) developed an android phone application for early detection of skin cancer using Faster R-CNN method and MobileNetV2 method. The methods were tested for skin cancer classification on a dataset containing 600 images with two different classes as melanoma and actinic keratosis. While testing the results on android application it was observed that MobileNetV2 model performed quite well with the overall accuracy of 86.3% at 0.005 learning rate.

A MobileNetV2 model embedded with Efficient Channel Attention was proposed by (Cheng et al.; 2021) for skin cancer classification which can identify and recognize the images on mobile or portable devices. HAM10000 image dataset was considered to carry out the research which contains 10015 skin lesion images with 7 classes. The classes were highly unbalanced, so to achieve equal distribution of images among all the classes, image transformation like cropping, random brightness, flipping and rotating was carried out. Significant trends and results were obtained for training loss and accuracy. The improved ECA-MobileNetV2 model gets an accuracy of 82.3% and fast recognition rate with less complex training model.

A multi class image classification model for the proper detection and appropriate use of face mask due to spreading of COVID-19 was proposed by (Rokhana et al.; 2021) using MobileNetV2 architecture. The analysis was done on a dataset containing 450 images with 3 classes which were correct use of face mask, incorrect use of face mask and no face mask. Softmax activation function and Adam's optimizer was used while training the model. The final model achieved the outstanding and efficient results with accuracy and f1-score as 97% in 265.94 seconds which is lesser as compared to other image classification models. The model was also tested for real-time using web-camera.

(Dong et al.; 2020) performed a comparative study analysis on two different image classification datasets to evaluate and compare the performance of Transfer learning MobileNetV1 and MobileNetV2 models. First dataset was Colorectal histology dataset with 8 classes to detect cancer and second was Sentinel satellite images dataset with 10 classes. After implementing both of the MobileNet model on each of the dataset, it was noticed that MobileNetV2 outperformed with better accuracy and training time.

Systemic sclerosis is rare but serious form of skin disease which if diagnosed in initial stage can be cured easily. A classification model using MobileNetV2 was proposed by (Akay et al.; 2021) for embedded mobile applications to classify SSc skin images and the performance was compared with state-of-the-art CNN model where MobileNetV2 outperformed over CNN. Both the models were tested on 2 datasets, one dataset with 2 classes normal verses SSc images and other dataset with 3 classes normal, early and late SSc images. Images were preprocessed using different augmentation techniques like cropping, flipping and enlarging. After fine tuning the model, MobileNetV2 achieved significant results over CNN model with 82.9% and 94.8% accuracy for first and second dataset respectively with lower computation time.

Early and proper diagnosis of Diabetic Retinopathy can prevent loss of vision. (Pamadi et al.; 2022) proposed two CNN based transfer learning MobileNetV2 models, one for binary and other for multinomial classification with 5 classes. GaussianBlur and addWeighted function was used for image preprocessing, also image augmentation was performed due to small dataset with limited images to achieve high accuracy. Adam's optimizer was used to reduce training loss. The final MobileNetV2 model was proved to be the best model when compared with 3 other models namely MobileNet, DenseNet169 and DenseNet121 with overall accuracy as 78% and 97% for multinomial and binomial classification respectively.

InceptionV3

Melanoma is one of the most dangerous type of skin cancer that can spread in different parts of body skin and must be diagnosed in early stages since delay in proper treatment can lead to death as well. The study proposed in (Sultana et al.; 2022) shows how well a transfer learning model using InceptionV3 can classify the skin cancer into 3 different classes. The dataset was pre-processed and the model was trained using Adams optimizer with batch size as 32 for 10 epochs. The model was tested on two different tools that is jupyter notebook and google colab with test accuracy obtained as 77.25% and 96.19% respectively. The computation time for the model was just 5 minutes which can be used easily for real time application by taking the dermoscopic image of skin lesion and using the model to identify and classify the image as melanoma or non-melanoma.

A transfer learning model using Google's Inception V3 was proposed by (Chang et al.; 2017) to classify and detect breast cancer. A BreaKHis dataset was used for the study with total 7909 biopsy images consisting of 2 classes labelled as benign with 2480 images and malignant with 5429 images. Due to insufficient amount of data for training the model, various data augmentation techniques like rotation, mirrored images and random image distortion was used and supplemented in the original dataset. The graph obtained for training accuracy and cross entropy shows significant results with accurate trends between training and validation data. Also, significant results were obtained for each of the classes for different cut-off values in the classification report. The InceptionV3 model produced accurate results to classify breast cancer with AUC value as 0.93

To maintain a healthy lifestyle, keeping track of eating habbits is one of the most essential thing. To analyze this (Rajayogi et al.; 2019) proposed various pre-trained transfer learning models like VGG19, VGG19, Google's InceptionV3 and ResNet. An Indian food dataset was taken for the experiment containing 20 classes with each class having 500 images. The entire dataset was splitted into training and testing set containing 8000 and 2000 sample images respectively. To minimize the loss function, stochastic gradient descent optimizer and Catergorical cross entropy function was used. To prevent overfitting dropout value of 0.2 was used and the model was trained in a batch of 32 for 30 epochs. The accuracy achieved on the test data was 69.91% for ResNet, 78.2% for VGG16, 78.8% for VGG19 and 87.9% for InceptionV3. InceptionV3 outperformed over other pre-trained transfer learning models providing higher accuracy to classify the food images.

A pre-trained transfer learning model using InceptionV3 was implemented by (Li et al.; 2018) to classify colorectal cancer based on MRI images. The data was obtained from Harbin Medical Cancer hospital with total of 619 images including 2 classes. Due to insufficient size of original data, data augmentation was performed. Back propagation algorithm was used to update the required parameters and also cross entropy function was used to decrease the loss during model training. The final model containing augmented images merged with original dataset attained the test accuracy of 94%.

(Nabil et al.; 2021) proposed an InceptionV3 model to classify the shape of face which can be of great benefit in fashion industry and beauty industry for hairstyle recommendation. CelebA dataset was selected for the analysis containing 5 different classes representing different shapes with 500 images belonging to each class. Another dataset with 300 images per class was selected and combined to train the model with better accuracy. The accuracy and loss graphs were significant with 94.9% accuracy.

DenseNet201

Early detection of melanoma, one of the most serious type of skin cancer at initial stage can raise the possibility of healing up to 99.2% survival rate which can be very time consuming if done by traditional approach. The study done in (Waweru et al.; 2020) proposes a DCNN based DenseNet201 model that can automatically detect melanoma from dermoscopy skin lesion images. The dataset taken for this was HAM10000 dataset containing 7 different classes which was splitted 60% for training and 20% each for validation and testing. Training the model for 150 epochs and altering the learning rate from 0.0001 to 0.01 after 2 epochs, they were able to achieve test accuracy of 86%. In addition they followed the approach of back propagation using cross entropy loss and adaptive momentum for optimizing the algorithm. Deployment of the model was done using Flask APP which is considered as python web framework.

Due to diverse characteristics of tumor cells, classification of breast cancer tumors is bit complex and challenging task. (Djouima et al.; 2022) proposes Deep Convolution GAN(Generative Adversial Network) based transfer learning DenseNet201 model to classify breast cancer tumors in 2 classes benign(minority class) and malignant(majority class). The considered BreaKHis dataset is highly imbalanced and therefore the data was balanced using data augmentation to produce significant results for both the classes. The model was trained using RMSprop optimizer with a learning rate of 0.0001 considering a batch of 35 for 40 epochs. Softmax was used as activation function and criteria of early stopping was implemented if there is no more improvement in model training results. The model was finally trained and evaluated for 4 magnification factors and yielding the test accuracies of 96% for 40x, 95% for 100x, 88% for 200x and 92% for 400x. The significant results clearly indicates that the proposed DenseNet201 model was very efficient in predicting the breast cancer classification.

A comparative analysis for brain tumor classification was done by (Sahaai et al.; 2022) by utilizing various pre-trained models like DenseNet201, Xception, VGG19 and InceptionV3. A dataset containing MRI images of brain tumor was selected consisting of 2 classes cancerous and non-cancerous. Various evaluation metrics like accuracy, precision, recall and f1-score were considered and compare and evaluate the performance of all the models. Each of the model was executed and their accuracy was compared for 10, 20 and 30 epochs. Results shows that DenseNet201 model outperformed over Xception, VGG19 and InceptionV3 by achieving the accuracy of 86.89%, 90.42% and 91.94% for 10, 20 and 30 epochs respectively.

Application of artificial intelligence in medical domain has transformed the manual approach of examining and diagonosing the disease. Different machine learning approaches have been used in the past but due to the complex process of feature extraction, researchers are now moving towards deep and transfer learning approaches. One such approach was followed by (Baldota et al.; 2021) by proposing a deep transfer learning model based on DenseNet201 to classify pancreatic cancer. The dataset consists of ultrasound pancreatic images with 3 classes cancerous, pancreatitis and healthy. The performance of DenseNet201 was analysed and compared with various other transfer learning models like DenseNet121, AlexNet, ResNet152, MobileNetV2, VGG19 and SqueezeNet1_1 with

DenseNet201 achieving highest accuracy of 99.93% and sensitivity as 99.87% which proves it as the best model to classify Pancreatic cancer.

(Rukhsar and Upadhyay; 2022) proposed a DenseNet201 model using transfer learning technique to classify and detect rice leaves disease among 3 different classes that are virus-caused tungro, bacteria-caused bacterial leaf blight and fungus-caused blast. A simple CNN model was also applied on same dataset to evaluate the performance of both the models. Results shows that DenseNet201 yielded accuracy of 96.09% which was comparatively very high as compared to accuracy of simple CNN model with 62.20% accuracy.

3 Research Methodology

A fundamental approach of KDD (Knowledge discovery in databases) methodology is followed in this project keeping in mind the objective of this research that how well a pre-trained transfer learning model can identify the skin cancer from dermoscopic skin lesion images and classify the type of skin lesion among 8 classes. It is a systematic process to analyse the raw data, extract knowledge from data, build and fine tune a model with its proper testing and performance evaluation.

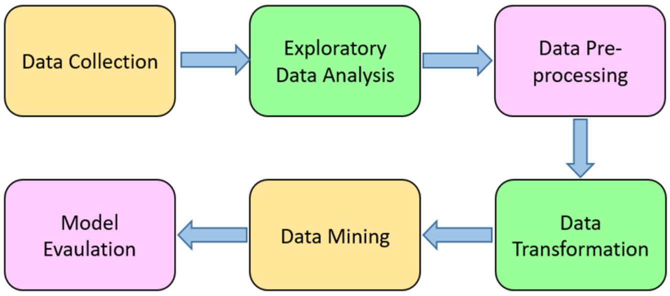

Figure 1: KDD Methodology for Melanoma Classification

3.1 Data Collection and Description

The dataset selected to carry out this thesis is downloaded from kaggle from ISIC (International Skin Imaging Collaboration) 2019 Skin Lesion images for classification which is publicly available dataset[1]. The dataset comprises of total 25,331 images belonging to 8 different classes. The data available in ISIC 2019 Skin Lesion images for classification dataset also includes training data samples from previous years that is ISIC challenge 2017 and 2018. The eight classes in the dataset includes AK (Actinic keratosis), BCC (Basal cell carcinoma),BKL (Benign keratosis), DF (Dermatofibroma), MEL (Melanoma), NV (Melanocytic nevus), SCC (Squamous cell carcinoma) and VASC (Vascular lesion). Generally all these classes lie in either of the 2 categories benign which is non-cancerous and malignant which is cancerous.

Actinic keratosis appears in the form of dry skin patches generally caused due to exposure to UV rays, Benign keratosis appears mostly due to skin aging, Dermatofibroma which is caused due to overgrowth of various cell type in skin's dermis layer, Nevus caused due to overgrowth of melanocytes cells in cluster and vascular benign are some of the classes

[1] https://www.kaggle.com/datasets/salviohexia/isic-2019-skin-lesion-images-for-classification

of skin lesion which falls in benign category and if treated properly on time may not lead to any serious cause. On the other hand Melanoma, Squamous cell carcinoma and Basal cell carcinoma which is the most serious form of skin cancer caused due to UV exposure comes under malignant category which if not diagnosed and treated at early stage can lead to death. Sample of skin lesion images of each class is shown in Figure 2.

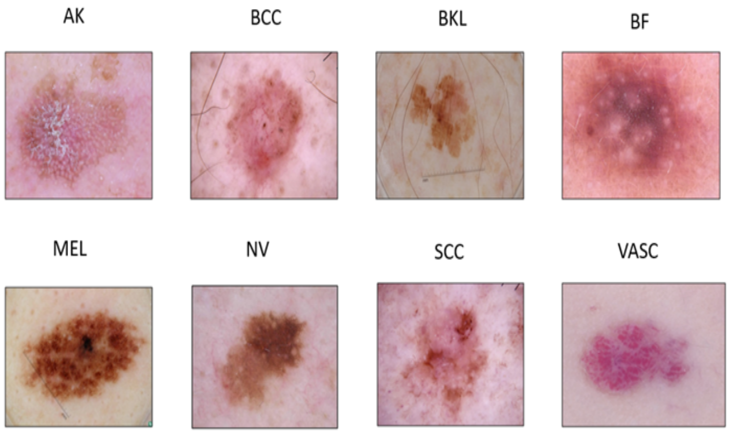

Figure 2: Sample image of each class

3.2 Exploratory Data Analysis

For the better understanding of skin lesion dataset, exploratory data analysis was done. EDA is usually performed to analyse the data in-depth and more precisely before reaching to any conclusions to make sure that the results produced after applying the proposed methodology are correct and can be applied in real time applications to meet the business requirements and objective of proposal. While analysing the dataset, several findings were made. Figure 3 shows the frequency of skin lesion in different regions of skin where anterior torso being the most affected part with count approximately 7000 and lateral torso being the least with count less than 200. This clearly indicates that front body parts like abdomen, chest and pelvis, thighs and head/next are more prone to lesion while palms, toes and side of the body are less likely to be affected by any type of skin lesion.

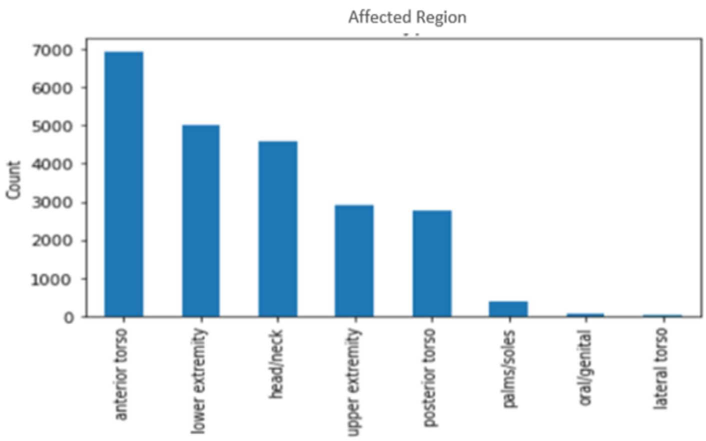

Figure 3: Affected region wise sample distribution

Figure 4 shows the ratio of skin lesion with respect to male and female. According to the data obtained from ISIC 2019 skin lesion images for classification, it can be clearly observed from the statistics that there is not much significant difference between affected male and female. Approximately 13000 males and 11000 females are exposed to skin lesions.

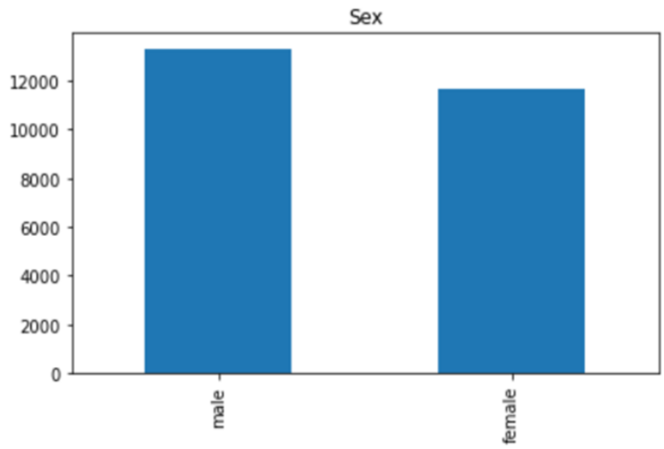

Figure 4: Sample distribution by gender

While analysing the data to classify melanoma skin cancer through lesion images belonging to 8 different classes, age is one of the most important parameter to analyse which age group is more prone to be affected by cancer. Figure 5 shows the statistics indicating that the skin lesions are experienced in both males and females in the age group between 40 and 70.

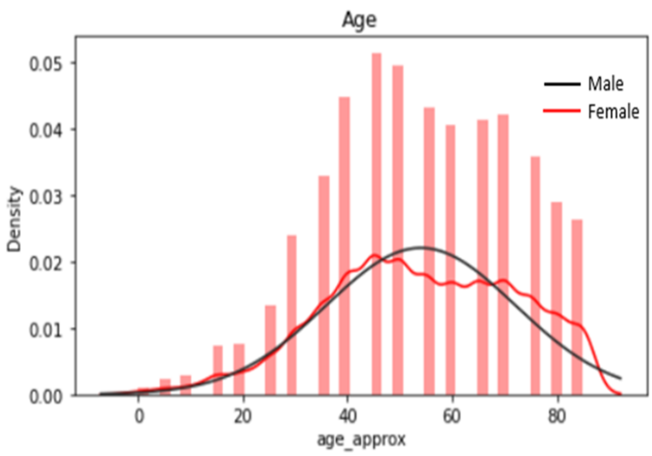

Figure 5: Density distribution for different age groups

3.3 Data Pre-processing

In the proposed research, three different pre-trained transfer learning models MobileNetV2, InceptionV3 and DenseNet201 are implemented among which MobileNetV2 uses some pre-processing techniques while InceptionV3 and DenseNet201 uses different approach for pre-processing. The dataset is highly imbalanced with 12,875 images in majority class (NV) and 239 images in minority class (DF). For MobileNetV2, dataframes are created for each of the 8 different classes and majority classes are downsampled by reducing the class size to 2500. Figure 6 shows the class size after downsampling the majority classes.

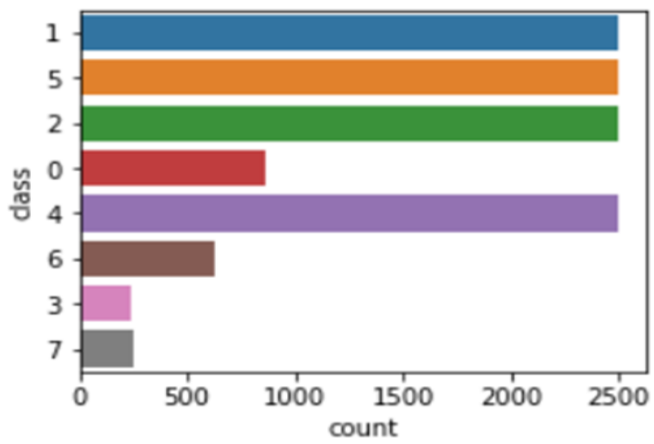

Figure 6: Class wise distribution after downsampling majority classes

The dataset was further splitted into 80% training and 20% test set. A RandomOverSampler() function was used on training dataset for upsampling the minority class to balance the dataset for training the final model.
In case of InceptionV3 and DenseNet201, to train the model, trim() function is used to trim the training dataframe so that no class will have more that maximum sample images. Considering the maximum sample size as 800 and minimum sample size as 227, the dataframe was trimmed and then the length of trimmed dataframe was 5063 images with 8 classes.

3.4 Data Transformation

Data transformation is one of the most important step in machine learning. To improve the quality of data and results associated with it, various data transformation techniques are used. Data augmentation is one of the most common and suitable approach to balance and increase the overall size of data for producing accurate results. Also the data is coverted in such formats which are compatible for different systems. For MobileNetV2 model, Keras ImageDataGenerator is used to create the batches of tensor images for the purpose of augmentation. Also, flow_from_dataframe() method is used to generate test, train and validate dataset in such a format that will be compatible to MobileNetV2 model. It takes dataframe as input and generates batches of augmented images. For InceptionV3 and DenseNet201 model, Keras ImageDataGenerator method is used to perform data augmentation using various techniques like horizontal flip, rotation, zoom range and width/height shift range.

3.5 Data Mining

For melanoma skin cancer classification and various other image classification problems, different machine learning and deep learning models have been implemented in the past. After carefully analysing different kinds of approaches opted for the image classification with reference to literature review, it was observed that transfer learning models produces best results for classification problems especially in the field of medical science since true and accurate results are of major ethical concern in this domain. This work focuses on implementing three different transfer learning models namely MobileNetV2, InceptionV3 and DenseNet201 for melanoma classification and determine the best model for this problem.

3.6 Model Evaluation

The performance of all the three models is evaluated based on the test accuracy obtained after training the transfer learning model, training and validation loss versus epoch, training and validation accuracy versus epoch and sensitivity (weighted average recall) obtained in classification report of each model. The accuracy and sensitivity can be calculated using the equations (1) and (2)

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

where,
TP (True Positives) means true prediction of positive cases
TN (True Negatives) means true prediction of negative cases
FP (False Positives) means false prediction of positive cases
FN (False Negatives) means false prediction of negative cases

4 Design Specification

A 3-tier design framework as shown in Figure 7 has been used to accomplish the proposed research for the classification of Melanoma from melanocyte cell images using transfer learning approach. Data layer is a layer in which the data stored at various sources can be accessed and used for pre-processing. Here, the dataset which is openly available on Kaggle is fetched and loaded in Google Colab pro, kaggle notebook and Jupyter notebook to perform exploratory data analysis, data pre-processing and transformation. Among these colab and kaggle uses GPU to speed up the performance of machine learning models. In Business logic layer, pre-trained transfer learning models are used to train the model followed by evaluation layer to assess the result of final model based on different evaluation metrices.

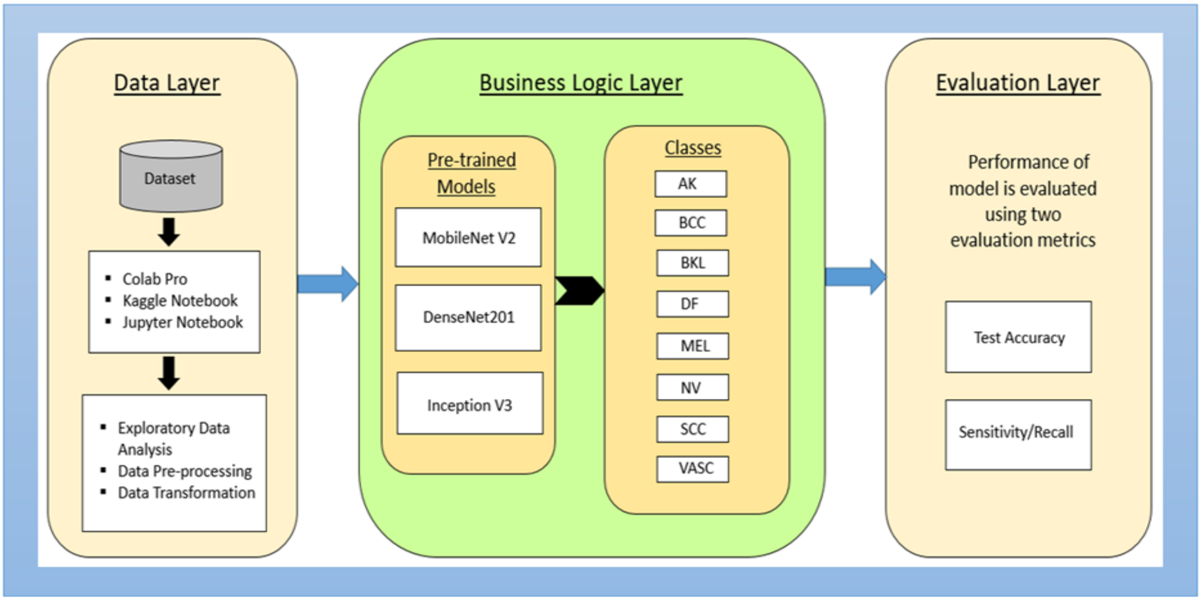

Figure 7: 3-tier Design Framework

4.1 Transfer Learning Approach

Transfer learning is a machine learning based research approach which is built over imagenet dataset that consists of 1000 different image classes using CNN architecture that aims to focuses on the idea of preserving the knowledge acquired while addressing one problem and its application to another analogous problem in related domain. Gathering a large dataset and analyzing it to the core to accomplish a new task can be a challenging task. Also, using small amount of training data to obtain a model with significant results can be another challenge. To overcome these issues, transfer learning is proved to be the best approach when compared to other methods. Transfer learning offers the advantage of minimizing model's training time and potentially lowering generalization error. Transfer learning follows the approach of weight initialization scheme since the weights in previously utilized layers could serve as the basis for training the model and then be modified to address the new issue. To fine tune the transfer learning models, additional dense, dropout and output layers can be added during model building step based on the type and size of dataset used.

4.2 Convolutional Neural Network

A deep neural network class known as CNN is most often used to analyse and evaluate visual imagery. A ConvNet requires substantially less pre-processing when compared to other classification techniques. The multi-level representation is the core idea of CNN performance with higher level features that provides more information about the data. Through the use of filters, a ConvNet may effectively seize the spatial and temporal dependencies in a visual image. Due to the CNN architecture's ability to reuse weights with reduced parameters, the image dataset is the best fit for it. A typical convolutional neural network architecture consists of an input and an output layer with one or more hidden layers as shown in Figure 8[2]. These hidden layers can be several convolutional,

[2]https://vitalflux.com/wp-content/uploads/2022/04/Typical-CNN-architecture.png

max pooling, normalization or fully connected layers.

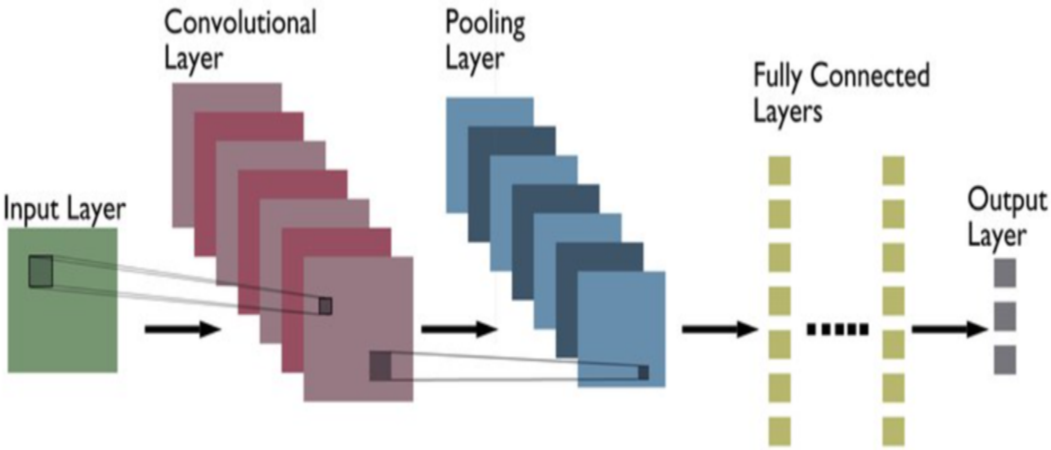

Figure 8: CNN Architecture

5 Implementation

The implementation of the entire proposal is achieved by using various technologies shown in Figure 9 which are discussed in this section. Python is the programming language which is used for this work. Due to the availability of wide range of libraries, programming simplicity and consistency, platform independence and access for various machine learning frameworks, python was used for entire implementation. Anaconda was used as an base platform to develop all the transfer learning models on Jupyter Notebook. Due to the large size of selected dataset which is 10GB, kaggle notebook is used to directly add kaggle dataset in the session and execute MobileNetV2 model. Google colab pro was opted to execute Inception V3 and DenseNet201 model since the execution of these two models was comparatively faster in colab when compared with Jupyter and Kaggle Notebook. MatplotLib library was imported and used to carry out Exploratary Data Analysis by creating visualisations to gain better insight of raw data. Keras Application programming interface which is basically a high level API of Tensorflow library is used for building all the three MobileNetV2, InceptionV3 and DenseNet201 models.

Figure 9: Technologies Used

The setup configuration as shown in Table 1 is required for the implementation.

Table 1: Setup Configuration

IDE	Google Colab Pro,Jupyter Notebook, Kaggle Notebook
Programming language	Python
Framework	Tensorflow
Modules	Matplotlib, Pandas, Numpy Seaborn, Scikit-learn, cv2, imblearn
Computation	GPU
Number of GPU	1
Type	Tesla P100-PCIE-16GB

5.1 Implementation of MobileNetV2 Model

MobileNetV2 is CNN based 53 layered transfer learning model which is basically designed to operate for mobile applications. Unlike MobileNet model, the foundation of MobileNetV2 is based on inverted residual structure which means the input for residual block and output obtained from it as a whole act as thin bottleneck layers to reduce the number of parameters and increase the efficiency of model. In order to remove non-linearity, lightweight depthwise convolutions are used by intermediate expansion layer.

For implementing MobileNetV2 model, initially all the required packages and libraries were imported. Dictionary was created for each of the class followed by downsampling of majority classes. The dataset was splitted into training and test set and then upsampling was performed on training data to balance the data for model building. The dataframes were converted into mobilenet compatible input format and data transformation was carried out. The model was build using pre-trained transfer learning MobileNetV2 model. For fine-tuning and improving the performance of model, 2 more additional dense layers were added with 128 nodes and 'relu' activation function in each layer. Here Rectified Linear Unit is used to transform the weighted input into output of that particular layer. As of now, it is considered the best activation function since the exponential growth that is required for the computation of neural network can be prevented using this. To prevent overfitting, a dropout layer with only 50 percent of active neurons is added. At last a dense layer with 8 nodes and 'softmax' activation function is added to store the output for making desired predictions. Softmax function is used here in output layer for the prediction of multiclass probability distribution. SparseCategoricalCrossentropy() is the loss function which is used while training and the parameter from_logits is set to false to avoid wastage of computation since a softmax output layer is already added while building the model. To reduce the loss by automatically adapting a learning rate for each particular input value, adam optimizer is used. The final model is trained for 30 epochs to evaluate the results on test dataset.

5.2 Implementation of DenseNet201 Model

DenseNet201 is 201 layers deep CNN architecture based model that belongs to DenseNet group. DenseNet models have several merits over other state-of-the-art models like vanishing gradient reduction, better feature propagation, model reuse and limited amount of trainable parameters (Huang et al.; 2017). Moreover, by utilizing limited amount of memory and computing power DenseNet models achieve significant results with high performance.

For implementing DenseNet201 model, after importing all the required libraries and performing data pre-processing by trimming the dataset so that none of the class will have more images than maximum sample images, data transformation by using various augmentation techniques such as flipping, rotation, zooming and many more in order to achieve balanced data, the dataset was splitted into test, train and validation sets. A DenseNet201 model using imagenet weights was created using keras API. Batch normalisation was used to speed up the model training process, additionally a dense layer with 256 modes, a dropout layer with 60% active neurons and an output dense layer with softmax as the activation function was added to fine tune the overall model. In Dense layer, to prevent overfitting relu activation function and various type of regularizers along with their suitable l1 and l2 parameter values like kernel regularizer to minimise weights, activity regularizer to minimise layer's output, bias regularizer to reduce bias was added. 'Categorical cross entropy' loss function with leaning rate 0.001 was used and the model was trained for 10 epochs. Moreover, Keras custom callback mechanism was also used in the approach to halt or continue the training process after desired number of epochs.

5.3 Implementation of InceptionV3 Model

Inception V3 is 48 layers deep well known CNN architecture based image classification model that uses Label Smoothing, auxillary classifier and Factorized 7 x 7 convolutions. For the implementation of Inception V3 model, similar approach as of DenseNet201 model was followed. Model was built by performing data pre-processing and desired transformation techniques, also few extra dense and dropout layer were added in order to increase the efficiency and performance of training model that can produce significant results on test data. Considering batch size as 20 and learning rate as 0.001, the model was trained for 10 epochs.

6 Evaluation

The purpose of this section is to provide a comprehensive analysis of the results and main findings of the study as well as the implications of these finding both from academic and practitioner perspective are presented. Only the most relevant results that support your research question and objectives shall be presented. Provide an in-depth and rigorous analysis of the results. Statistical tools should be used to critically evaluate and assess the experimental research outputs and levels of significance.

6.1 MobileNetV2 Results

The MobileNetV2 model was trained for 30 epochs followed by analyzing its training and validation accuracy at the end of 30th epoch. The Figure 10 clearly shows how the loss decreases and accuracy increases with the increase in each epoch. 77.7% and 55.7% of training and validation accuracy was obtained on the final trained model.

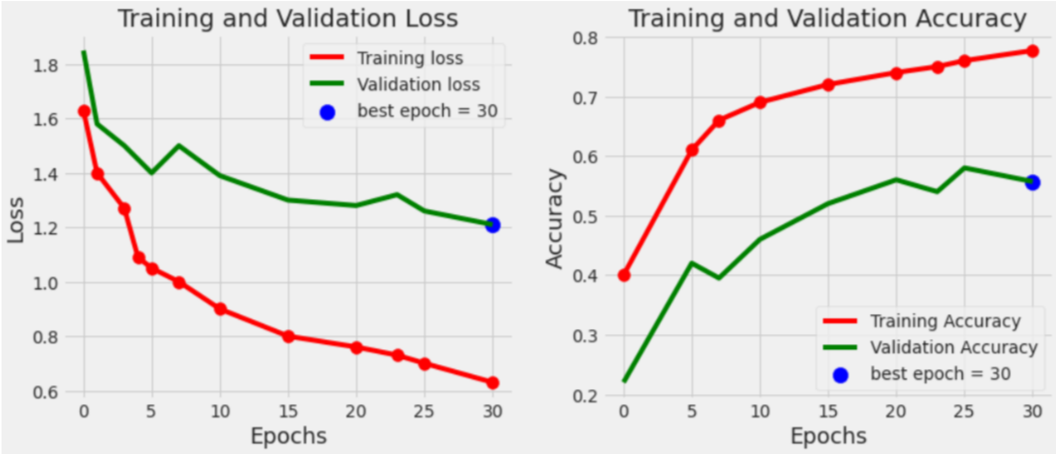

Figure 10: Learning Curves for MobileNetV2

Table 2 shows the classification report of MobileNetV2 model. The model obtained the test accuracy of 50%. Sensitivity or recall shows that for the classes NV and VASC, the model produced significant results that is it performed comparatively good for these classes to predict positive instances.

Table 2: Classfication Report for MobileNetV2

	Precision	**Recall**	**F1-score**
AK	0.29	0.57	0.38
BCC	0.58	0.50	0.54
BKL	0.48	0.47	0.47
DF	0.22	0.33	0.26
MEL	0.56	0.43	0.49
NV	0.68	0.59	0.63
SCC	0.28	0.44	0.34
VASC	0.45	0.66	0.54
Accuracy			0.50
Macro avg	0.44	0.50	0.46
Weighted avg	0.53	0.50	0.51

6.2 DenseNet201 Results

The DenseNet201 model was trained for 10 epochs followed by analyzing its training and validation accuracy at the end of 10th epoch. The Figure 11 clearly shows how the loss decreases till 8th epoch and attained a stable state and accuracy increases with the increase in each epoch. Also the minimum gap between training and validation loss curves shows a good model fit. 94% and 72% of training and validation accuracy was obtained on the final trained model.

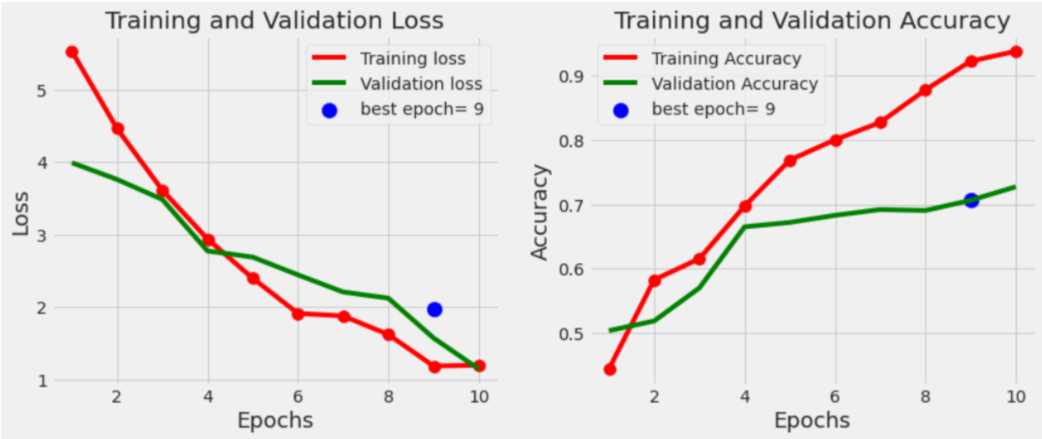

Figure 11: Learning Curves for DenseNet201

Table 3 shows the classification report of DenseNet201 model. The model obtained the test accuracy of 68.30%. Sensitivity or recall shows that for the classes AK, BCC, NV, MEL, SCC and VASC, the model produced significant results that is it performed comparatively good for these classes to predict positive instances.

Table 3: Classfication Report for DenseNet201

	Precision	**Recall**	**F1-score**
AK	0.36	0.77	0.50
BCC	0.72	0.81	0.76
BKL	0.56	0.46	0.51
DF	0.33	0.50	0.40
MEL	0.46	0.60	0.52
NV	0.88	0.72	0.79
SCC	0.50	0.43	0.46
VASC	0.66	1.00	0.80
Accuracy			0.68
Macro avg	0.56	0.66	0.59
Weighted avg	0.72	0.68	0.69

6.3 InceptionV3 Results

The Inception V3 model was trained for 10 epochs followed by analyzing its training and validation accuracy at the end of 10th epoch. The Figure 12 clearly shows the high fluctuation in training loss and accuracy between second and sixth epoch with the increase in each epoch. 96% and 68% of training and validation accuracy was obtained on the final trained model. Overfitting can be the reason for higher training accuracy.

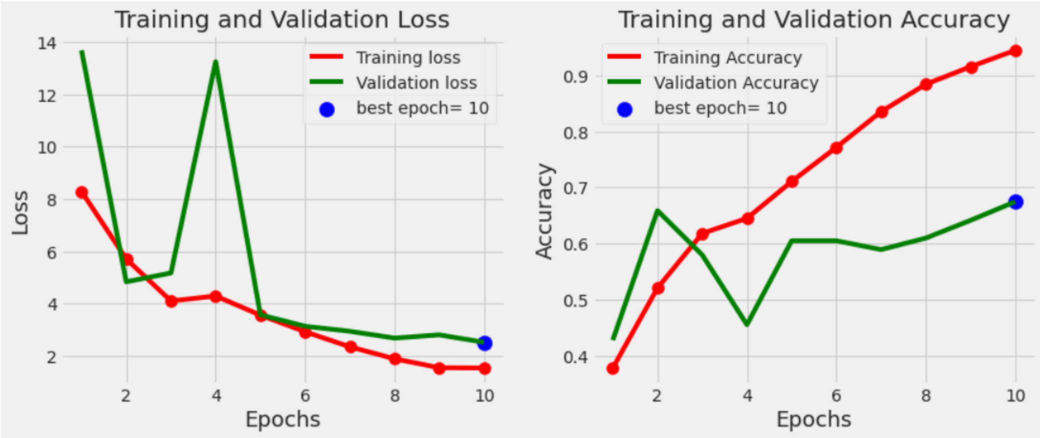

Figure 12: Learning Curves for InceptionV3

Table 4 shows the classification report of Inception V3 model. The model obtained the test accuracy of 64.04%. Sensitivity or recall shows that for the classes AK, BCC, NV and VASC, the model produced significant results that is it performed comparatively good for these classes to predict positive instances.

Table 4: Classfication Report for InceptionV3

	Precision	**Recall**	**F1-score**
AK	0.27	0.72	0.40
BCC	0.72	0.59	0.64
BKL	0.38	0.57	0.46
DF	0.00	0.00	0.00
MEL	0.57	0.50	0.53
NV	0.85	0.72	0.78
SCC	0.42	0.50	0.45
VASC	0.40	1.00	0.57
Accuracy			0.64
Macro avg	0.45	0.57	0.48
Weighted avg	0.69	0.64	0.65

6.4 Discussion

The research conducted in this work aims to classify melanoma skin cancer from melanocyte cell images belonging to eight different classes using 3 pre-trained transfer learning models and the results of the finals models are compared as shown in Table 5.

By analyzing the accuracy and learning curves of MobileNetV2 model, it is observed that this model achieved least test accuracy and sensitivity which is 0.50 and 0.50 respectively. Although the trends of the learning curves are significant without any overfitting or underfitting, but its less accuracy for test dataset shows its inability to predict positive results. InceptionV3 model achieved the overall test accuracy of 0.64 and sensitivity as 0.64 with lot of fluctuation in validation loss and accuracy. Batch size and the optimizer used while training can be one of the reason for this fluctuation. Results shows that

Table 5: Comparison of proposed Transfer learning models

Model	**Accuracy**	**Weighted average recall (or Sensitivity)**
MobileNetV2	0.50	0.50
InceptionV3	0.64	0.64
DenseNet201	0.68	0.68

DenseNet201 out-performed over other two models which test accuracy and sensitivity as 0.68. Moreover, its classification report shows that the model is capable of predicting significant results for 5 out of 8 classes. Learning curve shows minute overfitting due to higher training accuracy in early epochs which can be avoided using different optimization techniques but the good test accuracy shows its capability to predict accurate results and classify melanoma to higher extent.

7 Conclusion and Future Work

Melanoma Skin Cancer classification is one of the major research for which several researches and works have been already done in the past using different artificial intelligence and machine learning approaches and still this area requires more research to obtain improved results. The symmetry and semantics of the dermoscopic images were the most crucial discoveries. To remove unwanted artifacts from the given images, various image pre-processing and feature extraction techniques were adopted which enhanced the performance of the model. This research project emphasizes to implement different pre-trained transfer learning models which can significantly classify melanoma skin cancer from melanocyte cell images and identify true positives cases in order to prevent false positive and false negative results. Three transfer learning models MobileNetV2, InceptionV3 and DenseNet201 were implemented, out of which DenseNet201 achieved highest results with 68.3% test accuracy and sensitivity. The only limitation of this model was the higher training accuracy in the early epochs which could be due to overfitting of data. Downsampling the highly imbalanced dataset during model building step in order to achieve improved performance by balancing the data could be one of the possible reason for this.

The future work can be extended by Hyperparameter tuning of proposed transfer learning models using different optimization techniques. Moreover, the model can be trained by considering large training data and batch size. The pre-trained transfer learning models can be combined and an approach of ensemble model can be adopted to assess the performance of model. A novel technique of CapsuleNet that addresses feature relationships and require large training datasets can be implemented to evaluate the performance.

References

Adegun, A. A. and Viriri, S. (2020). Fcn-based densenet framework for automated detection and classification of skin lesions in dermoscopy images, *IEEE Access* **8**: 150377–150396.

Akay, M., Du, Y., Sershen, C. L., Wu, M., Chen, T. Y., Assassi, S., Mohan, C. and Akay, Y. M. (2021). Deep learning classification of systemic sclerosis skin using the mobilenetv2 model, *IEEE Open Journal of Engineering in Medicine and Biology* **2**: 104–110.

Amarathunga, A. A. L. C., Ellawala, E. P. W. C., Abeysekara, G. N. and Amalraj, C. R. J. (2015). Expert system for diagnosis of skin diseases, *INTERNATIONAL JOURNAL OF SCIENTIFIC TECHNOLOGY RESEARCH* **4**: 1.
URL: *www.ijstr.org*

Baldota, S., Sharma, S. and Malathy, C. (2021). Deep transfer learning for pancreatic cancer detection, *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–7.

Chang, J., Yu, J., Han, T., Chang, H.-j. and Park, E. (2017). A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer, *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1–4.

Cheng, Y. J., Lin, W., Liu, Y. Z. and Sun, L. (2021). Classification of skin diseases based on improved mobilenetv2, *2021 33rd Chinese Control and Decision Conference (CCDC)*, pp. 598–603.

Djouima, H., Zitouni, A., Megherbi, A. C. and Sbaa, S. (2022). Classification of breast cancer histopathological images using densnet201, *2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)*, pp. 1–6.

Dong, K., Zhou, C., Ruan, Y. and Li, Y. (2020). Mobilenetv2 model for image classification, *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pp. 476–480.

Fikile Gasa, Z. N., Owolawi, P. A., Mapayi, T. and Odeyemi, K. (2020). Mobilenet neural network skin disease detector with raspberry pi integrated to telegram, *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, pp. 1–5.

Hartanto, C. A. and Wibowo, A. (2020). Development of mobile skin cancer detection using faster r-cnn and mobilenet v2 model, *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, pp. 58–63.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. (2017). Densely connected convolutional networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269.

Li, J., Wang, P., Li, Y., Zhou, Y., Liu, X. and Luan, K. (2018). Transfer learning of pre- trained inception-v3 model for colorectal cancer lymph node metastasis classification, *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 1650–1654.

Nabil, M., Rady, M., Moussa, K., Wessam, M., Hossam, M., Yousri, R. and Darweesh, M. S. (2021). A preprocessing approach to improve the performance of inception v3-based face shape classification, *2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pp. 205–209.

Pamadi, A. M., Ravishankar, A., Anu Nithya, P., Jahnavi, G. and Kathavate, S. (2022). Diabetic retinopathy detection using mobilenetv2 architecture, *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, pp. 1–5.

Rajayogi, J. R., Manjunath, G. and Shobha, G. (2019). Indian food image classification with transfer learning, *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, pp. 1–4.

Rokhana, R., Herulambang, W. and Indraswari, R. (2021). Multi-class image classification based on mobilenetv2 for detecting the proper use of face mask, *2021 International Electronics Symposium (IES)*, pp. 636–641.

Rukhsar and Upadhyay, S. K. (2022). Rice leaves disease detection and classification using transfer learning technique, *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 2151–2156.

Sahaai, M. B., Jothilakshmi, G. R., Selva Kumar, R. and Praveen Kumar, S. (2022). Comparative analysis on brain tumor classification using deep learning models, *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, pp. 1–5.

Shalu and Kamboj, A. (2018). A color-based approach for melanoma skin cancer detection, *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, pp. 508–513.

Sultana, J., Saha, B., Khan, S., Sanjida, T., Hasan, M. and Khan, M. M. (2022). Identification and classification of melanoma using deep learning algorithm, *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, pp. 1–6.

Waweru, A. K., Ahmed, K., Miao, Y. and Kawan, P. (2020). Deep learning in skin lesion analysis towards cancer detection, *2020 24th International Conference Information Visualisation (IV)*, pp. 740–745.

Younis, H., Bhatti, M. H. and Azeem, M. (2019). Classification of skin cancer dermoscopy images using transfer learning, *2019 15th International Conference on Emerging Technologies (ICET)*, pp. 1–4.

Zapirain, B. G., Zorrilla, A. M., Oleagordia, I. R., Nuñez, G. and Abtane, A. (2009). Skin cancer parameterisation algorithm based on epiluminiscence image processing, *2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 236–241.