A Comparative Study To Classify Malignant Skin Lesion Using Mask-RCNN

Research in Computing
MSc in Data Analytics

Nikhil Vaidya
Student ID: x20245980

School of Computing
National College of Ireland

Supervisor: Prof. Qurrat Ul Ain

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Nikhil Vaidya
Student ID:	x20245980
Programme:	MSc in Data Analytics
Year:	2021
Module:	Research in Computing
Supervisor: &	Prof. Qurrat Ul Ain
Submission Due Date:	18/09/2022
Project Title:	**A Comparative Study To Classify Malignant Skin Lesion Using Mask-RCNN**
Word Count:	6280
Page Count:	19

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:	
Date:	18/09/2022

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):	

A Comparative Study To Classify Malignant Skin Lesion Using Mask-RCNN

Nikhil Vaidya
x20245980

Abstract

Rapidly progressing melanomas of the skin can be lethal if they go undetected for an extended period. Recent studies have found that 1/3 of all cancers are skin cancers. If it remains undiagnosed, the disease can cause severe harm or even be fatal for the patient. Thus, it is essential to detect melanoma early and begin treatment immediately. Conventional medical diagnosis typically requires painful and expensive skin samples taken from the patient. Melanoma may be detected quickly and effectively using contemporary AI (Artificial Intelligence) and deep learning techniques. This study employs a novel method for melanoma identification by applying automated custom annotation on Mask R-CNN with ResNet101/50 backbones to classify lesions. The publically accessible ISIC Dataset (International Skin Imaging Collaboration) is used to perform this task. Mean accuracy and recall (mAP and mAR) ratings are used to grade the model. Using our model, we got an mAP score of 78.6 and an mAR of 97.6. The effectiveness of the models is evaluated via a thorough comparison.

Keywords- Skin Lesion, Melanoma, Data Augmentation, Custom Annotation, Mask R-CNN, Deep Neural Networks, ResNet101/50 backbone

1 Introduction

The epidermis and dermis are the two outermost layers of the skin. When it comes to skin cancer, fair-skinned people are more likely to develop melanoma since they are more exposed than others to the sun's UV-A rays. In most situations, melanoma is often not diagnosed because the signs are disregarded, owing to a lack of information. As a result, it becomes more lethal as the tumor progresses further into the dermis and the lymph nodes. This problem will make cancerous cells spread and mutate to numerous body places, making it impossible to treat with clinical treatments. Early detection and diagnosis are essential for prompt treatment and increased chances of survival. In addition, diagnostic procedures that rely on observations have poor accuracy. However, thanks to advances in modern computer-aided techniques, artificial intelligence, and deep learning, segmentation can be used in imaging technologies to detect disease early. Extracting odd patterns or appearances in skin color, texture, and other parameters that are not revealed to the naked or assisted eye can help us speed up the detection process.

Various invasive and non-invasive research has already been done to detect malignant skin lesions. The most traditional is Raman spectroscopy for skin cancer diagnosis in Optical Biopsy. In the experiment that the researchers carried out for the study Zhao et al. (2008), they used a technique that was based on spectroscopy. Samples were taken from a total of 289 patients using a spectrometer, and after that, machine learning methods such as least mean square were applied to detect the cancerous skin lesion. Recent studies suggest that it may not always be ethical or practicable to perform intrusive methods of gathering images of skin lesions from patients and then use spectrometers to analyze the data. In medical science, neural networks have emerged as an increasingly helpful method for diagnosing various conditions over the previous several years. For example, researchers in study Alfaro et al. (2019) took advantage of data readily available to the public on skin lesions, such as HAM1000 and ISIC, to diagnose cancer using artificial neural networks (ANN). They concluded that the computer-aided method such as CNN works more effectively and produces better results.

Implementing the CNN method gives mode details filter maps with various features, making the result more accurate than the other models. For example, the authors in Huang et al. (2020) utilized the HAM10000 dataset, which contains around 10000 skin lesion photos and includes both malignant and noncancerous skin lesions. The classification of skin cancer was accomplished by utilizing CNN coupled with VGG16/19. The findings are also compared with those obtained using other neural networks; however, CNN paired with VGG16/19 achieved significantly better results than others.

Due to its ease of use, speed, adaptability, and accuracy, Mask R-CNN has become the model of choice for even the most challenging tasks, including skin cancer detection. This research uses Mask R-CNN packages with ResNet101 and ResNet50 backbone to classify skin lesions using image augmentation and fine-tune the model by changing various hyperparameters. The main goal is to determine whether there is any significant impact of the automated annotations file generated by the custom build annotation function on the performance of the classification. The results are then compared to a new image classification approach called CapsNet. The capsule net is chosen for the comparison factor because it is proposed to classify the image data better. The comparison has never been performed as it is a relatively new classification model.

The research is subdivided into below given sections:

- Section 2 - Related Work.
- Section 3 - Research Methodology.
- Section 4 - Design Specification
- Section 5 - Implementation
- Section 6 - Evaluation & Discussion
- Section 6 - Conclusion

1.1 Research Question and Objective

How do Mask R-CNN with different backbones perform when classifying melanoma using an automated function to create an annotation file?

The primary objective of this study is to use an automated annotation custom build function to evaluate the effectiveness of the MASK R-CNN model employing ResNet101 and ResNet50 as its backbone. The segmentation mask and images are fed into the counter calculation in the annotation function, producing shape attributes. The function's final output is an annotation file for use with Mask R-lesion CNN's classification. Both models are evaluated, and the findings are compared to those obtained using the cutting-edge CapsNet methods described in (Goceri; 2021). For this bench-marking exercise, we have specifically selected CapsNet because it's been claimed to outperform more conventional CNN models.

2 Related Work

A complete study of past work is evaluated and summarised in a specific domain as part of the literature review process. A literature review helps the researcher comprehend the topic under discussion and gain sufficient information on what still needs to be discovered in the field.

This research paper conducts an in-depth analysis of past studies that have been conducted on the classification of skin lesions utilizing both traditional and contemporary techniques as well as modern methods of artificial intelligence and deep learning.

The review of the relevant literature is further subdivided into the below-given sections:

- Classifying Skin Lesion using Traditional Methods
- Classifying Skin Lesion using Machine Learning
- Classifying Skin Lesion using Convolutional Layers (CNN)

2.1 Classifying Skin Lesion using Traditional Methods

In order to detect melanoma, spectroscopy, a non-invasive method of examining the skin's response to electromagnetic waves or microwaves, can be employed. Malignant cells can be identified using various electromagnetic radiation techniques, although researchers have had variable degrees of success with each method. In 2008, Raman spectroscopy was one of the methods used to detect cancerous tissue on the skin. For the study (Zhao et al.; 2008), a Raman spectrometer was employed on 289 individuals to collect and analyze the spectra. Machine learning techniques such as least square regression were used to differentiate the malignant tissue samples from the non-cancerous skin lesions after the spectra were collected. This experiment's results are statistically significant at 91% sensitivity and 75% specificity, where the melanoma was discovered in 37 out of 289 patients. In contrast, others were non-cancerous or with different cancer types.

A similar study was performed in (Li et al.; 2013), and the results were superior to those of (Zhao et al.; 2008). These results were obtained by capturing images of skin lesions and then using the SVN classification model to sort them into the correct subtypes. All 187 participants, aged 22 to 79, had their skin tissue scanned vertically and parallelly. For this classification, six images with 32X512 resolution were taken for each male and female patient. Of these, 19 samples were discovered to be malignant. The SVM classifier model performed brilliantly, obtaining an accuracy of 92 percent coupled with a sensitivity of 100 percent and a specificity of 92 percent after incorporating statistical methods such as mean and standard deviation for feature extraction.

In medical practices, Raman spectroscopy devices are not often employed since they take a while to catch the spectra from the devices. In the study (Zeng et al.; 2016), it has been hypothesized that the spectrometer can detect malignant tissue in real-time using a new technology that uses specialized optical probes. Nine hundred people volunteered for the study, and samples were taken by measuring and photographing skin lesions. Compared to the old method, which required several minutes, the new method could capture the spectra in seconds. Then, various statistical techniques were applied to analyze the spectra acquired to distinguish between cancerous and non-cancerous skin lesions. (Zeng et al.; 2016) Sensitivity and specificity of 90% and 73%, respectively, are obtained using the partial least squares approach, which is followed by multiple regression and genetic algorithms.

Later sections of this paper will further explore the neural networks in detecting cancerous skin lesions.

2.2 Classifying Skin Lesion using Machine Learning

This section explores how current technology has evolved to identify skin cancer using computer-assisted machine learning and neural networks. International Skin Imaging Collaboration (ISIC) has hosted a competition since 2016 and shared an enormous public dataset of skin lesion scans, which has helped several academics develop unique techniques for identifying melanoma early on.

The TNM (Tumor(T), Nymph(N), and Metastasis(M)) method is used to determine the stage of melanoma on the epidermis. The cancer stage can be predicted using the TNM technique, which considers the tumors, lymph, and distant metastases as a variable. The skin lesion data from 27,882 skin cancer patients were considered and divided into two datasets: the training dataset and the test dataset for machine learning purposes. (Todorova; 2020) The accuracy and R-Squared value of various approaches, including Decision Tree, Support Vector, and Ensemble Boosted Tree algorithms, are examined when determining cutaneous melanoma. With 99.06% accuracy and the highest R-squared value of 0.93, the decision tree with a medium tree value performed much better than the other models in classifying cutaneous melanoma. Thirty-six features were gathered in (Javaid et al.; 2021), and the most significant ones were segmented and classified using the threshold technique. The images were enhanced by pre-processing, and various techniques were used, including HOG (histogram of gradient), principal component analysis, and pixel reduction, to reduce the image size to 96X128 pixels before putting it into a model. As the data was highly imbalanced, it was necessary to apply oversampling techniques like SMOTE to balance the dataset. (Javaid et al.; 2021) The random forest model surpassed all other models with a 93 percent success rate.

In order to determine which algorithm is most efficient in detecting melanoma cells, dermoscopy data from the ISIC dataset to evaluate the findings of multiple algorithms is used in (Janney et al.; 2018). After performing the data collection and the segmentation, features are extracted using Naive Bayes and Artificial Neural networks, which are then applied. Images were then processed using GLCM techniques to extract texture and wavelength features, which were then exported in .xls format. The data acquired from this method was fed into the Matlab-based classification models. (Janney et al.; 2018) A comparison is made between all the models implemented using accuracy, sensitivity, and specificity evaluation techniques. Finally, the ANN fared the best, with findings that were 89% correct among all models. Another study where an unusual method for automatically detecting melanomas (ALDS) has been discovered in (Farooq et al.; 2016). For this purpose, the probability-based technique was used to segment the data, and then classification algorithms were implemented, such as the support vector and neural network classifier. In the beginning, the data was with a resolution of 765X573. However, after the images have been cleaned and segmented, they are polished, and the features are expanded to be used in the further stages of the model implementation. Segmentation was also aided by a gray-level technique and watershed algorithms, with promising results. After using the models, the final results showed acceptable accuracy of around 80%. (Farooq et al.; 2016)

Medical researchers have relied heavily on neural networks in the previous several years to help them discover new diseases. For example, the research uses multiple publicly available datasets, such as HAM10000, ISIC, and PH2, to implement machine learning models. In addition, there have been encouraging results from neural networks like the artificial neural network (ANN) and the (RNN) recurrent neural network. After examining most papers, convolutional neural networks were found to be the most widely used approach for medical prediction. In Section 2.3, we will delve deep into the convolutional network research to detect the skin lesion that's already been done and achieved significant results.

2.3 Classifying Skin Lesion using Convolutional Layers (CNN)

In this section, we will explored various CNN models implemented for classification of Melanoma.

2.3.1 Classification using Traditional CNN's

With the publicly available dataset, Human Against Machine (10000), a new proposed model for identifying malignant skin lesions is used to identify cancerous lesions and incorporate VGG16/19 in the same way as earlier studies. (Aburaed et al.; 2020) A total of 10,000 photos of people of various age groups and genders are included in the dataset, each with a different cancer set. The dataset was reduced to around 7thousand after being segmented and preprocessed to reduce the issues caused by unbalanced data. The revolutionary strategy recommends a deep CNN(Convolutional Nueral Network) with two convolution layers and maximum pooling of three before dividing the data in an 80/20 ratio for training and testing (Aburaed et al.; 2020). Compared to other models, the suggested system outperformed them by 99%and resulted in lower losses than the competition, and the results were superior.
In (Jusman et al.; 2021), a malignant skin lesion is more classified using the HAM10000 data. The comparison is based on MLP and VGG16 using evaluation metrics such as accuracy and performance evaluated using a novel CNN approach. (Jusman et al.; 2021) Two dense and four convolutional layers make up the specialized CNN models. As a result of using VGG16 and Imagenet, the dataset had to be scaled down to 128X128

because the VGG16 only accepts images with dimensions of 224*224. New CNN models surpassed MLP (Multilayer Perception) in terms of performance, but they took much more processing time to train because of their complexity. While this accuracy was considered acceptable, the models could still accurately detect melanomas at 87 percent.

In (Jain et al.; 2021), the ideal technique combines ensemble learning with a primary CNN model. With the help of segmentation methods, numerous CNN approaches can help the detection process generate the most accurate model possible. The dataset comprised 44,108 jpg images, which were analyzed and then divided into two groups with an 80 to 20 ratio for testing and training dataset. Images are segmented using a k-means cluster for the model used in the simulation to be employed in the process. The Keras image generator is also applied in this work to enhance the image quality. Xceptionnet version 3 (with 15 epochs), densenet201 (30 epochs), and VGG16 (30 epochs) were all examined in (Jain et al.; 2021). A total of three algorithms were put through their paces in the preceding section: Though Densenet could not equal the performance of the VGG16 or Xceptionnet models, it did attain a 95% accuracy rate, matching the other two models.

In (Chen et al.; 2020) again, a CNN model in conjunction with EfficientNet was constructed, which used a combination of CNN and ensemble learning in solving this problem statement. Using EfficientNet's scalable neural networks, photos can be enhanced and used in the classification stage to generate more efficient results. EfficientNet beats earlier research approaches when utilized in conjunction with CNN for segmentation and classification. Six basic EfficientNet models were created to solve the issue arising from highly skewed data and evaluate alternative parameter designs. Their predictions were combined to arrive at a solution. On the Kaggle platform, a dataset was available for researchers to use. The proposed model outperformed the VGG16/19 by a statistically significant margin with an AUC score of 0.95 (Chen et al.; 2020).

In the following subsection, the studies done on a more improved version of CNN models are studied called Mask R-CNN. Mask R-CNN is an extended and improved version of the Fast R-CNN model with an extra branch of classification and mask segmentation.

2.3.2 Classification using Mask R-CNN

In study (Huang et al.; 2020), an updated Mask RCNN model is utilized to segment images of skin lesions. In order to carry out the segmentation process, this study uses the publicly available dataset from ISIC (International Skin Imaging Collaboration). The dataset has a total of 500 photos, including 238 melanoma images and 262 nevi images. These images have been separated into train and test datasets. The dimensions of the photos range from 962*1022 pixels across and 767 *762 pixels. The model used in this investigation was put into action in two stages: In the first step, the object proposal was provided by RPN (Region Proposal Network), and in the second stage, the Fast R-CNN model was used to forecast the segment mask. In addition, the researchers modified and altered the hyperparameter in the suggested model. For example, they reduce the weight decay parameter to a specific value. For example, the NMS (Non-Maximum Suppression)algorithm was employed in stage 2 to determine which candidate box had the most favorable candidates. The model worked effectively, as evidenced by its ability to achieve an accuracy and recall level of 90% and 91%, respectively.

A similar study is performed in this research for segmenting the skin lesion by researchers in (Alfaro et al.; 2019). On the 2017 dataset from the International Skin Image Collaboratory (ISIC), two distinct models, including U-Net and an enhanced Mask R-CNN model, were used. Both systems are used with their respective hyperparameters, and the outcomes are contrasted using evaluation matrices based on the Jaccard index and dice. Approximately 2000 train pictures are included in the dataset, along with 150 validation and 600 test images. This research's findings indicate that the Mask R-CNN model's performance was considerably superior to that of the U-NET model. The models are trained in three stages, with the first stage consisting of training for 160 epochs, the second stage consisting of training for 180 epochs, and the third stage consisting of training for 200 epochs. The Jaccard index and dice for the Mask R-CNN model are 0.78 and 0.79, whereas the corresponding values for the U-NET model are 0.58 and 0.68.

2.4 Summary of Literature

From hardware-based techniques to contemporary neural networks are examined in the literature review. Several fascinating findings and methodologies are revealed, and it is shown that the influence of integrating different neural network models on melanoma diagnosis can be recognized by studying many articles. It is shown in the 2 section how a range of spectroscopic and statistical methods are used to identify skin lesion types and cancer. In addition, in section 2.2, many studies such as (Todorova; 2020), (Javaid et al.; 2021), (Hasan et al.; 2019), and (Senan et al.; 2021) contributed significantly to our understanding of computer-assisted machine learning in this domain. Then in section 2.3, a brief history of current detection methods such as CNN models and how numerous machine learning approaches and neural networks might improve the classification result. A close look at convolutional networks in Section 2.3.1 reveals how novel techniques for convolution neural networks may help us improve the detection of malignant skin lesions. The section then explores the Mask R-CNN approaches for skin lesion classification in studies (Huang et al.; 2020) and (Alfaro et al.; 2019).

Finally, it has been shown that deep learning algorithms can aid in diagnosing various illnesses, but they also have their shortcomings. For example, in most instances, researchers examining skin lesions will employ the Mask R-CNN method for the segmentation task. A simple annotation file with all the image labels and shape attributes is necessary for Mask R-CNN, an improved version of the Fast R-CNN model with an additional branch for classification. Annotations for use in the Mask R-CNN model for object classification are often created manually with a dedicated software package like VIA (VGG Image Annotator). If the dataset is huge, it might be challenging to perform such an operation. Instead, this study employs a transfer learning approach to train the model, and an automated function is developed to produce unique annotation files. In the discussion section, we compare the results and draw some inferences about the model's overall performance based on our findings.

3 Research Methodology

This research follows the KDD data mining approach as it is the best approach to follow in medical studies. It provides a structured approach to conducting any research with steps involving data selection, preprocessing, transformation, and extracting meaningful information from the data. The study is performed on the ISIC skin lesion dataset to create an intuitive approach to classifying skin lesions to help the medical professional in the early detection of melanoma. As mentioned in the literature review, deep-learning models are best for detecting melanoma early. Transfer learning-based Mask R-CNN with two different backbones, ResNet101 and ResNet50, is implemented in this study. **Figure 2.** illustrates the research methodology used. The subsections given below will further explain each steps that are followed in the research.

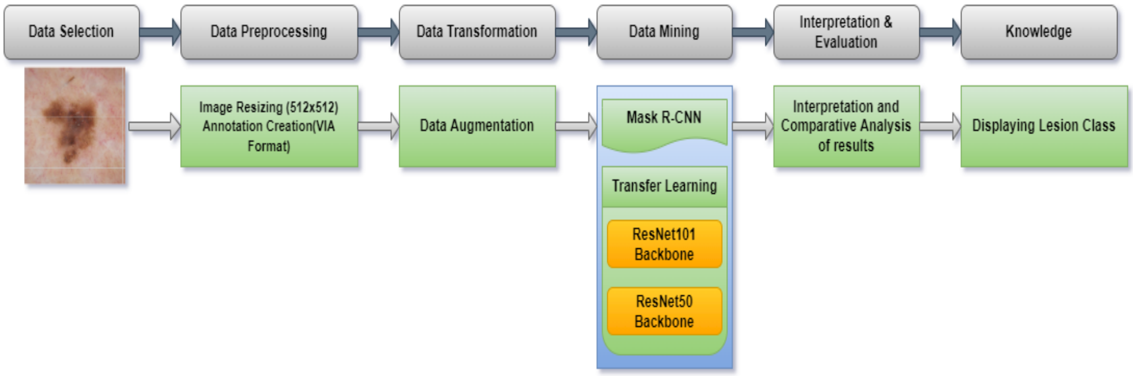

Figure 1: Process Flow (KDD Approach).(Fiala; 2005)

3.1 Data Selection

The research is conducted using a publicly available dataset from the ”ISIC (International Skin Imaging Collaboration) -Melanoma Segmentation and Detection Challenge.” The department has held a lesion classification competition every year from 2016 till 2020. The dataset provided has multiple skin lesion images in .jpg format and their segmentation in *.png format. Ground truth files are also provided in .csv format for train and test datasets, containing the image names and their diagnosis. All the images and segmentation have varying degrees of resolution, and a total of 1279 images, including 900 train and 379 test images, are present. The images are taken from patients for medical diagnosis where no sensitive data about the patient is disclosed while making it public. Because of this, several medical universities and associations use public data to stimulate more computer-aided research.

3.2 Data Pre-processing

As mentioned in the previous section, research uses data available to the public and provided by the Skin Imaging Collaboration. The dataset was downloaded from the ISIC website and contained images and their corresponding masks. The dataset has varying resolutions; for example, some images are 1024x767, and some are 1024x768 and soon. To normalize the data and improve the model's performance and efficiency, the images, and their corresponding masks are resized to 515x512 before modeling. The images are saved in imagesV2 and the maskV2 folder in both train and Val datasets and later copied to google-colab for faster processing. The images and mask are then used to generate VGG Image Annotator(VIA) file with the name via_region_data.json. The JSON file has x_point and y_point coordinates of the images acquired from the segmentation. The labels are also present in the malignant or benign JSON file, which are acquired from the growth truth files named Training_GroundTruth.csv and Test_GroundTruth.csv. A total of 1279 images are used for research, and all the images are in .jpg format.

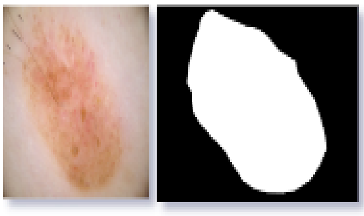

Figure 2: Lesion and Segmentation Mask.

3.3 Data Transformation

3.3.1 Data Augmenting

The photos are already in the.jpg format, and thus the conversion is unnecessary. However, augmentation is performed to enhance the model's performance and efficiency. For example, in the augmentation step, the lesion pictures are rotated, brightened, and modified with contrast, sharpness, and Gaussian blurring. **Figure 3.** shows the augmented image of skin lesion.

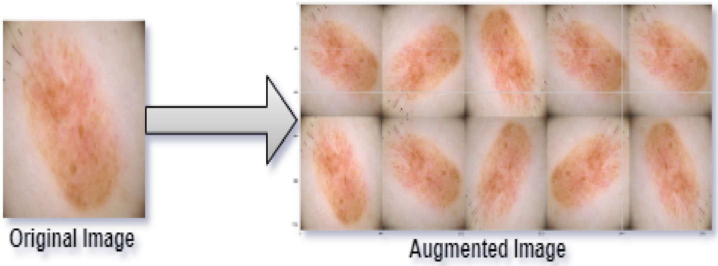

Figure 3: Augmentation on Lesion Image.

3.3.2 Data Splitting

The dataset is split into two sections: Train and Valid dataset, including 900 pictures and an annotation file (via region data.json) in the training dataset. In contrast, the validation dataset comprises 379 JPEG photos and an annotation file.

3.4 Modelling Approach

As the literature review studies, deep learning techniques such as Mask R-CNN with various backbones can effectively classify lesion lesions. (Huang et al.; 2020). Mask R-CNN was introduced as an extension of Fast RCNN and is used significantly in image classification problems. As a novel approach in this research, Mask R-CNN with backbone Resnet101 and Resnet50 is implemented where a custom-made annotation file is used for classification. The results of both models are later compared with each other and with the CapsNet implemented in Goceri (2021) to determine whether our custom model can perform better. CapsNet is chosen for comparison as it is a relatively new image classification technique and is suggested to give better insights into the problem.

Figure 4. shows the stages in which we performed the study.

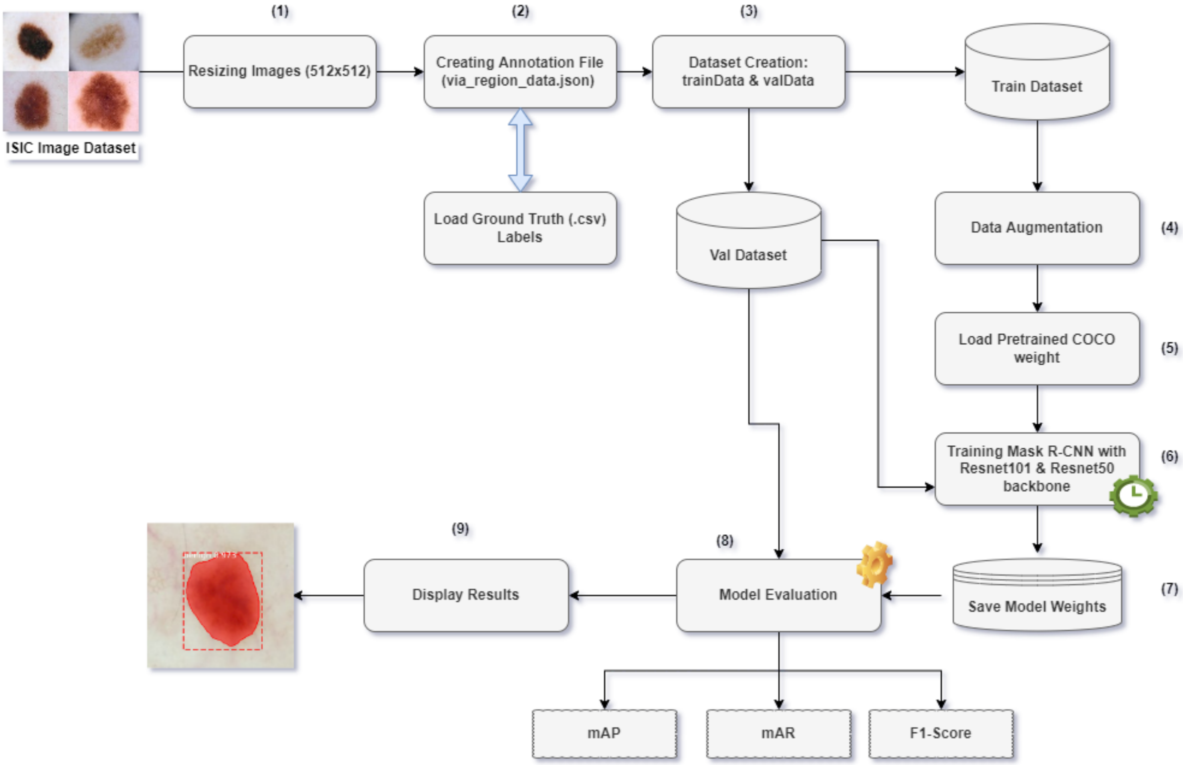
ISIC Image Dataset
(1)
Resizing Images (512x512)
(2)
Creating Annotation File (via_region_data.json)
Load Ground Truth (.csv) Labels
(3)
Dataset Creation: trainData & valData
Train Dataset
Val Dataset
Data Augmentation
(4)
Load Pretrained COCO weight
(5)
Training Mask R-CNN with Resnet101 & Resnet50 backbone
(6)
Save Model Weights
(7)
Model Evaluation
(8)
Display Results
(9)
mAP
mAR
F1-Score

Figure 4: Proposed architecture

- Firstly, to maintain the consistency of the dataset, images and the segmented masks(.png) are selected and resized to 512x512.
- In the second step custom annotation file is created where the Training_GroundTruth.csv and Test_GroundTruth.csv files are used to assign labels and generate via_region_data.json.
- The third step is creating training and validation datasets to be used in the model using prepare() method.
- Augmentation is performed on the training dataset before using it in the model training.
- Pre-trained COCO weights are downloaded and loaded into the model as part of transfer learning.

- Model training is implemented where training and validation datasets are used.
- The generated weights are stored at each epoch, and a weight file is created at the end.
- An epoch with minimum validation loss is selected and used for evaluation, prediction, and visualization.

4 Design Specification

The Mask R-CNN model is a variant of the R-CNN family. In fact, it is a more robust variant of the Fast RCNN model, with the ability to make accurate predictions in segmentation and classification tasks. The Mask RCNN consists of two main phases: In the first phase, features are extracted using the ResNet101/50 model, and a feature map is generated. The RPN (Region Proposal Network) receives the feature map as input, allowing it to scan for and propose areas where the object, in this example, a skin lesion, is present. In the second phase, the FCN (Fully Connected Convolutional Network) is used to create the multi-class objects, bounding boxes, and masks based on the proposed regions provided in the first phase.

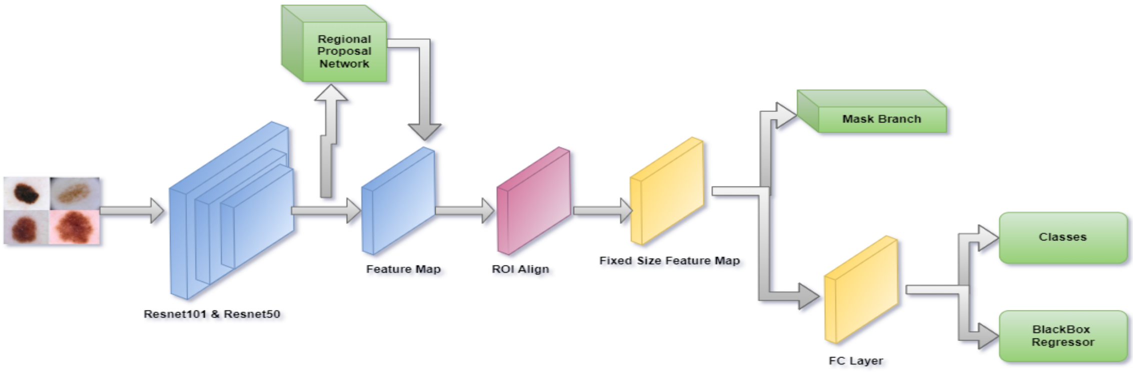

Figure 5: ResNet50/101 Based Lesion Classification Architecture.

- **Backbone and Residual Network**: ResNet101 and ResNet50 backbone are used in this research to classify skin lesions. Resnet101 has 101 layers, and Resnet50 has 50-layer deep neural networks, which will be compared in this study. The RPN(Region Proposal Network) is a neural network that scans the feature map provided by the backbone model using anchors. A total of 9 anchors will be created along with predefined boxes and scales. This task determines the object area and the bounding boxes required.

- **RoIAlign**: RoI-Align is a modified version of RoI-Pool, which is present in the Fast RCNN model. RoI-Align generates multiple bounding boxes overlapped on one another. These features are then fed to the FC layer, which uses a softmax classifier to classify skin lesions.

4.1 Transfer Learning

This study employs a transfer learning strategy wherein we utilize weights of the pre-trained COCO dataset to hone our model's performance. After downloading the COCO weights, they are kept in a particular directory at Google's Colab. In addition, the weight file for the best epoch that was created after training the model is used for the classification and prediction phase. **Figure 6.** shows the flow of the transfer learning method used in the research.

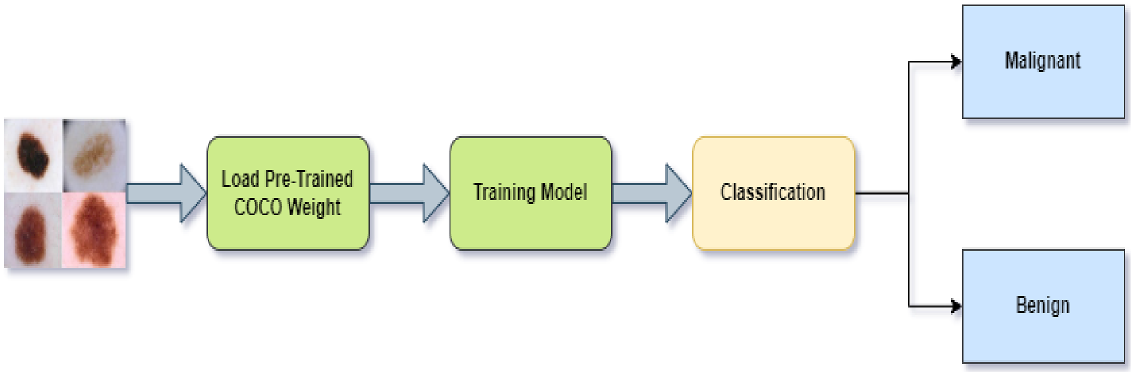

Figure 6: Transfer Learning Mask-RCNN Approach.

5 Implementation

This chapter provides an overview of the project's implementation. From basic setup to data processing and model training, it covers all the processes and tools required to complete the task.

5.1 Environment Setup

In this research, the Google Colab Pro version is used for training the model on Python 3.7.13 because new deep learning techniques require high processing power, which Google Colab provides. This version of Colab comes with 124GB of disk, 25GB RAM, and Intel(R) Xeon(R) CPU @ 2.20GHz processor. Tensorflow version 1.13.1 and Keras version 2.1.5 is used to implement the task.

5.2 Data Setup and Processing

The dataset was downloaded from the archive section of the 'ISIC Skin Lesion Challenge' [1] and contains two folders: train and test. We treated the test data from this website as the validation data for our research as the ground truth values were also given. These two folders are subdivided into images and masks folders, which are downloaded and mounted to google drive. The ground truth files were also downloaded for train and val datasets in .csv formats, such as Training_GroundTruth.csv, which contains the label information.

[1] https://challenge.isic-archive.com/data/

The images and masks are then resized to 512x512 resolutions. In addition, resized images were stored into imagesV2 and masks in the masksV2 folder for both the train and validation dataset. Our custom annotation function then used resized images to generate an annotation file named via_region_data.json to train the model. In addition, the Mask RCNN library [2] is mounted on the colab local disk, and datasettrain datasetval folders are created. All the resized images and the annotation file for both the train and validation dataset are copied to colab local directory folder as explained to make the training process faster. Augmentations such as rotation, flipping, gaussian blur, and sharpening are also performed while training the model.

5.3 Model Training with Transfer Learning

The Mask R-CNM model is performed using the transfer learning technique where we downloaded and used pre-trained COCO dataset weights [3] for training the model. Model is trained with a variation of Hyper Parameter such as SGD(Stochastic Gradient Descent) optimizer is used in our research as it is the best suited for Mask RCNN models and is also suggested by the algorithm's developers. This model has two classes: Malignant and benign, to classify skin lesions. The batch size is set to 1, and steps per epoch are set as 300 with 150 validation steps. Resnet101 with backbone stride [4, 8, 16, 32, 64] and Resnet50 model is implemented in the Mask RCNN model. A changed rpn anchor values of [8,16,32,64,128] are used to improve lesion detection and classification. Image minimum and maximum dimensions are set to 512 as the original images are also the same size. Below given table describes the flow of the training model 1:

Epoch Param	**Learning Rate Value**
1-2	0.001
3-14	0.0001
15-22	0.0001
23-30	0.0005

Table 1: Learning rate decay

The model is trained for a total of 30 epochs in four steps: First, we trained the model for two epochs using head layers and augmentation. The model is then trained for 12 more epochs with a learning rate of 0.0001 in the second stage by setting the epochs parameter to 14. The model was then trained for another 8 epochs in the third stage by adjusting the epoch param to 22 and using the same learning rate of 0.0001. In the last stage, we trained the model for 8 epochs with an epoch param of 30 and a learning rate of 0.0005. The learning rate is determined by trial and error, and the best epoch value obtained through experimentation is chosen for evaluation, with no overfitting of the model. **Table 2** below shows set of configuration used in the for ResNet101/50 model.

[2] https://github.com/matterport/Mask$_{R}CNN$

[3] https://github.com/matterport/Mask_RCNN/releases/download/v2.0/mask_rcnn_coco.h5

Hyperparameter	**Value**
NAME	Lesion
BACKBONE	resnet101/50
BATCH_SIZE	1
NUM_CLASSES	3
POOL_SIZE	7
BACKBONE_STRIDES	[4, 8, 16, 32, 64]
RPN_ANCHOR_SCALES	(8, 16, 32, 64, 128)
IMAGES_PER_GPU	1
IMAGE_MAX_DIM	512
IMAGE_MIN_DIM	512
STEPS_PER_EPOCH	300
VALIDATION_STEPS	150
LEARNING_RATE	0.001, 0.0001 & 0,0005
DETECTION_MAX_CONFIDENCE	100
DETECTION_MIN_CONFIDENCE	0.7
IMAGE_SHAPE	[512 512 3]
WEIGHT_DECAY	0.0001
MASK_POOL_SIZE	14
MASK_SHAPE	[28, 28]

Table 2: Training Configuration Used in ResNet101/50 Model.

6 Evaluation

6.1 Model Evaluation

Evaluation is the most crucial part of analyzing the model's overall performance and efficiency. In this study, the performance of both the ResNet101 and ResNet50 models will be evaluated three metrics: mAP(Mean Average Precision), mAR(Mean Average Recall), and F-1 score. Mask R-CNN libraries provide a function to calculate the average precision (APs) for the model for each epoch. A mean of all the collected APs is taken by building a custom function in this study and calculating the mAP(mean average precision) score. Both the models are trained for 30epochs, and loss graphs generated after that are examined for over-fitting. Evaluation is performed in two parts, and details are explained below in subsections.

6.1.1 Study 1: ResNet101 Model

The ResNet101 backbone architecture using transfer learning was performed in the first study. The model is executed for a total number of 30 epochs. The weight and losses obtained are saved at the end of each epoch. Figure 7 shows the loss train loss vs. validation loss graph. We can see that the loss values are decreasing gradually, with the lowest validation loss obtained at the 25th epoch, which is 0.40. These epoch weights are stored in the log directory and is chosen as the best epoch for this model, which is further used for running the model in inference mode and prediction.

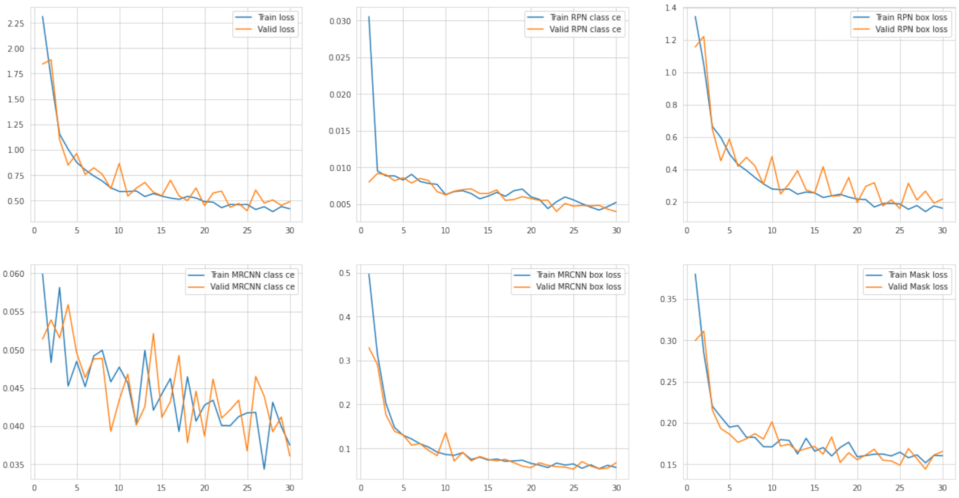

Figure 7: Loss Graphs of ResNet101 Model

After the weights are loaded, and the inference model is implemented, the performance metrics are calculated. The model received an mAP score of 78.6%, mAR score of 97.6%, and an F1 score of 87.05% for the training data. For test data, the metrics values received are mAP score of 76.3%, mAR score of 97%, and F1 score of 85%. It is noted that the mAP score for both sets is higher than the threshold of mAP, which is 0.5 (50%). It is suggested that if a model has an mAP value higher than the threshold, it is a good fit model, and our method is able to achieve that.

6.1.2 Study 2: ResNet50 Model

The second experiment used transfer learning on the ResNet50 backbone architecture. Again, we run the model for 30 iterations, saving the weights and losses after each iteration. Figure 7 depicts the validation loss vs train loss. By observing the loss values over time, we can see that they are decreasing, with the lowest validation loss found at the 21st epoch being 1.21.

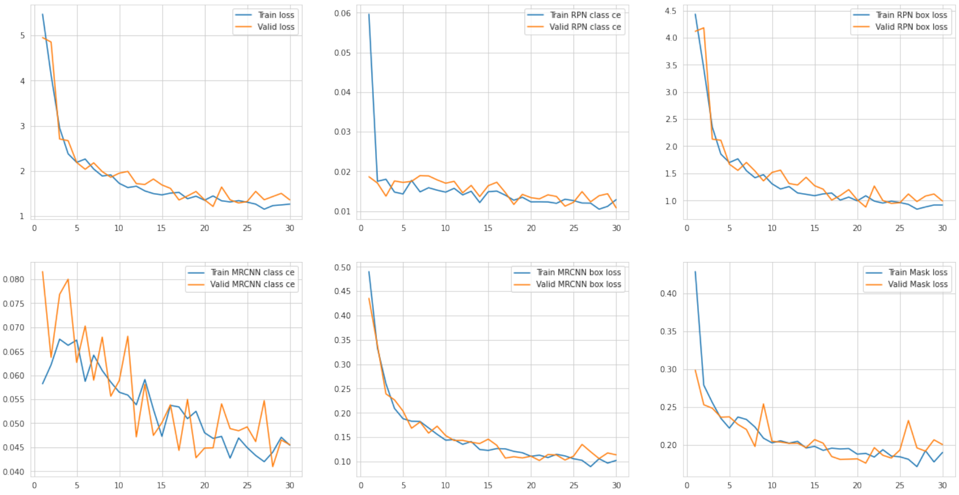

Figure 8: Loss Graphs of ResNet50 Model

Once the weights have been loaded, and the inference model has been applied, the performance metrics for resnet50 is calculated. For the training data, the model provided a poor mAP score of 56.6%, mAR score of 85.8%, and F1 score of 68.1%. Measured values for validation data include an mAP score of 59.4%, an mAR score of 88%, and an F1 score of 70%. The scores obtained from this model are not optimal as compared to the ResNet101 model.

6.1.3 Results

In order to make accurate predictions, we use the ResNet101 model, which performed better across all of our evaluation measures. Some random images from the validation dataset were used to test the model. One of the images in which the model successfully distinguishes between a malignant and benign skin lesion is displayed in Figure 9. Furthermore, the model can generate masks and bounding boxes consistent with the classification.

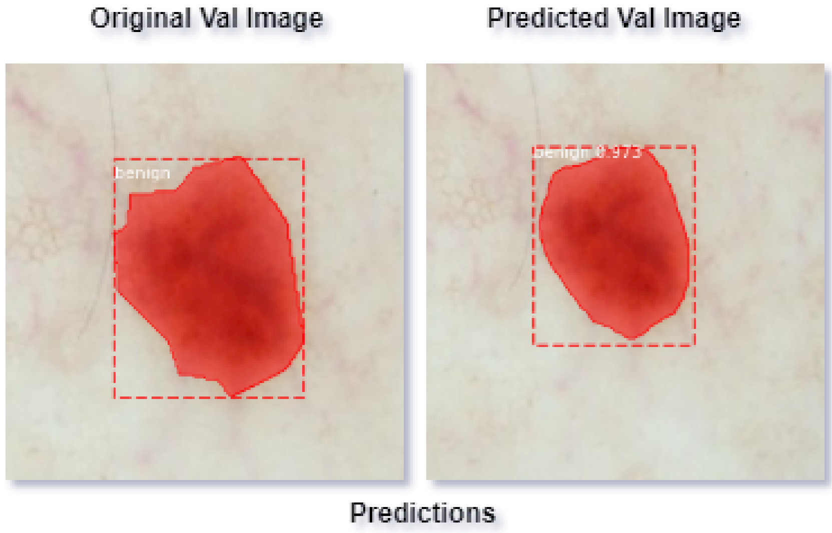

Figure 9: Prediction done on the validation dataset

6.2 Discussion & Comparison

The study's goals are to (1) assess how well the Mask R CNN model works in skin lesion classification and (2) determine whether or not its efficiency and accuracy can be enhanced by using automated built annotation files. In this research, the Mask-RCNN model with backbones ResNet101 and Resnet50 is implemented with a custom build function to create an annotation file using images and their corresponding masks. The results are then compared with the new CapsNet technique (Goceri; 2021). CapsNet model is chosen for comparison as it is a relatively newer classification model and is suggested to be performing better than most CNN models.

The models are compared by calculating three metrics: mAP(Mean Average Precision), mAR(Mean Average Recall), and F1-Score. In this research, the model is executed for a total of 9000 iterations (300 steps per epoch x 30 epochs) using different settings of hyperparameters such as RPN scales, image dimensions, and decreasing learning rate. As the dataset is slightly imbalanced, augmentation techniques such as rotation, gaussian blur, sharpening, and brightness are also implemented on the dataset to improve the results. An mAP score of 78.6% and mAR score of 97.6% was achieved in the ResNet101 model, whereas an mAP score of 56.6%, and the ResNet50 model achieved an mAR score of 85.8%. This suggests that the ResNet101 backbone models perform better by a significant margin out of both implemented models.

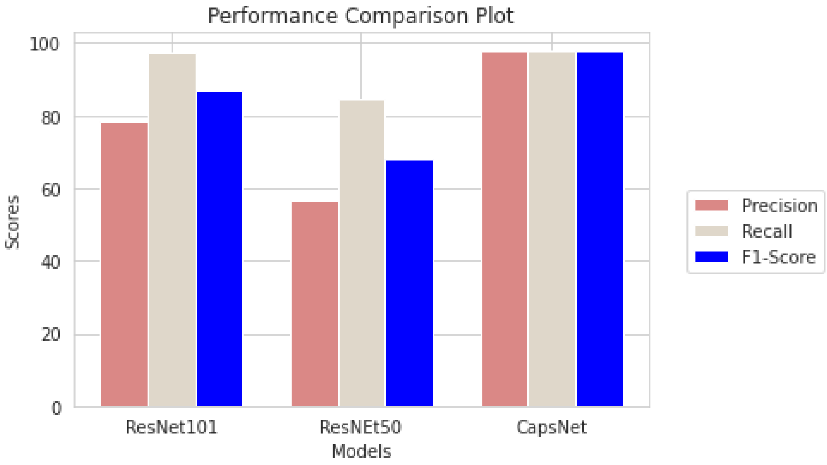

Figure 10: Comparison Plot Implemented Models vs CapsNet

Figure 10 shows the comparison plot for the models. Here we can see that the precision of CapsNet was significantly higher as compared to the ResNet101 and ResNet50 models, with a value of 98% in all three metrics, as discussed. Although we were able to improve the performance of the ResNet101 model by implementing image augmentation, it still achieved a lower score than the CapsNet model in (Goceri; 2021). The reason behind these lower scores is: (1) shape attributes acquired by our custom-built function were not able to extract exact coordinates (2) the amount of dataset used in this study was less as compared to the (Goceri; 2021); and (3) resource limitation as after running the model for a significant time the resources available on Colab were getting drained used and resulting in frequent server disconnections. The implemented model was able to classify the skin lesion perfectly but could not achieve the same or better score than CapsNet. Changes in the design of the custom annotation function to extract the shape feature accurately and acquiring more data to train can further improve the performance.

7 Conclusion

Detection and early diagnosis of Melanoma is an essential task to avoid low survival rates. The main objective of this research was to assess the performance impact of custom build annotations for classifying Melanoma using Mask R-CNN. The model was implemented with two different backbones, Resnet101 and ResNet50, by changing various hyperparameters available in the Mask R-CNN configuration. A set of data augmentation techniques were also performed in training the models. The implemented model was then evaluated, and the performance metrics were then compared with a new classification CapsNet technique. The CapsNet is used as a comparison model because of

its proposed accuracy in the classification of image data. The implemented model with the Resnet101 backbone achieved an mAP (Mean Average Precision) score of 78.6%, which is significant compared to the Resnet50 backbone, which scored low 56.6%. Due to the three reasons debated in the discussion section of this report, including resources limitation, system design(custom annotation function), and dataset, the model could not perform well compared to CapsNet (Goceri; 2021).
The tasks involved in future work include enhancing the automatic customized function to accurately recognize shape features, running the model with more balanced data, and using better resources to enhance the model's performance. The SGD optimizer used in this study is just one option; alternative optimizers and backbones, such as ResNet152 can also be utilized to improve the model's performance.

8 Acknowledgment

I would like to express my gratitude to Prof. Qurrat Ul Ain for her support and feedback during the duration of this research. I am grateful to her for continually encouraging and motivating me to improve my research.

References

Aburaed, N., Panthakkan, A., Al-Saad, M., Amin, S. A. and Mansoor, W. (2020). Deep convolutional neural network (dcnn) for skin cancer classification, *2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pp. 1–4.

Alfaro, E., Fonseca, X. B., Albornoz, E. M., Martínez, C. E. and Ramrez, S. C. (2019). A brief analysis of u-net and mask r-cnn for skin lesion segmentation, *2019 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, pp. 000123–000126.

Chen, D., Ziyuan, Z., Ji, H. and Huang, Y. (2020). Melanoma classification using deep convolutional neural networks with ensemble scheme, *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pp. 363–366.

Farooq, M. A., Azhar, M. A. M. and Raza, R. H. (2016). Automatic lesion detection system (alds) for skin cancer classification using svm and neural classifiers, *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 301–308.

Fiala, D. (2005). Web mining and its applications to researchers support, *Technical Report No DCSE/TR-2005-06* .

Goceri, E. (2021). Capsule neural networks in classification of skin lesions, *International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing*, pp. 29–36.

Hasan, M. Z., Shoumik, S. and Zahan, N. (2019). Integrated use of rough sets and artificial neural network for skin cancer disease classification, *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, pp. 1–4.

Huang, C., Yu, A., Wang, Y. and He, H. (2020). Skin lesion segmentation based on mask r-cnn, *2020 International Conference on Virtual Reality and Visualization (ICVRV)*, pp. 63–67.

Jain, M., Jain, M., Faizan, M. and Nehra, N. (2021). Melanoma classification using deep learning architectures and transfer learning, *2021 International Conference on Industrial Electronics Research and Applications (ICIERA)*, pp. 1–5.

Janney, B. J., Roslin, S. E. and Shelcy, M. J. (2018). A comparative analysis of skin cancer detection based on svm, ann and naive bayes classifier, *2018 International Conference on Recent Innovations in Electrical, Electronics Communication Engineering (ICRIEECE)*, pp. 1694–1699.

Javaid, A., Sadiq, M. and Akram, F. (2021). Skin cancer classification using image processing and machine learning, *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, pp. 439–444.

Jusman, Y., Firdiantika, I. M., Dharmawan, D. A. and Purwanto, K. (2021). Performance of multi layer perceptron and deep neural networks in skin cancer classification, *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, pp. 534–538.

Li, L., Zhang, Q., Ding, Y., Jiang, H., Thiers, B. T. and Wang, J. Z. (2013). A computer-aided spectroscopic system for early diagnosis of melanoma, *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, pp. 145–150.

Senan, E. M., Jadhav, M. E. and Kadam, A. (2021). Classification of ph2 images for early detection of skin diseases, *2021 6th International Conference for Convergence in Technology (I2CT)*, pp. 1–7.

Todorova, M. (2020). Application of machine learning methods for determining the stage of cancer, *2020 International Conference Automatics and Informatics (ICAI)*, pp. 1–4.

Zeng, H., Zhao, J., Short, M. A., McLean, D. I., Lam, S., McGregor, H. C., Kalia, S., McWilliams, A., Wang, W. and Lui, H. (2016). Real-time in vivo tissue raman spectroscopy for early cancer detection, *2016 Asia Communications and Photonics Conference (ACP)*, pp. 1–3.

Zhao, J., Lui, H., McLean, D. I. and Zeng, H. (2008). Real-time raman spectroscopy for non-invasive skin cancer detection - preliminary results, *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3107–3109.