Integrating Deep Learning Algorithms into a Web Application for Accurate Melanoma Skin Cancer Detection

MSc Research Project
MSc in Data Analytics

Bhagyalakshmi Shridhar Bichchal
Student ID: x23109149

School of Computing
National College of Ireland

Supervisor: Victor del Rosal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Bhagyalakshmi Shridhar Bichchal
Student ID:	x23109149
Programme:	MSc in Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Victor del Rosal
Submission Due Date:	12/08/2024
Project Title:	Integrating Deep Learning Algorithms into a Web Application for Accurate Melanoma Skin Cancer Detection
Word Count:	6597
Page Count:	23

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:	Bhagyalakshmi Shridhar Bichchal
Date:	11th August 2024

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):	

Integrating Deep Learning Algorithms into a Web Application for Accurate Melanoma Skin Cancer Detection

Bhagyalakshmi Shridhar Bichchal
x23109149

Abstract

Melanoma, an aggressive form of skin cancer, is known to be one of the deadliest forms of skin cancer if not detected accurately. Early detection is paramount, as early detection of melanoma offers a much better prognosis. However, traditional diagnostic techniques, such as visual diagnosis or dermoscopy, rely heavily on the experience of the examiner and can result in irregular diagnostic results or a delay in diagnosis. This study investigates the possibility of using deep learning models to recognize melanoma, which may enhance diagnostic preciseness. In this research, four deep learning models were implemented and evaluated including Custom CNN, MobileNet, VGG-19, and DenseNet121 using a variety of metrics, such as accuracy, precision, recall, and F1-score. Among the models, VGG-19 delivered the best results, achieving 88.75% accuracy on the test dataset. The primary contributions of this research are the detailed comparison of these models for melanoma classification and the successful deployment of the VGG-19 model as a web application for melanoma diagnosis. The benefits of this work include the development of a user-friendly web application using Python and Flask where the user can upload an image for an instant diagnosis, providing an accessible and accurate tool for the early detection of melanoma, and potentially saving lives.

1 Introduction

Skin cancer has been identified as one of the deadliest incurable types of cancer that can be categorized as melanoma cancer and non-melanoma cancer. The contribution of knowledge on the melanoma type of skin cancer is more fatal or life-threatening compared to another type. The disease mechanism for skin cancers has been determined with the enormous growth of skin cells, although it depends on nature as well as the intensity and may infiltrate to different regions in the body. The main cause of skin cancer according to dermatologists is the continuous exposure of the sense organ to the environment, especially toxins. The vulnerability lies underneath leads to an unpredictable condition if left undetected. As per the statistical report presented from the UK survey, nearly 46,000 cases of both types of skin cancers are reported every year Bhatt et al. (2023).

A focus on Melanoma skin cancer has been identified with the origination from melanin-producing cells that are primarily found in the dermal layer (inner layer of skin) and hair follicles. Although an incurable disease when it reaches advanced stages, melanoma is usually curable if detected early, thus a proper diagnosis is required. It has also been investigated that regardless of the rise in the rate of this skin cancer type globally, there are approximately 94-98% survival chances through proper diagnosis and treatment. Patients in stage-I condition, if diagnosed properly, generally have a survival rate of up to 10 years, while a patient in stage IV has a 10-15% chance of survival for 10 years Bhatt et al. (2023). Therefore, this determination has scored comprehensive insights into the necessity of recognizing this particular skin cancer type with more enhanced diagnosis.

1.1 Motivation

Since the beginning, a biopsy procedure is a common technique to detect any type of cancer. Therefore, experts initially perform a biopsy to detect skin cancer. The procedure focuses on the sample extraction from a "putative skin lesion" and performs medical tests to test the malignancy. Despite the suitability of the test, the method is time-consuming and an invasive procedure. In recent times, a computer-assisted diagnosis enables a flexible approach to rapidly identifying symptoms of Melanomas affordably and conveniently. While understanding the importance of the computerized diagnosis process, image processing techniques have gained wide recognition due to their non-invasive nature and flexibility. Image processing and classification help experts in tracking the changes within the patient's condition, thus providing data for training, presentation, and comparison of those images. The above understanding therefore elucidates the contribution of the current study in Melanoma skin cancer detection. In this regard, a dataset containing 10000 images of different stages of melanomas is used and further classified. Amid this approach, a deep learning model is used to determine the accuracy level in the optimal detection of skin cancer. The models used in this research are Custom CNN, MobileNet, VGG-19, and DenseNet121. To identify the most optimal model each of these models will be evaluated based on accuracy, precision, recall, and F1-Score. The best-performing model will be deployed to a web application for Melonama Skin Cancer Detection in real-time.

1.2 Research Objectives

- To advance the use of AI in skin cancer detection by demonstrating the practical application of deep learning in a real-world setting.
- To assess the performance of Custom CNN, MobileNet, VGG-19, and DenseNet121 in detecting melanoma, using metrics like accuracy, precision, recall, and F1-score.
- To determine the most accurate model for melanoma detection, with a focus on balancing performance across all key metrics.
- To implement the top-performing model in a web application that provides instant diagnosis from uploaded skin images.

1.3 Research Question

- How can deep learning models be optimized and applied for the accurate and accessible detection of melanoma skin cancer, and which model among Custom CNN, MobileNet, VGG-19, and DenseNet121 offers the best balance of accuracy, precision, recall, and F1-score for real-world deployment in a web-based diagnostic tool?

2 Related Work

Melanoma skin cancer is one of the major public health concerns that have made significant progress in the medical research paradigm. Upon drawing focus on this perspective, empirical evidence has been gathered and evaluated in this chapter indicating that, a further research directive is deployed to address the limitations. The purpose of this evidence is indeed crucial to assess and understand the progress made so far to highlight important aspects related to computational diagnosis and image processing techniques for efficient melanoma skin cancer detection.

2.1 Dataset identification and its importance in Melanoma Skin Cancer Detection

The detection of melanoma skin cancer has become a vital process in clinical practice for experts. From the previous chapter, it has been determined that there are approximately 94-98% survival chances for patients diagnosed with Stag-I cancer. Despite this information, it is still a matter of consideration as to how the disease can be detected with higher accuracy. While focusing on this information, evidence stated that traditional methods have become ineffective in the efficient detection of the disease Alquran et al. (2017). Therefore, computational intelligence as well as image-processing methods has gained importance. Information presented by Javid (2022) demonstrates the importance of relevant datasets such as the “melanoma skin cancer dataset” containing 10,000 images. With the deadly nature of the disease, it is increasingly imperative to early detection of the disease and reduce the global mortality rate. Therefore, the above dataset holds importance in helping experts train suitable models to classify melanoma from different stages.

A rapid enhancement of health innovation is changing healthcare and clinical practice with continuous accessibility to improved technology-based applications. In this focus, Wen et al. (2022) informed the importance of image processing techniques that extensively classify images through availed datasets for the detection of skin cancer. Amid this consideration, the study has introduced publicly available datasets that are used to train various state-of-the-art methods and neural architectures. Despite this information, Wen et al. (2022) further explained that there is still limited evidence identified in the total dataset count as well as their respective content, thus indicating a necessary study implication in the matter. On the contrary, Naeem et al. (2020) explained that melanoma skin cancer detection using different classification models requires real-world datasets to summarize highly accurate results that can ensure a better detection outcome and treatment intervention. Thus, this overall information suggests that skin cancer detection with higher accuracy requires suitable image-processing datasets that contain high-resolution images and can train classifiers more efficiently.

2.2 Melanoma Skin Cancer Detection Based on Existing Medical Diagnostic Procedures

Skin cancer has become one of the most prevalent types of cancer that has been continuously threatening global public health. Considering its types, skin cancer can be melanomas and non-melanomas. According to the information presented by Zghal and

Derbel (2018), despite the fact melanoma is less frequently observed among people, it is one of the specific reasons for mortality. Therefore, medical diagnosis to identify this cancer type has become a concern for global medical experts to prevent the loss of lives. Over the years, various approaches have been registered by medical experts to make a greater assumption and treatment emphasis for melanoma skin cancer. However, the diagnostic process using manual processes and traditional medical procedures provides less accurate results. Amid this concern, an improvement with computer-aided diagnosis (CAD) has been introduced to assist medical experts in lesion detection using human images and enable a detection process in the early stage by augmenting the likelihood of patients' survival. In this regard, Zghal and Derbel (2018) introduced a method based on classifying lesions with dermoscopy images by following ABCD rules. The rule-based method followed different steps where pre-processing, stage segmentation, feature extraction, and classification to yield "total dermoscopy value" were performed. As per the result obtained, the accuracy obtained is approximately 90%, thus reflecting the reliability of the testing method.

In another study presented by Tajeddin and Asl (2018), a range of detection procedures has been followed where each step is focused on a different model. As per the implication of the study, a propagation model was used for segmentation purposes while the feature extraction process was followed by gathering peripheral information on skin lesions. Using Daugman's transformation model, the peripheral lesion area is mapped to extract textual features. For the detection and classification process, a "sequential feature selection" algorithm has been used in ranking and classifying features, thus introducing a distinct method of melanoma skin cancer detection with an accuracy of 97%. Melanoma skin cancer is highly responsible for public mortality, where it constitutes approximately 70% of deaths globally. The cause of this disease is rooted in the development of malignant melanocytes. Therefore, the detection of the cancer type is important in the early stage. Sreelatha et al. (2019) in their study have presented survey-based outcomes that have identified several risk factors like skin color, freckles, sunburn, and aging. Over the years, many health campaigns and intervention measures have been introduced by healthcare experts, which are indeed effective.

Findings from the above study show that "public health campaigns" have become suitable to gather extensive data for the diagnostic process. From this perspective, it can also be stated that using information from these campaigns, a feasible detection method can be obtained that provides extensive detection outcomes regarding melanoma skin cancer. Over the years, the image-processing technique has received extensive attention, which is further assisted through computer-aided diagnosis procedures. Ananth and Therese (2020) explained that converting an image to a computerized structure is performed to extract useful data-oriented features suitable for the model training and classification of the data. In Melanoma skin cancer, the above study has presented a survey of some datasets based on which evidence suggested that the ATLAS dataset and PH2 dataset are commonly used in the investigation process. A further emphasis shows that computer-aided diagnosis has been used for years to make proficient decisions through optimized feature selection from image datasets. Thus, this information suggests that besides existing medical procedures, computerized techniques for skin cancer detection have also gained significant interest in dermoscopic imaging, which serves a better outcome.

2.3 Melanoma Skin Cancer Detection Based on State-of-the-Art Methods

The prediction of cancer stages is a rising area of focus in research where modern techniques have highlighted the importance of supervised methods or learning approaches. As Bhatt et al. (2023) stated machine learning or state-of-the-art methods are typically categorized into supervised and unsupervised techniques as well as reinforcement learning. Among them, the supervised learning approach enabled experts to deploy different algorithms to classify the type based on various conditional decisions and probabilities. Based on the study focus and implications, it is, therefore, integral to demonstrate that machine learning models successfully aid in the early detection of melanoma skin cancer and reduce the workload of specialists with a simultaneous enhancement of lesion diagnosis. Melanoma has been identified as a curable yet rapidly spreading cancer type that has become a concern for global public health. According to the information presented by Javaid et al. (2021), it has been identified that melanoma is one of the worst types of cancer that appears as a skin lesion on the surface although it further extends deeper into the dermal layer of the skin. Based on the study's emphasis, traditional methods such as biopsies are expensive and time-consuming. Therefore, progress toward adopting the machine learning method to detect skin lesions serves a better detection result.

In the above study, different classifiers such as Random Forest (RF), Support Vector Machine (SVM), and Quadratic Discriminant were used and tested using the ISIC-ISBI 2016 dataset. As per the resultant observation, maximum accuracy has been achieved through Random Forest with a level estimated to be 93.89%. In another study presented by Mahmoud and Soliman (2024), the information demonstrated that traditional methods are time-consuming, costly, and invasive procedures. Therefore, the focus on improved methods such as artificial intelligence (AI), and different algorithms has presented enhanced detection significance for melanoma skin cancer. As per the study evidence, "Adaptive Snake" (AS) and "Region Growing" (RG) algorithms have been applied in practice and further compared in terms of accuracy and efficiency. Experimental observation shows that Adaptive Snake has gained a higher accuracy level, which is estimated to be 96% than the Region Growing model. Besides, the "artificial neural network" is used for the classification of images of skin lesions which has obtained an accuracy of 94% and specificity of 95.83% respectively.

Over the years, there has been extensive research performed on incidences of melanoma skin cancer that have been increasing rapidly. The early detection of skin cancer, however, presents a potential outcome through lesion skin-image analysis. In this regard, the information stated by Strzelecki et al. (2024) has identified the relevance of image-processing techniques that support imaging-specific diagnostics of neoplasm hold importance with greater dynamism. Based on the evidence provided by the study, the artificial intelligence-based recognition of neoplastic lesions is an emerging trend that supports medical diagnosis with greater relevance. Thus, it can be stated that the analysis of skin lesion-based images using modern technologies such as AI has proven effective compared to traditional methods. Cancer is indeed a complex as well as intricate disease that keeps specialists struggling to identify feebleness and rudimentary characteristics. Therefore, effective diagnosis and discovery of treatments to detect cancer in the early stage are essential. Based on the information presented by Kontogianni and Maglogiannis (2020), common techniques in melanoma detection involve dermoscopy, which is a non-invasive procedure and performs examinations based on the optical system such as a magnifying glass with a light source, thus promoting extensive feature visualization during diagnosis. However, an advancement to more improved methods, which depends on a computer-assisted imaging approach, has served a better outcome. As per the evidence provided by Kontogianni and Maglogiannis (2020), the computer-based system provides the optimum result. However, the study-based insights into the image dermoscopy method indicate that the method is ineffective in detecting the early stage of melanoma. Thus, it has been identified that to determine malignant melanoma, more improved methods need to be introduced that can serve a positive outcome with real-world datasets.

2.4 Melanoma Skin Cancer Detection Based on Deep Learning Methods

While understanding the relevance and limitations of previous methods, research has extended evidence on the further improvement in the detection process of melanoma skin cancer. Upon understanding this many studies have come into focus that have positively expanded knowledge on distinct detection processes of melanoma. A progression observed in ozone layer depletion has become a significant threat to public and ecological health. With prolonged exposure to ultraviolet radiation, the condition has escalated the risk of developing skin cancer, especially melanoma. It was previously discussed that continuous sunburn is a significant cause of melanoma skin cancer Bhatt et al. (2023). Evidence-based information presented by Orhan and Yavşan (2023) explained that the risk of exposure to ultraviolet radiation is indicated through developing melanoma which therefore needs proper monitoring and a successful approach to medical treatment. In this regard, the above study has presented effective strategies for medical diagnosis to eliminate the risk of unpredictable life-altering events. Understanding the focus, Orhan and Yavşan (2023) demonstrated the importance of various datasets containing skin lesion images, where these datasets underwent training, validation as well as a testing process with different models such as ResNet, MobileNet, and VGGIB. As per the indication of the model suitability, the MobileNet provides an accuracy of 84.94%. Indeed, the model is effective; however, the accuracy level indicates the need for more improvement.

In another study presented by Singh et al. (2023), information has served to deeper insights into understanding the importance of improved diagnosis for melanoma skin cancer. Indeed many studies have contributed significant information regarding the validation and effectiveness of models; however, very few have minutely disclosed facts on the importance of enhanced techniques in the diagnosis process, precisely in terms of accuracy. In the above study, a hybrid configuration has been considered where a "fuzzy logic-based" image segmentation method is an ensemble with an advanced deep infrastructure - deep "Convolutional Neural Network" (DCNN). The DCNN model employs an improved algorithm, "You Look Only Once" (YOLO) that serves the application of CNN architecture Singh et al. (2023). As per the result obtained from the experiment, it has been observed that the model provides a higher accuracy compared to pre-existing models and is trained with approximately 2000 as well as 8695 skin lesion images from the ISIC-2017 and ISIC-2018 datasets. The critical evaluation of the information has presented an understanding of the need to consider a valid and effective dataset based on which the model can be trained and tested. Since the focus is on a medical diagnosis condition, the selection of a suitable model and dataset for training is mandatory.

The effect of skin cancer is evident globally with high escalation in the rate every year. Based on the statistics provided by Balaha and Hassan (2023), in the US, approximately 3.5 million individuals are diagnosed with melanoma annually. Moreover, the survival rate highly reduces with the progression. Under such a health crisis, identifying a suitable automated melanoma detection approach is a promising solution. Balaha and Hassan (2023) therefore, introduce various pre-trained CNN frameworks which were trained and tested through image segmentation and classification process. Based on the reported result, DenseNet201 is one of the best models with an accuracy of 94.16% identified when "skin cancer segmentation and classification" is used. However, using ISIC_2019 and ISIC-2020 datasets, MobileNet has achieved 98.27% accuracy while with the "Melanoma Classification HAM-10K" dataset; the accuracy level reached 98.83% respectively. Overall, it can be stated that MobileNet is one of the best deep-learning models used in the diagnosis of melanoma although the dataset is a crucial factor in achieving maximum efficiency. Zafar et al. (2023) explained that computed-assisted diagnosis of melanoma recognizes approaches relevant to early recognition, which is succeeded through the integration of various ML, DL, and computer-enhanced vision techniques. However, amid all this information, choosing proper medical samples containing certain uncommon skin lesion characteristics is essential to examining the survey outcome on the model.

In humans, the most common form of neoplasms that have been identified is skin cancer. Even though it is divided into two types, melanoma skin cancer is a more vulnerable condition than non-melanoma. In this regard, the information provided by Mehr and Ameri (2022) explained that the threatening cause of melanoma is likely identified from newly registered cases of approximately 324,635 from recent years in nearly 185 countries. Therefore, the detection of this disease holds specific importance, where conventional modalities show ineffective identification outcomes. Therefore a contemporary solution is introduced by the above study, which is based on a deep learning framework. The model used is Inception-ResNet-v2 - a CNN model that has been pre-trained with lesion images. The classification result obtained identifies the model accuracy by 89.3% when discriminating lesion images based on skin condition, while 94.5% in classifying the image as a malignant or benign lesion. Contrastingly, another study presented by Waheed et al. (2023) explained that a CNN model when pre-trained with large datasets containing malignant and non-malignant skin lesion images provides a potential discrimination result, which is more accurate than existing state-of-the-art methods. Thus, the overall information suggests that deep learning models, especially CNN architecture provide promising solutions in the detection of skin cancer than existing methods.

2.5 Gaps Identified in the Existing Literature

Current research on melanoma recognition tends to depend on traditional methods, such as visual examination and dermoscopy, which may lead to inconsistency and might be time-consuming. Although several papers have investigated machine learning and deep learning techniques, many of them focused on single models or smaller datasets, without necessarily drawing comprehensive comparisons across different architectures. Furthermore, there has been at present little focus on actually distributing these models in usable applications that could enhance timely medical diagnoses. To address these gaps, we initialise our research by evaluating multiple deep learning models - Custom CNN, MobileNet, VGG-19, and DenseNet121 - on an abundant dataset to provide a comprehensive comparison of their results. Here, we select VGG-19 as the most effective and deploy it into a practical web application utilising Flask that enables instant melanoma recognition services in a user-friendly manner. This could not only enhance diagnostic accuracy, but brings advanced Artificial Intelligence tools closer to both clinicians and patients, contributing to the practicalities of post-reading and image analysis.

3 Methodology

Melanoma is a type of deadly skin cancer and accounts only for 4% of all skin cancers and it is the reason for 75% of skin cancer deaths. If the diagnosis can be treated earlier then it can be cured and the melanoma increases deeper into the skin to the other part of the body if left untreated. The cause of melanoma is the presence of melanocytes in any part of the body and the main cause is the high exposure of skin to UV rays. The early detection and cure for melanoma can save many lives. This methodology diagram for detecting Melanoma Skin Cancer using image processing and deep learning models for accurate detection is shown in Figure 1. The research is carried out by following a certain set of steps, which are essential to follow. Each Step plays a crucial role, which has been discussed below in further subsections.

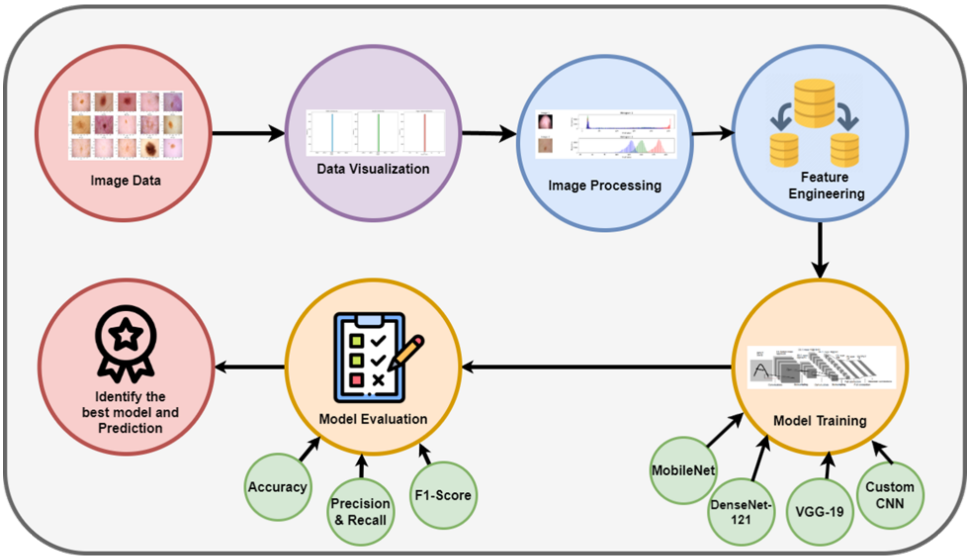

Figure 1: Methodology diagram for detecting Melanoma Skin Cancer

3.1 Data Description

In the research, the data is taken from Kaggle for detecting melanoma skin cancer that contains 10000 images which is used for implementing deep learning models *Melanoma Skin Cancer Dataset of 10000 Images — kaggle.com* (n.d.). The dataset consists of 9600 images for training the model and 1000 images for evaluating the model. The size of the dataset is around 103.28 MB. There are two classes in the melanoma skin cancer dataset benign and malignant. In training data, there are 5000 images for benign and 4605 images for malignant and there is equal distribution of classes in test data.

3.2 Data Analysis and Visualization

Data analysis and visualization are important in dealing with image data as they help one to view and understand the structures and points in the data. Visualization is also useful in understanding challenges that are within imaging data analysis and results, making it easy to detect patterns, unusual phenomena, and other things that may need more attention. In this research, the data analysis is carried out plotting the distribution of training and testing data distribution for each class.

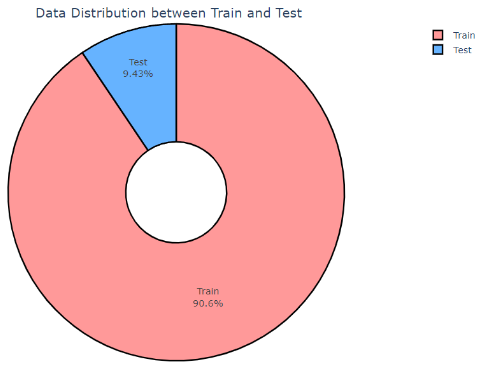

Figure 2: Data Distribution between Train and Test Set

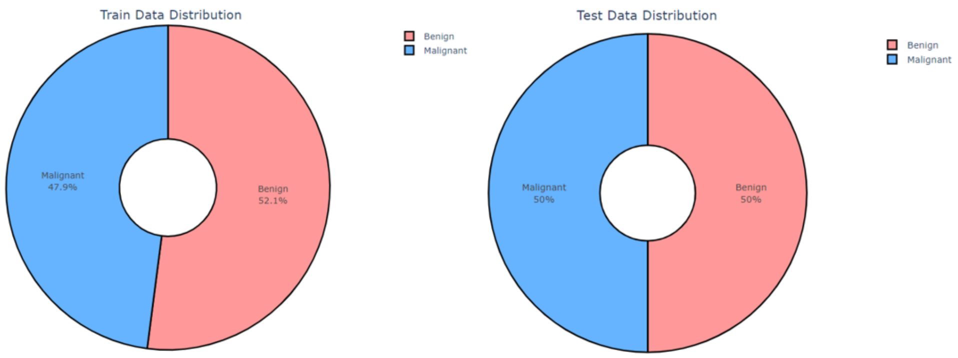

Figure 3: Train and Test Data Distribution with respect to Labels

Two pie charts are visualized for the distribution of training and testing data and the distribution of classes in the training data which is shown in Figure 2 and Figure 3. It can be seen that 90.6% of the data is used as training data and remaining data is used for evaluating the model and 52.1% images are benign classes in the training data and the remaining images are of class malignant. Whereas for the test data distribution it has been identified that dataset is balanced with respect to labels.

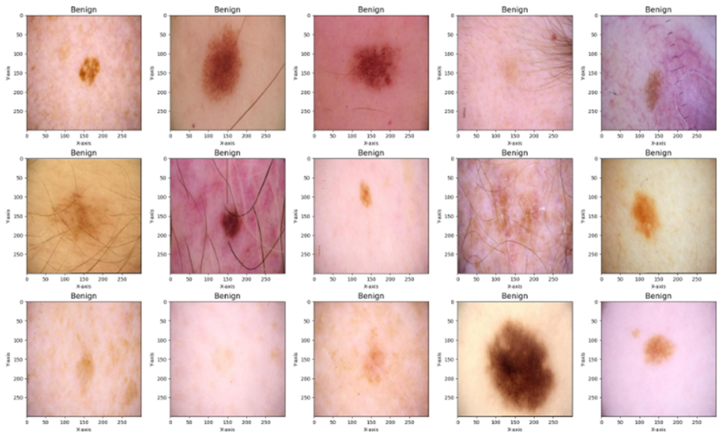

Figure 4: Random Images from Benign class

Some random images of benign class are visualized to understand the data which is shown in Figure 4.

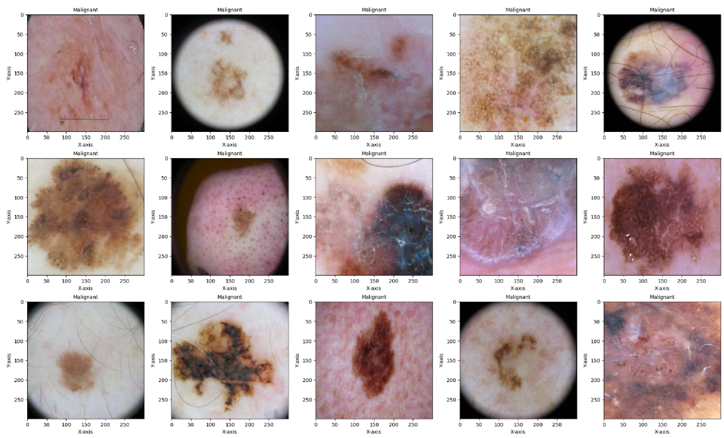

Figure 5: Random Images of Malignant class

Some random images of malignant class are visualized to understand the data which is shown in Figure 5.

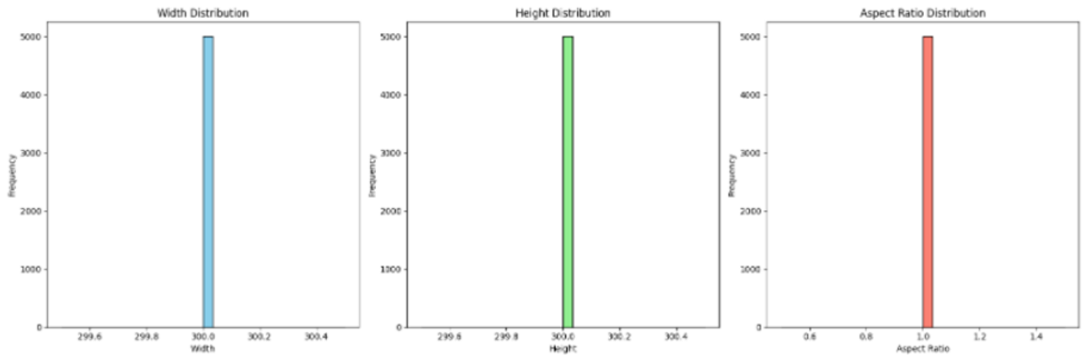

Figure 6: Histograms for the width, height and aspect ratio distribution of Images

A histogram is plotted for the width, height, and aspect ratio shown in Figure 6 for the benign class that calculates the dimensions and helps in understanding variations in the size and shape of the images.

3.3 Image Processing

Image Preprocessing is very important for increasing the quality and usefulness of the image by removing noise and sharpening the image. It is also useful in extracting only the important features of the images, making it easier to analyze and interpret them. In this research, the image sharpening technique is applied to the data by utilizing a convolutional kernel that increases the edges and fine details in the images. Sharpening of the images helps in highlighting the subtle features and irregularities of the skin lesions which is essential in detecting melanoma. The increase in the edges and contrast of the images makes it easy to find the characteristics patterns and textures that are associated with melanoma.

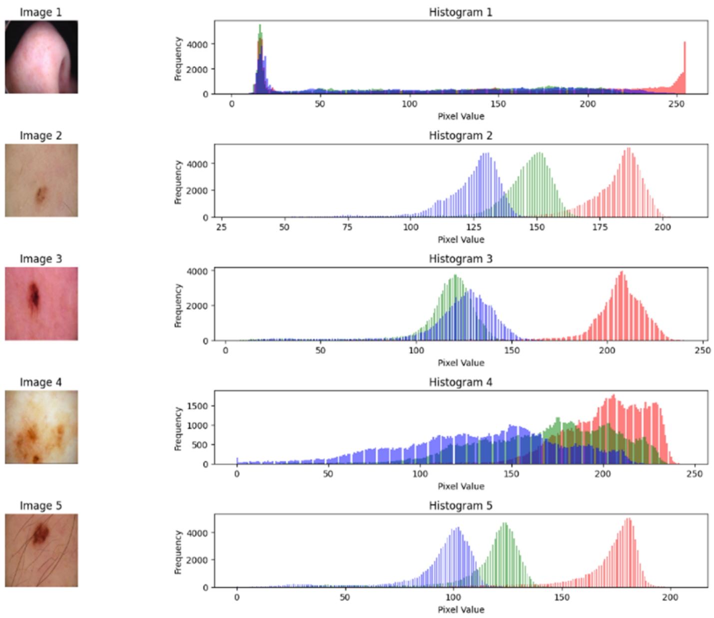

Figure 7: Images before increasing the sharpness and contrast of the images

Images and the intensity are visualized before increasing the sharpness and contrast of the images as shown in Figure 7.

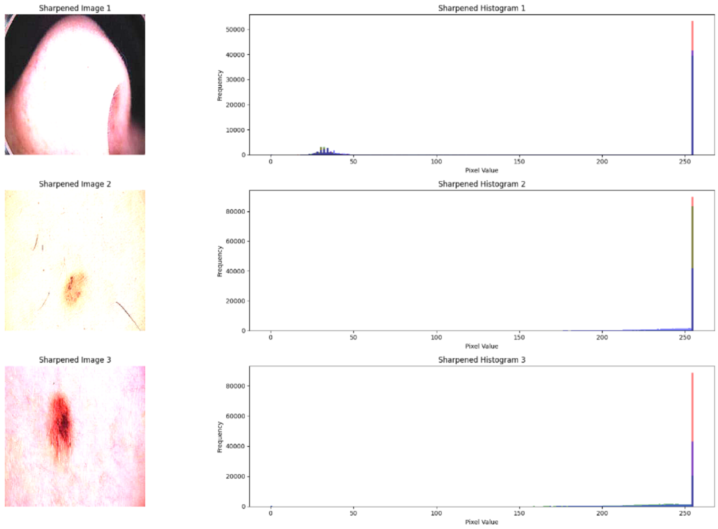

Figure 8: Images after increasing the sharpness and contrast of the images

Images and the intensity are visualized after increasing the sharpness and contrast of the images as shown in Figure 8.

3.4 Feature Engineering

Feature engineering is the process of getting, choosing, and transforming raw image data to be able to get meaningful data or informative data to assist in the training of other models. Some of the features that are captured have the important characteristics of the images that are most suitable for the detection of melanoma skin cancer. In the research, the image sharpening technique is applied to the training images to enhance the edges and details in the images by leveraging convolutional kernels. By increasing the features helps in model training more effectively to learn the difference between benign and malignant skin lesions. The images are generated for training, validation, and testing data that includes sharpening of images. The image generated helps to recognize melanoma-specific features such as irregular shapes, borders, and color patterns, which helps in improving detection accuracy.

3.5 Model Training

Predicting the data depends mostly on the process of model training. During the model training phase, the model learns and stores features to targets and maps them to the targets by changing its internal parameters. Model training enables them to predict the result, assists in incorporating the patterns, and enhances the model's accuracy. The research is carried out for the detection of melanoma by leveraging the power of a deep convolutional network for accurate results. The deep convolutional networks are VGG-19, MobileNet, DenseNet-121, and the custom CNN model. We will discuss about the architectural overview of each model in Section 4

3.6 Model Evaluation

The last step in the process of evaluating the algorithm is determining results after the data has been trained on the algorithm. Hence, there is a need to assess the result of the algorithm to have an understanding of the effectiveness of the algorithm in managing the data. The key performance metrics for detecting melanoma used are accuracy- used when classes are balanced and give instances that are classified correctly, precision-fraction of correct positive predictions, recall- indicates actual positives which are correctly identified in instances where high false negatives are recorded and F1-score is the harmonic mean of precision and recall. These metrics are calculated on the test data for the algorithm and then the result produced by the algorithm is compared with actual results on the test data set. The model evaluation step helps us to identify the most optimal model for Melanoma skin cancer detection.

4 Design Specification

The research focuses on the classification task for predicting melanoma which is taken by implementing four deep learning algorithms: Custom Convolutional Neural Network, VGG-19, MobileNet, and DenseNet-121.

4.1 VGG-19

VGG-19 is a type of deep convolutional neural network (CNN) having 19 layers, that includes 16 convolutional layers followed by three fully connected layers. VGG-19 is simple and uses filters of 3x3 and multiple layers for capturing the complex features from the images. VGG-19 is highly effective for detecting melanoma as it can learn hierarchical features such as irregular shapes and color distributions in skin lesions that can distinguish between benign and malignant skin lesions accurately. A schematic diagram of VGG-19 is shown in Figure 9.

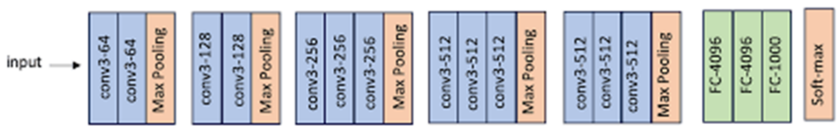

Figure 9: VGG-19 Architecture Diagram Ali et al. (2021)

4.2 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a deep learning algorithm that is used to learn spatial hierarchies of features from the input images such as edges, textures, and complex patterns. For the detection of melanoma, custom CNNs are particularly used because CNNs can capture and distinguish between benign and malignant skin lesions effectively. CNN models learn characteristic features from melanoma like irregular borders, color variations, and asymmetry that lead to accurate results. The custom CNN model in the research has 16 layers that consist of 4 convolutional layers with ReLu activation followed by a MaxPooling layer and Batch Normalization layer. The architecture contains one

Dropout layer for regularization and a flattened layer for converting the feature maps to 1D vectors, 3 dense layers are added after the flattening layer, and as the problem is binary classification, a sigmoid activation function is added to the last layer. An architectural diagram of CNN is shown in Figure 9.

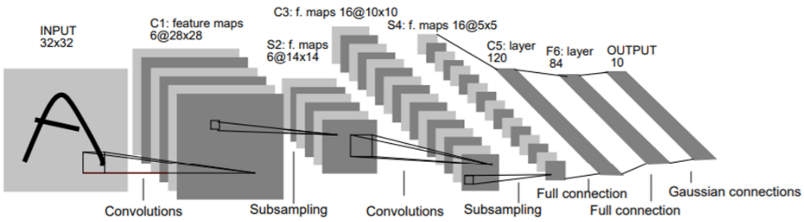

Figure 10: Architectural Diagram of CNN Hidaka and Kurita (2017)

4.3 MobileNet

MobileNet architecture can be defined as the collection of efficient convolutional network (CNN) architectures designed specifically for use on mobile and embedded devices. These architectures help navigate the limitations of computational resources while retaining high accuracy in image classification and relative tasks. MobileNet core innovation lies in the utilization of depth-wise separable convolutions, a great technique that significantly reduces or diminishes computational burden and memory usage. Segregating spatial and channel-wise convolutions provides a lightweight alternative to conventional convolutional layers, enabling MobileNet to achieve an impressive performance with minimal computational resources. A schematic diagram of MobileNet is shown in Figure 10.

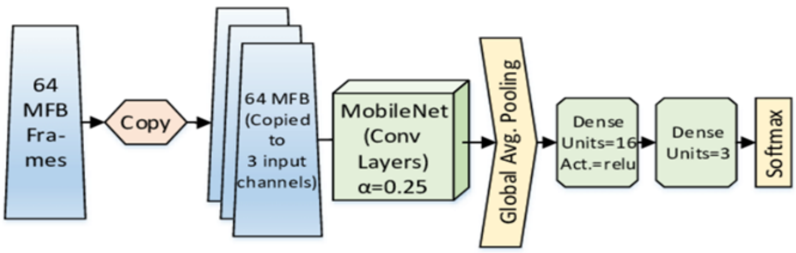

Figure 11: Architectural Diagram of MobileNet Hussain and Haque (2018)

4.4 DenseNet-121

DenseNet-121 is a type of deep convolutional neural network having 121 layers that are characterized by its dense connectivity, where input is received by each layer from all preceding layers. For detecting melanoma, DenseNet-121 is considered highly effective because it captures rich features from both shallow and deep layers. This makes the model differentiate between skin lesion images such as irregular patterns, textures, and colors. An architectural diagram of DenseNet-121 is shown in Figure 12.

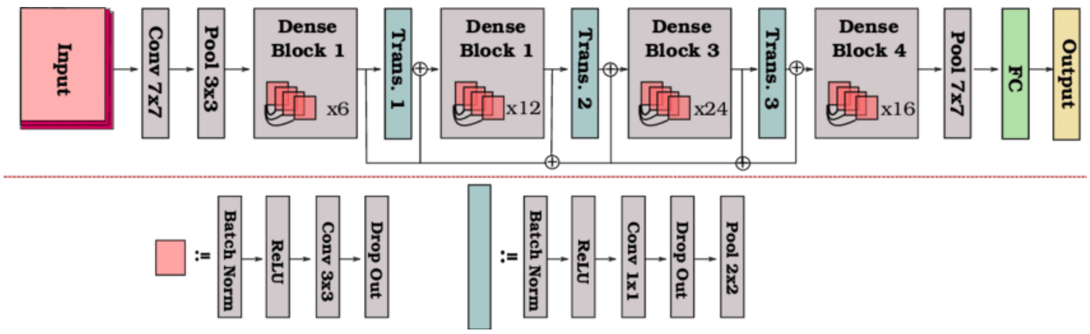

Figure 12: Architectural Diagram of DenseNet-121 Raju et al. (2023)

5 Implementation

For this project, we created a melanoma detection system based on deep learning, using a variety of tools, technologies, and libraries and yielding high accuracy and performance. The entire project was developed using Python as the primary programming language because it is versatile and has an ecosystem of libraries for data science, machine learning, and web development. There were several libraries that we used in our work. To begin, we used pandas for data manipulation, including loading, cleaning, and preprocessing, which was essential for our dataset. numpy was crucial for numerical operations as it is efficient at handling large matrices and arrays that are necessary for training and evaluating models. To visualize data and model performance, we applied Matplotlib for static plots and Seaborn for slightly more advanced and aesthetically pleasing plots. cv2 (OpenCV) played an essential role in our image processing tasks, such as resizing and augmentation, to ensure our skin lesion images were in the format that the model required. To save and load the trained models, we used Pickle to allow us to store our VGG-19 after it was trained, to be used in the web application. As a high-level API, Keras was used to construct and train our deep learning models Custom CNN, VGG-19, MobileNet, and DenseNet121, while TensorFlow served as the backend framework. For creating interactive plots for exploring the evaluation metrics for our models, we used Plotly. Finally, scikit-learn (sklearn) was used to determine accuracy, precision, recall, and F1-score, in addition to splitting the data into training and testing data. the tqdm library created progress bars that provided insights into model training progress and data processing tasks. Our melanoma detection web application needed to be easy to use as well as available and accessible, so we have used Flask, which is a lightweight web framework in Python. The web application allows users to upload skin images and receive quick predictions as to whether the skin image is likely to be benign or malignant as shown in Figure 13 Figure 14.

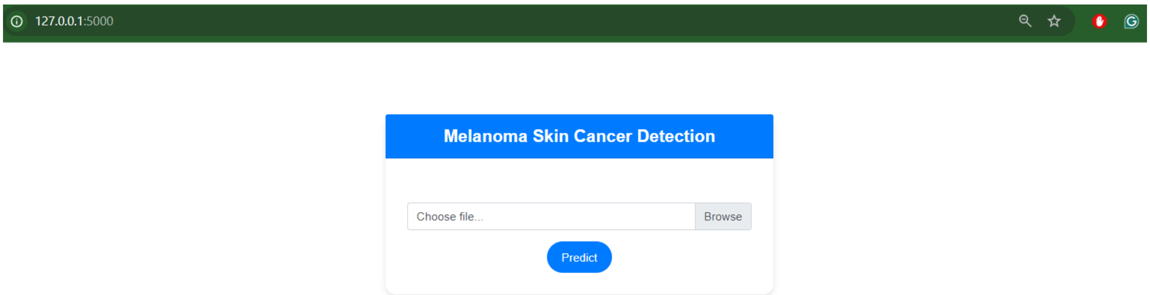

Figure 13: Web Application Interface to Upload the Image of Melanoma Skin Cancer

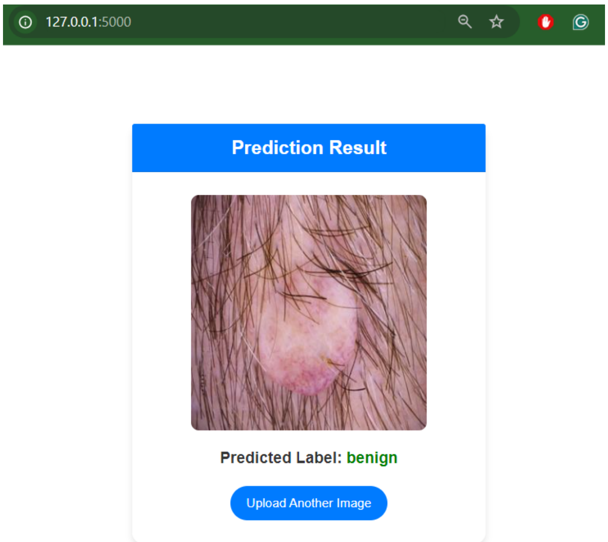

Figure 14: Web Application Predicting Results based on Input Image

The web application uses VGG-19 model in the backend, which is the most accurate model in our study, was implemented within the web application, and we used Pickle to load the model and keep it stored to prevent the need for retraining and have efficient performance. The web application has a good and easy-to-use interface. It provides real-time predictions, which would make it a very useful resource for early melanoma detection. We did everything in Python, from processing the data, training the model, and deploying it as a web application, which shows that the language can do a lot and is adaptive and versatile for developing AI solutions.

6 Evaluation

The research is carried out leveraging the power of deep learning algorithms on the model, the deep learning algorithms are custom CNN, DenseNet-121, VGG-19, and MobileNet. Since the problem is to detect melanoma having two classes benign and malignant. The model is evaluated on key performance metrics: Accuracy, Precision, F1-Score, and Recall. These metrics then compare with each other to find which model is suitable for predicting the target.

6.1 Evaluation Based on Accuracy

Accuracy tends to quantify the level of error of a given classifier by comparing the number of times that an instance has been classified as it is to the total number of instances. It offers the summary of the performance of the classifier both in the positive instances as well as the negative ones. In the research, out of all the models, VGG-19 attained the highest accuracy of 88.75% which indicates that VGG-19 correctly classifies the instance. While DenseNet-121 achieved an accuracy of 86.82% which is slightly lower than VGG-19. On the other hand, custom CNN and MobileNet give lower accuracy of 85.94% and 85.53% respectively which are less effective compared to VGG-19 and DenseNet-121. VGG-19 emerges better in terms of accuracy. Comparative analysis between the models based on accuracy is shown in Figure 15.

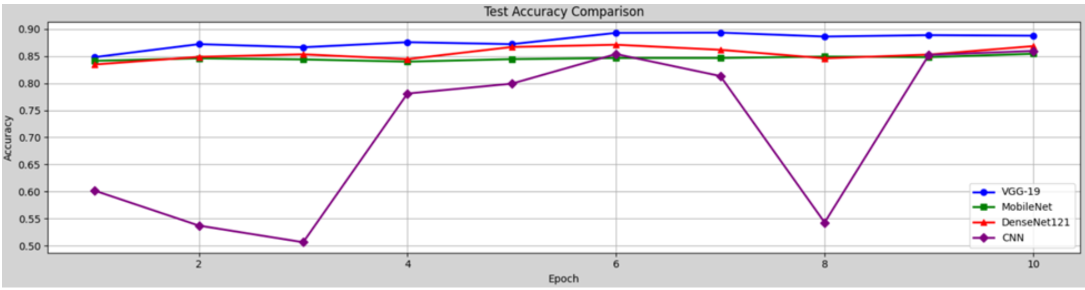

Figure 15: Accuracy comparison among Deep Learning Models

6.2 Evaluation Based on Precision

Precision gives the ratio of the truly positive instances that have been correctly predicted as positive by the model to the total number of instances that are seen as positive ones. In other words, precision measures the number of cases that is positive among the cases that are identified as positive. In this research, custom CNN attains the highest precision of 94.77% which indicates custom CNN has a high rate of correctly predicting positive cases out of all the cases it predicted as positive. Although DenseNet-121 and VGG-19 attain similar precision values of 86.86% and 86.75% respectively. On the other hand, MobileNet achieved a lower precision of 85.10% among models. Based on precision, it can be said that the custom CNN model is very precise in predicting melanoma. A comparison is analyzed between the models by plotting a line chart based on precision is shown in Figure 16.

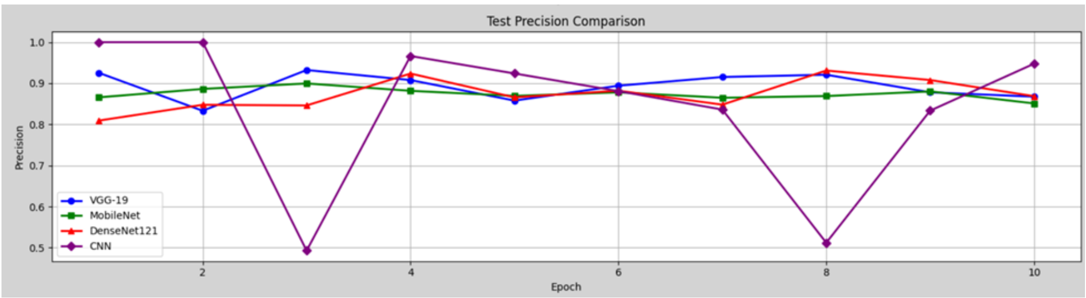

Figure 16: Precision Score Comparison among Deep Learning Models

6.3 Evaluation Based on Recall

Recall measures the percentage of the actual positive cases as distinguished by the classification model on the given data set. It estimates the model's capability of properly identifying the positive classes among all other true positive instances in the dataset. In the study, it is found that VGG-19 attains the highest recall of 90.36% which indicates VGG-19 detects the highest proportion of actual positive cases. Compared to VGG-19, DenseNet-121 and MobileNet have low recall values of 85.45% and 84.36% respectively. The lowest recall values are attained by CNN with 74.81% which means that CNN miss a significant number of positive cases. A comparative analysis of models based on recall is visualized in Figure 17.

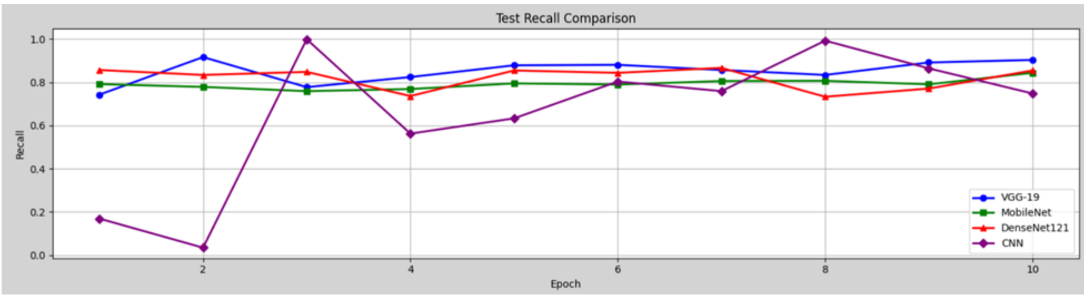

Figure 17: Recall Score Comparison among Deep Learning Models

6.4 Evaluation Based on F1-Score

F1-score can be defined as the composite metrics that combine both precision and recall(sensitivity) of a classification model and provide a single value for both precision and recall. Out of all the models in the study, VGG-19 achieved the highest F1-Score of 88.51% which represents a strong balance between precision and recall. DenseNet-121 attains a slightly lower F1-Score of 86.15% and the F1-Score attained by MobileNet is 84.73% which indicates a reasonable balance but less than DenseNet-121 and VGG-19. Despite having the high precision value of custom CNN due to the lower recall CNN achieved the lowest F1-Score of 83.61%. A scatter polar chart is visualized in Figure 18 for the comparison of models based on the F1-score.

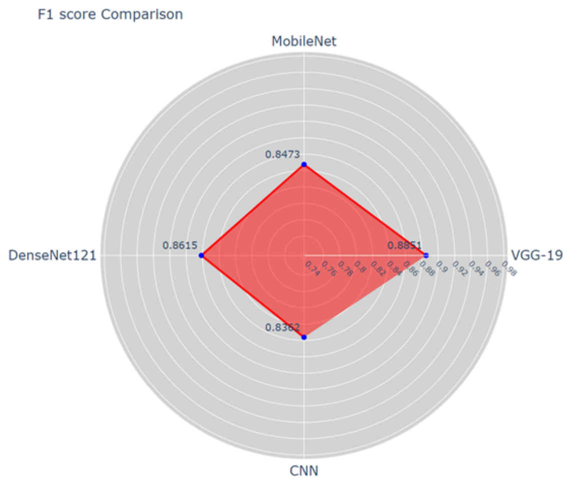

Figure 18: F1-Score Comparison among Deep Learning Models

6.5 Discussion

In this experiment, deep convolutional neural networks are employed to detect melanoma with two classes benign and malignant. The deep convolutional networks are custom CNN models with 16 layers, VGG-19, DenseNet-121, and MobileNet. These models are evaluated on four key performance metrics: accuracy, precision, recall, and F1-score for comparing the better generalizability of the models. Out of all the models VGG-19 outperformed better in terms of accuracy, recall, and F1-Score which makes it a better choice for detecting melanoma where precision and recall are crucial. On the other, DenseNet-121 has slightly lower metrics but performs well across all metrics. While MobileNet maintains a decent balance between precision and recall although it has lower performance compared to VGG-19 and DenseNet-121. The Custom CNN model has the greater precision value but the lowest recall that represents CNN is precise in terms of predicting the instances but fails to predict most of the true positive cases which makes it less effective for detecting melanoma where recall is critical. Overall VGG-19 emerged as the most balanced model for identifying melanoma. A line plot for the performance of the VGG-19 model for the training accuracy and validation accuracy and training loss and validation loss is visualized in Figure 19.

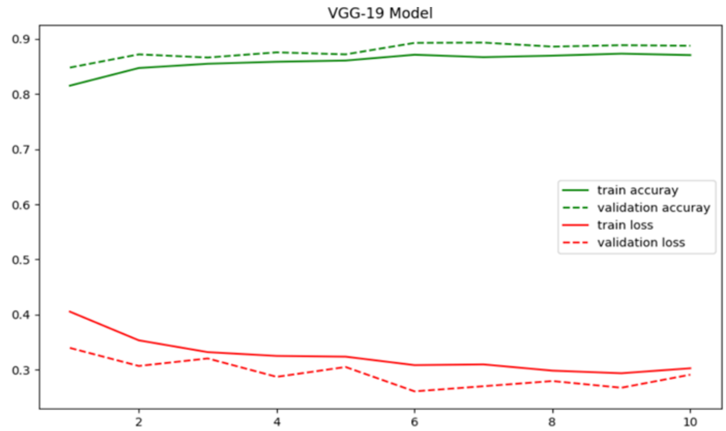

Figure 19: Metrics for VGG-19 Model

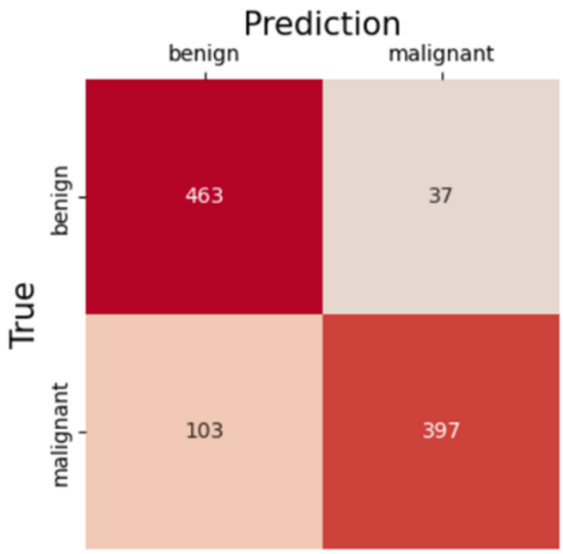

Figure 20: Confusion matrix for VGG-19

The confusion matrix is visualized in Figure 20 for VGG-19 which talks about the number of classes that are predicted correctly and it tells the true positive rates which model predicted correctly and the false negatives that were identified incorrectly by the model. The model identifies 463 benign classes correctly and 103 benign are incorrectly predicted. 37 cases as malignant are incorrectly predicted as they are benign actually. There are 397 cases of malignant are correctly predicted.

7 Conclusion and Future Work

In this project, we were successfully able to implement and evaluate four distinct deep learning models, namely Custom CNN, MobileNet, VGG-19, and DenseNet121, for melanoma skin cancer detection. Our extensive evaluation, which included performance metrics such as accuracy, precision, recall, and F1-score, determined that VGG-19 outperformed each model and demonstrated the highest accuracy of 88.75% on the test dataset. While

Custom CNN showed favorable precision, it performed poorly in other vital metrics such as recall and F1-score, and appeared to be less effective overall. As a result, VGG-19 was chosen for deployment and was incorporated into a Python and Flask-based web application. The web application enables users to upload skin images and receive an instant diagnosis that indicates whether the skin image is benign or malignant. This deployment demonstrates the potential for deep learning to be used for the early detection of melanoma, which would clearly have an immense impact on the outcomes for patients.

For future work, we can focus on further refining the VGG-19 model, as well as investigating other complex architectures like EfficientNet or ResNet for improving accuracy and generalization. Augmenting the dataset by including more diverse images can improve the model's robustness when applied to different populations. Implementing global transparent AI (XAI) techniques would provide transparency in the AI decision-making process and build trust for medical applications. Our Web application utility can be expanded by developing a mobile application and integrating the tool with Electronic Health Record (EHR) systems. Furthermore, a clinical trial could be conducted to verify the AI model in real-world scenarios and in a clinical setting with the EHR. In regards to enhancing scope and use of the AI diagnostic tool, another phase of work could be carried out in a different setting to identify other skin conditions.

References

Ali, L., Alnajjar, F., Jassmi, H., Gochoo, M., Khan, W. and Serhani, M. (2021). Performance evaluation of deep cnn-based crack detection and localization techniques for concrete structures, *Sensors* **21**: 1688.

Alquran, H., Qasmieh, I. A., Alqudah, A. M., Alhammouri, S., Alawneh, E., Abughazaleh, A. and Hasayen, F. (2017). The melanoma skin cancer detection and classification using support vector machine, *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1–5.
URL: *https://doi.org/10.1109/AEECT.2017.8257738*

Ananth, C. and Therese, M. J. (2020). A survey on melanoma: Skin cancer through computerized diagnosis, *SSRN Electronic Journal* .
URL: *https://doi.org/10.2139/ssrn.3551811*

Balaha, H. M. and Hassan, A. E. S. (2023). Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm, *Neural Computing and Applications* **35**(1): 815–853.
URL: *https://doi.org/10.1007/s00521-022-07762-9*

Bhatt, H., Shah, V., Shah, K., Shah, R. and Shah, M. (2023). State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review, *Intelligent Medicine*, Vol. 3, Elsevier, pp. 180–190.
URL: *https://doi.org/10.1016/j.imed.2022.08.004*

Hidaka, A. and Kurita, T. (2017). Consecutive dimensionality reduction by canonical correlation analysis for visualization of convolutional neural networks, Vol. 2017, pp. 160–167.

Hussain, M. S. and Haque, M. (2018). Swishnet: A fast convolutional neural network for speech, music and noise classification and segmentation.

Javaid, A., Sadiq, M. and Akram, F. (2021). Skin cancer classification using image processing and machine learning, *Proceedings of 18th International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, pp. 439–444.
URL: *https://doi.org/10.1109/IBCAST51254.2021.9393198*

Javid, M. H. (2022). Melanoma skin cancer dataset of 10000 images.
URL: *https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images*

Kontogianni, G. and Maglogiannis, I. (2020). A review on state-of-the-art computer-based approaches for the early recognition of malignant melanoma, *Studies in Computational Intelligence*, Vol. 891, Springer, Berlin, Heidelberg, pp. 81–101.
URL: *https://doi.org/10.1007/978-3-662-61114-2*$_6$

Mahmoud, N. M. and Soliman, A. M. (2024). Early automated detection system for skin cancer diagnosis using artificial intelligent techniques, *Scientific Reports* **14**(1): 1–10.
URL: *https://doi.org/10.1038/s41598-024-59783-0*

Mehr, R. A. and Ameri, A. (2022). Skin cancer detection based on deep learning, *Journal of Biomedical Physics and Engineering* **12**(6): 559–568.
URL: *https://doi.org/10.31661/jbpe.v0i0.2207-1517*

Melanoma Skin Cancer Dataset of 10000 Images — kaggle.com (n.d.). `https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images/data`. [Accessed 11-08-2024].

Naeem, A., Farooq, M. S., Khelifi, A. and Abid, A. (2020). Malignant melanoma classification using deep learning: Datasets, performance measurements, challenges and opportunities, *IEEE Access* **8**: 110575–110597.
URL: *https://doi.org/10.1109/ACCESS.2020.3001507*

Orhan, H. and Yavşan, E. (2023). Artificial intelligence-assisted detection model for melanoma diagnosis using deep learning techniques, *Mathematical Modelling and Numerical Simulation with Applications* **3**(2): 159–169.
URL: *https://doi.org/10.53391/mmnsa.1311943*

Raju, C., Ashoka and Prakash, B. (2023). Hybridtransfernet: Advancing soil image classification through comprehensive evaluation of hybrid transfer learning.

Singh, S. K., Abolghasemi, V. and Anisi, M. H. (2023). Fuzzy logic with deep learning for detection of skin cancer, *Applied Sciences (Switzerland)* **13**(15): 8927.
URL: *https://doi.org/10.3390/app13158927*

Sreelatha, T., Subramanyam, M. V. and Giri Prasad, M. N. (2019). A survey work on early detection methods of melanoma skin cancer, *Research Journal of Pharmacy and Technology* **12**(5): 2589–2596.
URL: *https://doi.org/10.5958/0974-360X.2019.00435.9*

Strzelecki, M., Kociołek, M., Strakowska, M., Kozłowski, M., Grzybowski, A. and Szczypiński, P. M. (2024). Artificial intelligence in the detection of skin cancer: State of the art: Artifical intelligence and skin cancer, *Clinics in Dermatology* **42**(3): 280–295.
URL: *https://doi.org/10.1016/j.clindermatol.2023.12.022*

Tajeddin, N. Z. and Asl, B. M. (2018). Melanoma recognition in dermoscopy images using lesion's peripheral region information, *Computer Methods and Programs in Biomedicine* **163**: 143–153.
URL: *https://doi.org/10.1016/j.cmpb.2018.05.005*

Waheed, S. R., Saadi, S. M., Rahim, M. S. M., Suaib, N. M., Najjar, F. H., Adnan, M. M. and Salim, A. A. (2023). Melanoma skin cancer classification based on cnn deep learning algorithms, *Malaysian Journal of Fundamental and Applied Sciences* **19**(3): 299–305.
URL: *https://doi.org/10.11113/mjfas.v19n3.2900*

Wen, D., Khan, S. M., Xu, A. J., Ibrahim, H., Smith, L., Caballero, J., Zepeda, L., de Blas Perez, C., Denniston, A. K., Liu, X. and Matin, R. N. (2022). Characteristics of publicly available skin cancer image datasets: a systematic review, *The Lancet Digital Health* **4**(1): e64–e74.
URL: *https://doi.org/10.1016/S2589-7500(21)00252-1*

Zafar, M., Sharif, M. I., Sharif, M. I., Kadry, S., Bukhari, S. A. C. and Rauf, H. T. (2023). Skin lesion analysis and cancer detection based on machine/deep learning techniques: A comprehensive survey, *Life* **13**(1): 146.
URL: *https://doi.org/10.3390/life13010146*

Zghal, N. S. and Derbel, N. (2018). Melanoma skin cancer detection based on image processing, *Current Medical Imaging (Formerly Current Medical Imaging Reviews)* **16**(1): 50–58.
URL: *https://doi.org/10.2174/1573405614666180911120546*