DAYANANDA SAGAR UNIVERSITY

KUDLU GATE, BANGALORE – 560114, Karnataka, India

Bachelor of Technology

in

COMPUTER SCIENCE AND ENGINEERING

PROJECT REPORT

on

ANALYZING TRANSFER LEARNING APPROACHES FOR MELANOMA DETECTION: A COMPARATIVE STUDY

By

Ritika D Chandavarkar – ENG21CS0334

Sohan Nath - ENG21CS0406

Under the supervision of

Dr. Pramod Kumar Naik
Associate Professor (CSE)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,
SCHOOL OF ENGINEERING
DAYANANDA SAGAR UNIVERSITY

(2023-2024)

Analyzing Transfer Learning Approaches for Melanoma Detection: A Comparative Study

Ritika D Chandavarkar (ENG21CS0334) and Sohan Nath (ENG21CS0406)

Dayananda Sagar University

Abstract: This project investigates the efficacy of various transfer learning methods and techniques utilizing an array of popular convolutional neural network architectures, including AlexNet, VGG 16, ResNet 50, ResNet 101, GoogleNet, VGG 19, DenseNet201, InceptionResNetV2, InceptionV3, DenseNet121, and EfficientNetB1, for the automated detection of benign and malignant skin lesions. Our study employs a labeled dataset of skin lesion images, fine-tuning each model to classify lesions. We compare the performance metrics such as the accuracy value, precision value, recall value, and F1 score among the various models. Additionally, we assess computational efficiency and resource requirements, considering factors such as training time and hardware specifications. Through rigorous experimentation, we provide insights into the strengths and limitations of each architecture for this specific task. Our results not only highlight the efficacy of transfer learning in accurately classifying lesions but also reveal nuanced performance differences between the tested architectures. These findings contribute to the understanding of deep learning applications in dermatology and early skin cancer detection, providing valuable guidance for future research and clinical implementation.

1. Introduction

Melanoma is a formidable form of skin cancer arising from melanocytes. It presents itself as a substantial public health challenge globally due to its high mortality rates in advanced stages. Early diagnosis is paramount for successful treatment and prognosis, underscoring the urgent need for accurate and efficient detection methods. Dermoscopic imaging, which provides magnified views of skin lesions, has emerged as a valuable adjunct to clinical examination for melanoma diagnosis. However, dealing with dermoscopic images can be intricate and needs specialized expertise.

In recent years, deep learning techniques, particularly convolutional neural networks (CNNs), have revolutionized medical image analysis by automatically extracting discriminative features from large datasets and achieving high accuracy in classification tasks. CNNs offer the potential to assist clinicians in melanoma detection, thereby improving diagnostic accuracy and patient outcomes.

In this paper, we investigate alternative convolutional neural network (CNN) architectures and transfer learning approaches for the automated identification of skin lesions as benign or malignant, delving into the field of dermatology. Transfer learning is a technique that shows promise in applying knowledge from one domain to another. It may reduce the requirement for a significant amount of labeled data in medical imaging tasks by using pre-trained models on large-scale datasets that have been tailored for particular tasks.

Our investigation focuses on several prominent CNN architectures: AlexNet, VGG16, ResNet50, ResNet101, GoogleNet, VGG19, DenseNet201, InceptionResNetV2, InceptionV3, DenseNet121, and EfficientNetB1. Each architecture possesses distinct characteristics in terms of depth, complexity, and computational efficiency, making them suitable candidates for comparison in our study.

One of the first CNN designs, AlexNet, rose to prominence after winning the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Multiple convolutional and fully connected layers make up its design, which facilitates efficient feature extraction and categorization.

VGG16, another influential architecture, is characterized by its uniform structure, consisting of 13 convolutional layers followed by three fully connected layers.VGG16 has demonstrated exceptional performance in various image classification tasks despite its simple architecture.

We also explored deeper architectures such as ResNet50 and ResNet101, which presented the cutting-edge idea of residual learning to solve the deep network vanishing gradient issue. These residual connections facilitate the training of significantly deeper models while maintaining computational efficiency.

GoogleNet, notable for its inception modules, emphasizes computational efficiency by introducing parallel convolutional pathways within each module. This design enables the network to capture features at multiple scales efficiently.

DenseNet201, DenseNet121, and EfficientNetB1 represent architectures that prioritize feature reuse and parameter efficiency. DenseNet architectures establish dense connections between layers, promoting feature propagation throughout the network, while EfficientNetB1 employs a compound scaling method to optimize both accuracy and computational cost.

InceptionResNetV2 and InceptionV3 architectures integrate inception modules with additional computational optimizations, resulting in networks capable of capturing intricate spatial hierarchies in images.

Through rigorous experimentation with these diverse architectures, our study aims to compare their effectiveness in automated lesion classification. We will assess performance metrics which include precision, accuracy, F1 score and recall across the different models, shedding light on their relative strengths and limitations. By elucidating the comparative performance and practical considerations associated with each model, our research contributes to a deeper understanding of transfer learning in dermatological applications. Additionally, the project differents these models on computational efficiency, and trade-off of these models.

2. Related Works and Trends in Melanoma Detection

In recent years, researchers have intensified efforts towards devising techniques for earlier detection of Melanoma (Me) and other skin lesions (SL). The focus has primarily been on leveraging neural networks (NNs), with notable trends observed in system design for detection, segmentation, and classification tasks. A significant proportion of recent literature emphasizes NN-based approaches, while classic classifiers such as KNN and SVM continue to be explored.

Notably, studies have highlighted several emerging trends in NN utilization for Me detection, including the adaptation of single CNN architectures tailored specifically for Me, the integration of multiple CNNs for enhanced performance, and the combination of CNNs with other classifiers.

Furthermore, the adoption of transfer learning (TL) techniques has been prevalent, with researchers often tweaking and fine-tuning the CNN models on diverse datasets. Additionally, there has been a surge in the development and modification of CNN architectures, ranging from traditional models like AlexNet, VGG, and GoogLeNet to designs such as Xception, and DenseNet. This trend reflects a transition towards increasingly complex and deep networks, often incorporating residual connections to enhance performance metrics.

Overall, the literature underscores the diverse approaches employed in designing Me detection systems, ranging from single CNN models to hybrid architectures combining multiple CNNs with complementary classifiers or techniques.

3. Proposed Work

By this work, we envision and propose to utilize a range of prominent CNN architectures, including AlexNet and VGG-16, alongside ResNet50, ResNet101, GoogleNet, VGG19, DenseNet201, InceptionResNetV2, InceptionV3, DenseNet121, and EfficientNetB1, for the detection of benign and malignant melanoma lesions using dermoscopic images. These models have demonstrated exceptional performance in image classification tasks and offer promising potential for automated melanoma diagnosis. We aim to leverage the unique architectural characteristics of each model and compare their performance in melanoma detection.

3.1. Using Alexnet

For our utilization of AlexNet, we'll tap into its five convolutional layers and the fully connected layers.

This architecture, pioneered by Krizhevsky et al. in 2012, introduces innovative elements like rectified

linear units (ReLU) for activation and dropout regularization to ward off overfitting. By harnessing the hierarchical features learned within AlexNet, our goal is to evaluate how well it discerns between benign and malignant melanoma lesions. AlexNet consists of several layers designed to extract important and minute features from input images:

1. Convolutional Layers:

AlexNet begins with 5 convolutional layers. They are responsible for detecting various features like edges,

textures, and patterns within the input images. These convolutional layers use kernels or filters to iterate across the image, extracting relevant features using operations such as edge detection and gradient computation.

$$z = \sum_{m=0}^{F-1} \sum_{n=0}^{F-1} \sum_{c=0}^{C_{in}-1} W_{m,n,c,k}^{(l)} \cdot x_{i+m,j+n,c}^{(l-1)} + b_k^{(l)}$$

2. ReLU Activation:

ReLU introduces non-linearity into the network by replacing negative pixel values with zero, thereby allowing the network to learn complex relationships between features and improve its ability to capture nonlinear patterns in the data

3. Max Pooling Layers:

AlexNet incorporates max-pooling layers to downsample the maps, reducing spatial dimensions while maintaining important features. Max pooling involves selecting the biggest value between a fixed-size window, effectively summarizing most relevant information within each region of the feature map.

$$a_{i,j,k}^{(l)} = \max_{m,n} \left\{ a_{m,n,k}^{(l-1)} \right\}$$

4. Fully Connected Layers:

AlexNet includes three fully connected layers at the end, known as dense layers. The layers integrate the spatial information extracted by the convolutional layers and aggregated it to make high-level decisions about the input image. These layers connect the neurons in one layer to every neuron in the other layer,

$$z_k^{(l)} = \sum_{i=1}^{N_{in}} W_{i,k}^{(l)} \cdot a_i^{(l-1)} + b_k^{(l)}$$

5. Softmax Layer:

This tier converts the raw product of the former fully affiliated tier into probabilities for each class. From softmax, VGG-16 produces a probability distribution over the different classes.

$$Soft\max(zi) = \frac{e^z i}{\sum_{j=1}^{K} e^z j}$$

Each layer in AlexNet plays an important role in the network's ability to learn hierarchical representations of visual patterns and make precise and accurate predictions.. By leveraging these layers, we aim to assess AlexNet's effectiveness in discriminating between benign and malignant melanoma lesions, ultimately contributing to the advancement of automated melanoma diagnosis systems.

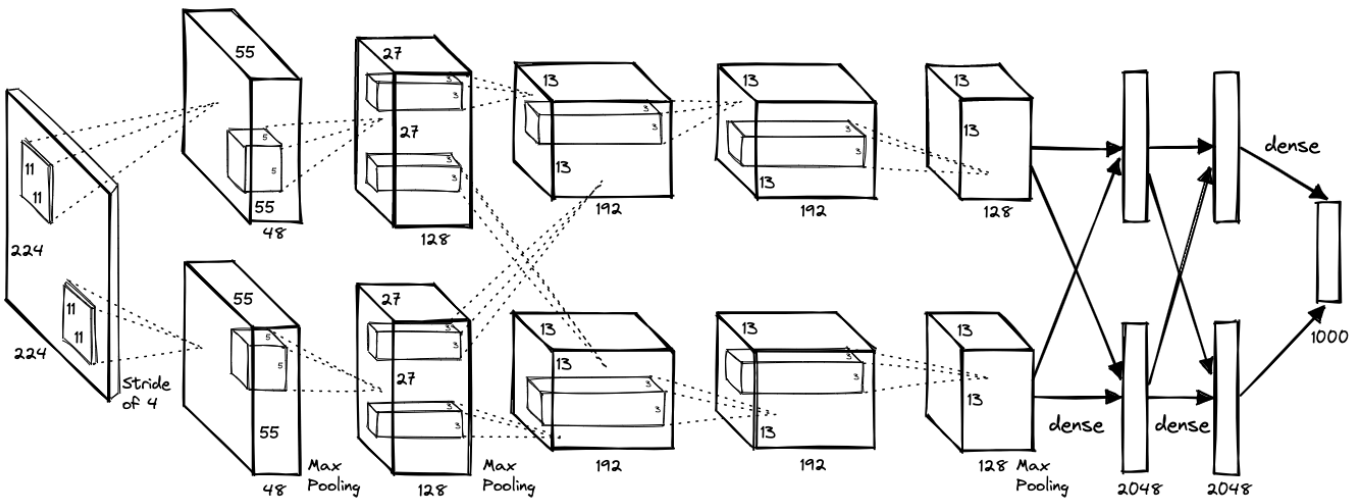

Figure 1. AlexNet Architecture

3.2. Using VGG-16

VGG-16 is another influential convolutional neural network (CNN) architecture that has made significant contributions to the field of image classification. Let's explore its architecture in more detail:

1. Convolutional Layers:

VGG-16 comprises 13 convolutional layers, arranged in a sequential manner. These layers consist of 3x3 convolutional filters on which ReLU is applied. The use of multiple convolutional layers allows VGG-16 to capture increasingly complex features in the input images.

$$X_{i+1} = \sigma(Wi * Xi + bi)$$

2. Max Pooling Layers:

By lowering the spatial dimensions of the feature maps, max pooling contributes to the network's increased computational efficiency while maintaining key features. VGG-16 employs max pooling with a 2x2 window with 2 strides. resulting in downsampling.

.

$$A_{i,j,k}^{[l+1]} = \max_{(p,q)\in R} A_{(i\times s+p),(j\times s+q),k}^{[l]}$$

3. Fully Connected Layers:

VGG-16 includes three fully connected layers. These dense layers integrate the spatial information given by the previous layer and aggregate them to make high-level decisions about the input image. The fully connected layers enable VGG-16 to capture complex relationships between features and perform fine-grained classification.

$$X_{i+1} = \sigma(Wi \cdot Xi + bi)$$

4. Softmax Layer:

This tier converts the raw product of the former fully affiliated tier into probabilities for each class. From softmax, VGG-16 produces a probability distribution over the different classes.

$$Soft\max(zi) = \frac{e^z i}{\sum_{j=1}^{K} e^z j}$$

5. Uniform Architecture:

One of the distinctive features of VGG-16 is its uniform architecture, where the convolutional layers consist of multiple 3x3 filters stacked on top of each other. This design principle simplifies the architecture and makes it easier to understand and implement. Despite its simplicity, VGG-16 has often demonstrated exceptional performance in image classification.

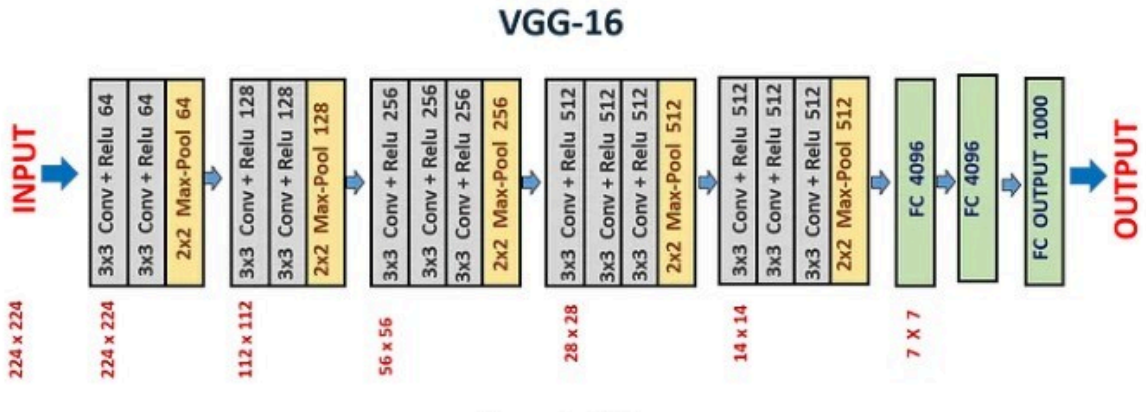

Figure 2. VGG16 Architecture

3.3 Using ResNet50 and ResNet101

ResNet50 and ResNet101 represent pivotal milestones in the evolution of convolutional neural network (CNN) architectures, particularly renowned for their innovative use of residual connections. In this section, we explore the architectural intricacies of ResNet50 and ResNet101.

1. Convolutional Layers:

ResNet50 and ResNet101 architecture uses residual blocks. These blocks include multiple convolutional layers. They have skip connections that avoid certain layers and the vanishing gradient problem. Residual blocks allow you to learn more complex elements in the input images.

$$Conv(x) = ReLU(W * x + b)$$

2. Max Pooling Layers:

Both ResNet50 and ResNet101 architectures employ max-pooling layers with a similar configuration to VGG-16, aiding in spatial dimension reduction while preserving essential features.

$$A_{i,j,k}^{[l+1]} = \max_{(p,q)\in R} A_{(i\times s+p),(j\times s+q),k}^{[l]}$$

3. Fully Connected Layers:

ResNet50 and ResNet101 architectures include fully connected layers at the end. The fully connected layers enable VGG-16 to capture complex relationships between features and perform fine-grained classification.

$$FC(x) = ReLU(Wx + b)$$

4. Softmax Layer:

A softmax layer is typically appended to the final fully connected layer in ResNet50 and ResNet101 architectures. This layer computes class probabilities, allowing the network to make predictions about the input image.

$$Soft\max(zi) = \frac{e^z i}{\sum_{j=1}^{K} e^z j}$$

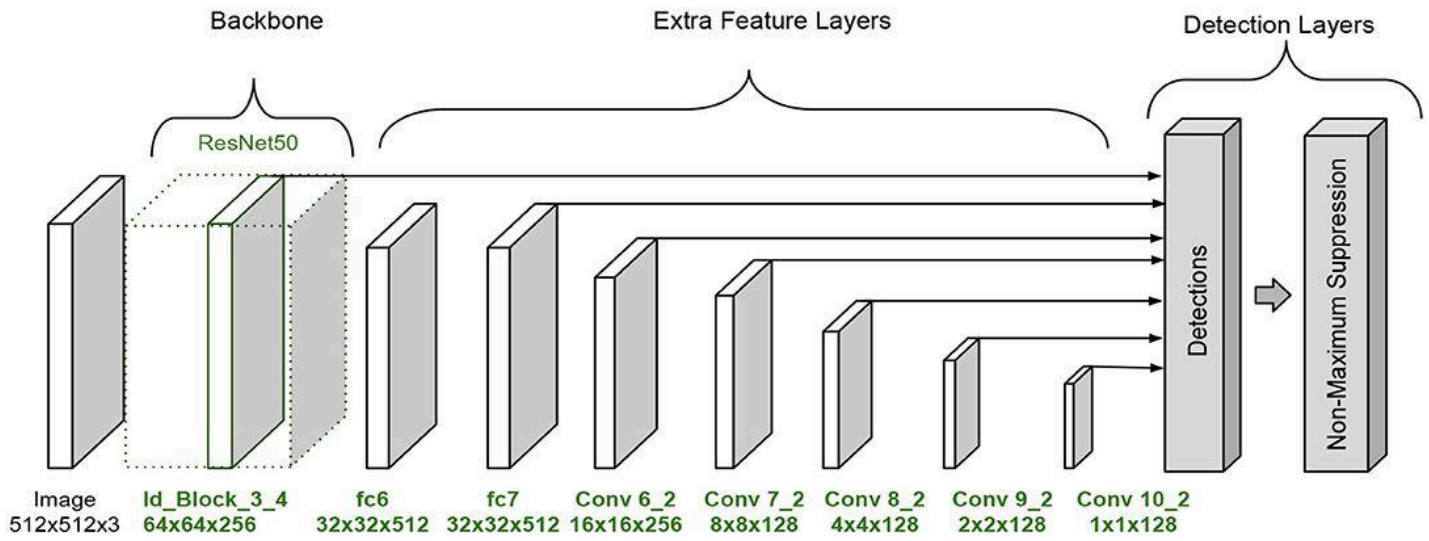

Figure 3. ResNet50 Architecture

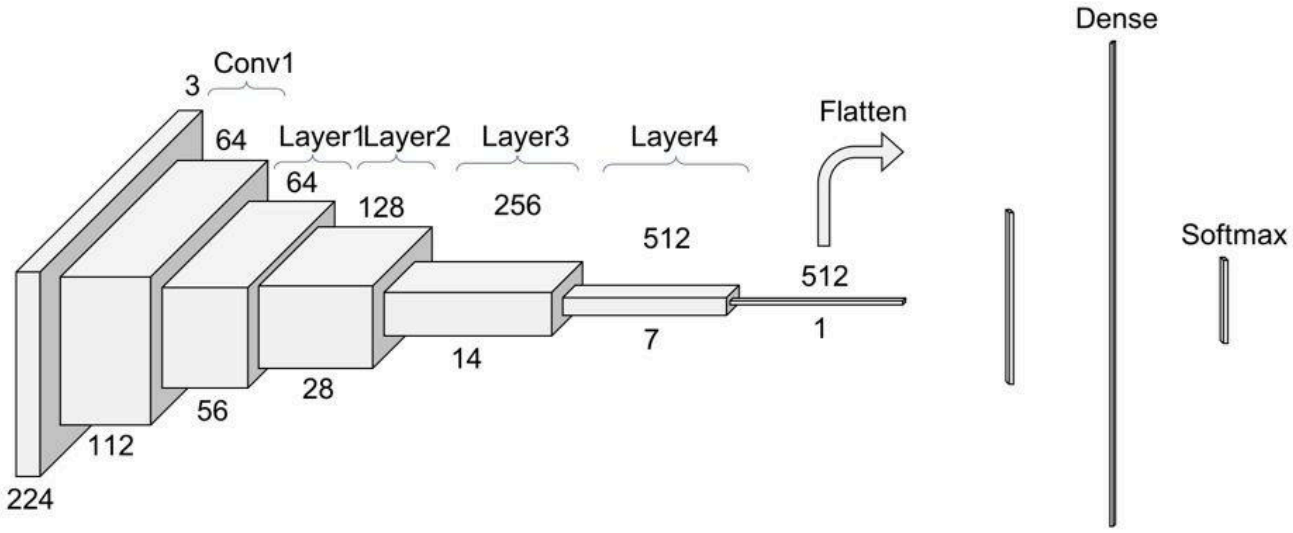

Figure 4. ResNet101 Architecture

3.4. Using VGG19

VGG19 is a CNN with layers from VGG16 with additional layers. VGG19 is renowned for its simplicity and effectiveness in image classification. The architecture and layers of VGG19 includes:

1. Convolutional Layers:

VGG19 comprises 16 convolutional layers, organized into blocks. They employ 3x3 filters with a stride of 1 and 0 padding, enabling the network to extract intricate features from input images.

$$X_{i+1} = \sigma(Wi * Xi + bi)$$

2. Activation Functions:

Following each convolutional layer in VGG19, a rectified linear unit (ReLU) activation function is applied. ReLU introduces non-linearity into the network by replacing negative pixel values with zero.

3. Max Pooling Layers:

Interspersed between the convolutional layers are five max pooling layers. By lowering the spatial dimensions of the feature maps, max pooling contributes to the network's increased computational efficiency while maintaining key features. VGG-16 employs max pooling with a 2x2 window with 2 strides. resulting in downsampling.

$$A_{i,j,k}^{[l+1]} = \max_{(p,q)\in R} A_{(i\times s+p),(j\times s+q),k}^{[l]}$$

4. Fully Connected Layers:

In the end, there are three fully connected layers, also known as dense layers which integrate the spatial information extracted by the convolutional layers. In a fully connected layer, each neuron in one layer is connected to each neuron in the following layer, enabling complex feature combinations and classification.

$$X_{i+1} = \sigma(Wi \cdot Xi + bi)$$

5. Softmax Layer:

The final layer of VGG19 is a softmax layer. This tier converts the raw product of the former fully affiliated tier into probabilities for each class. From softmax, VGG-16 produces a probability distribution over the different classes.

$$Soft\max(zi) = \frac{e^z i}{\sum_{j=1}^{K} e^z j}$$

In summary, VGG19 is characterized by its deep architecture, comprising multiple convolutional layers followed by fully connected layers and a softmax output layer. VGG19 has demonstrated exceptional performance in image classification tasks, making it a popular choice for various computer vision applications.

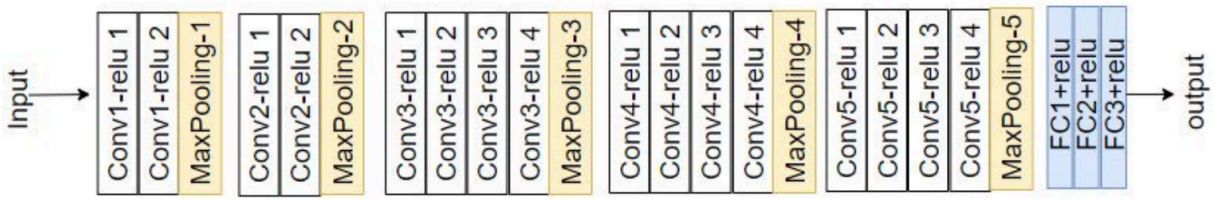

Figure 5. VGG19 Architecture

3.5. Using Inception Resnet V2

Inception ResNet V2 is a convolutional neural network architecture that combines the advantages of both Resnet and Inception Architecture. Developed by Google researchers in 2016, Inception-ResNet-v2 aims to achieve improved accuracy and efficiency in image classification tasks. Let's explore the architecture and layers of Inception-ResNet-v2:

1. Stem Block:

Inception-ResNet-v2 begins with a stem block, which serves as the initial feature extraction module. The stem block contains a number of convolutional layers and activation functions such as ReLU and batch normalization. This module extracts low-level features from the input image and prepares it for further processing.

$$Output = Conv_{7x7stride2}(Input) \rightarrow MaxPool_{3x3stride2}(Output) \rightarrow Conv_{3x3reduce}(Output)$$

2. Inception Blocks:

The core of Inception-ResNet-v2 architecture comprises multiple inception blocks, which are responsible for extracting hierarchical features from the input data. Each inception block consists of parallel pathways, known as inception modules, that perform convolutions with different kernel sizes to capture features at various spatial scales. Additionally, these blocks incorporate dimensionality reduction techniques, such as 1x1 convolutions, to reduce computational complexity while preserving valuable information.

$$Output = Concat([Conv_{1x1}, Conv_{3x3reduce}, Conv_{3x3}, Conv_{5x5reduce}, Conv_{5x5}, MaxPool_{3x3stride1}])$$

3. Residual Connections:

One of the key innovations of Inception ResNet V2 is the integration of connections that are residual within the inception blocks. Residual connections, inspired by the ResNet architecture, enable the network to mitigate the vanishing gradient problem and make way for the training of deeper networks. By integrating skip connections that bypass certain layers, residual connections allow for the direct flow of gradients during backpropagation, promoting more effective feature learning and model convergence.

$$Output = Input + BlockOutput$$

4. Reduction Blocks:

In addition to inception blocks, Inception-ResNet-v2 incorporates reduction blocks at specific intervals within the network. These reduction blocks serve to downsize the spatial dimensions of the maps while incrementing the number of channels. This downsampling aids in reducing the computational burden and improving the efficiency of the network.

5. Auxiliary Classifiers:

Inception-ResNet-v2 may also include auxiliary classifiers, which are additional branches connected to intermediate layers of the network. These auxiliary classifiers aid in training by providing additional supervision signals during the training process. By incorporating auxiliary classifiers, Inception-ResNet-v2 encourages the propagation of gradients through the network, facilitating more stable and efficient training.

6. Soft-max layer and Global Average Pooling

At the end, Inception-ResNet-v2 includes a global average pooling layer followed by a softmax layer. global average pooling layer accumulates feature maps across spatial dimensions, reducing them to a one-dimensional vector. Subsequently, the softmax layer changes the vector into a probability distribution of the different classes, authorizing the network to make accurate predictions.

$$Output = AveragePool(Input)$$

$$Soft\max(zi) = \frac{e^z i}{\sum_{j=1}^{K} e^z j}$$

In summary, Inception-ResNet-v2 combines the strengths of the Inception and ResNet architectures, incorporating parallel convolutional pathways, residual connections, and reduction blocks to achieve precise and accurate performance. Inception-ResNet-v2 has demonstrated remarkable accuracy and efficiency.

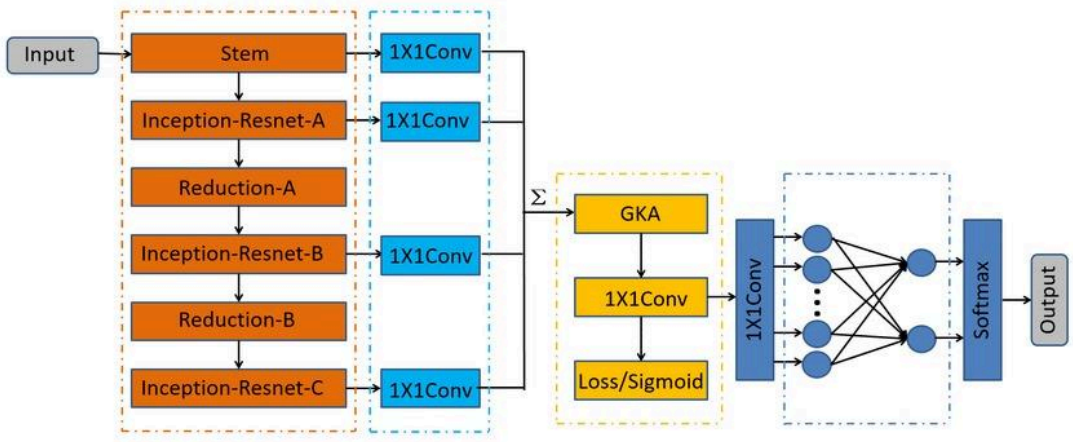

Figure 6. Inception-ResNet-V2 Architecture

3.6. Using Inception-V3

Inception-v3 is a convolutional neural network (CNN) architecture developed by Google researchers. Inception-v3 is renowned for its efficiency and accuracy in image classification. Let's explore the architecture and layers of Inception-v3:

1. Stem Block:
Inception-v3 begins with a stem block, which serves as the initial feature extraction module. The stem block usually includes multiple layers of convolution, batch normalization, and activation functions like ReLU. The low-level feature extraction module takes the input image and pre-processes it.

$$Output = Conv_{7x7stride2}(Input) \rightarrow MaxPool_{3x3stride2}(Output) \rightarrow Conv_{3x3reduce}(Output)$$

2. Inception Blocks:
The core of Inception-v3 architecture comprises multiple inception blocks, which are responsible for extracting hierarchical features from the input data. Each inception block consists of parallel pathways, known as inception modules, that perform convolutions with different kernel sizes to capture features at various spatial scales. Additionally, these blocks incorporate dimensionality reduction techniques, to reduce complexity while preserving valuable information.

$$Output = Concat([Branch1x1, Branch3x3, Branch5x5, BranchPool])$$

3. Inception Modules:
Within each inception block, several inception modules are stacked. Inception modules are composed of parallel layers with different sizes for filtering, which helps the network to capture a wide variety. Additionally, the modules incorporate pooling operations and concatenation layers to aggregate information from multiple pathways.

$$Output = InceptionBlock(Input)$$

4. Reduction Blocks:
At specific intervals within the network, Inception-v3 includes reduction blocks. These blocks help downsample the dimensions of the feature maps while increasing the number of channels. This eventually reduces the computational burden and improves efficiency.

$$Output = Concat([Conv_{3x3stride2}, MaxPool_{3x3stride2}])$$

5. Auxiliary Classifiers:
This model may also incorporate auxiliary classifiers, which are additional branches connected to intermediate layers of the network. These auxiliary classifiers provide supervision signals for training. This helps Inception-v3 encourage the propagation of gradients through the network, leading to more stability.

$$Output = Soft\max(FC(Pool(Conv(Input))))$$

6. Global Average Pooling and Softmax Layer:
Towards the end of the network, Inception-v3 typically includes a global average pooling layer followed by a softmax layer.

$$Output = AveragePool(Input)$$

$$Soft\max(zi) = \frac{e^z i}{\sum_{j=1}^{K} e^z j}$$

In summary, Inception-v3 is characterized by its efficient and scalable architecture, leveraging parallel convolutional pathways and dimensionality reduction techniques to achieve state-of-the-art performance in image classification tasks..

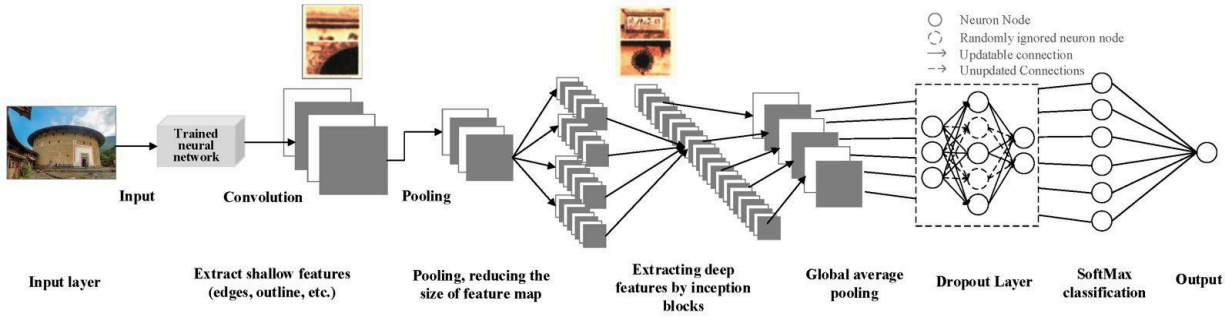

Figure 7. Inception-V3 Architecture

3.7. Using DenseNet-121

Huang et al. presented DenseNet-121, a convolutional neural network (CNN) architecture, in 2017. DenseNet is known for its densely connected architecture, in which all layers are feed-forward connected to the other layers. This connectivity pattern facilitates feature reuse and encourages the propagation of gradients throughout the network.

Let's explore the architecture and layers of DenseNet-121:

1. Dense Blocks:
The core of DenseNet-121 architecture comprises of multiple dense blocks. Within a dense block, each layer receives feature maps from all preceding layers as input and passes its own feature maps to all subsequent layers as output.

2. Convolutional Layers:
Within each dense block, densely connected convolutional layers perform feature extraction on the input data. These convolutional layers typically employ 3x3 filters with a stride of 1 and zero padding. Like this, the model captures hierarchical features of raising complexity.

$$X_{i+1} = \sigma(Wi * Xi + bi)$$

3. Batch Normalization and ReLU Activation:
Within the dense blocks, rectified linear unit (ReLU) activation functions and batch normalization are applied after every convolutional layer. By normalizing each layer's activations, batch normalization lowers internal covariate shift and speeds up convergence during training. ReLU activation gives the network non-linearity, which enables it to recognize intricate relationships in the data.

$$\mu_c = \frac{1}{N}\sum_{i=1}^{N} x_{i,c}$$

$$\sigma_c^2 = \frac{1}{N}\sum_{i=1}^{N}(x_{i,c} - \mu_c)^2$$

$$x_{i,c} = \frac{x_{i,c} - \mu_c}{\sqrt{\sigma_c^2 + \in}}$$

$$y_{i,c} = \gamma_c x_{i,c} + \beta_c$$

4. Transition Layers:
To control the growth of feature maps and reduce computational complexity, DenseNet-121 includes transition layers between dense blocks. These transition layers typically consist of batch normalization, 1x1 convolutional layers, and average pooling operations.

5. Global Average Pooling and Softmax Layer:
Towards the end of the network, the model includes global average pooling and a softmax layer. Global average pooling aggregates feature maps across spatial dimensions, reducing them to a one-dimensional vector. The softmax layer converts the vector into a probability distribution, helping to make predictions.

$$z = GlobalAveragePooling\,(x)$$

$$Soft\max(zi) = \frac{e^z i}{\sum_{j=1}^{K} e^z j}$$

In summary, DenseNet-121 is characterized by its densely connected architecture, leveraging dense blocks to facilitate feature reuse and enhance information flow within the network. By exploiting the benefits of dense connectivity and hierarchical feature learning, DenseNet-121 achieves good implementation in image classification tasks..

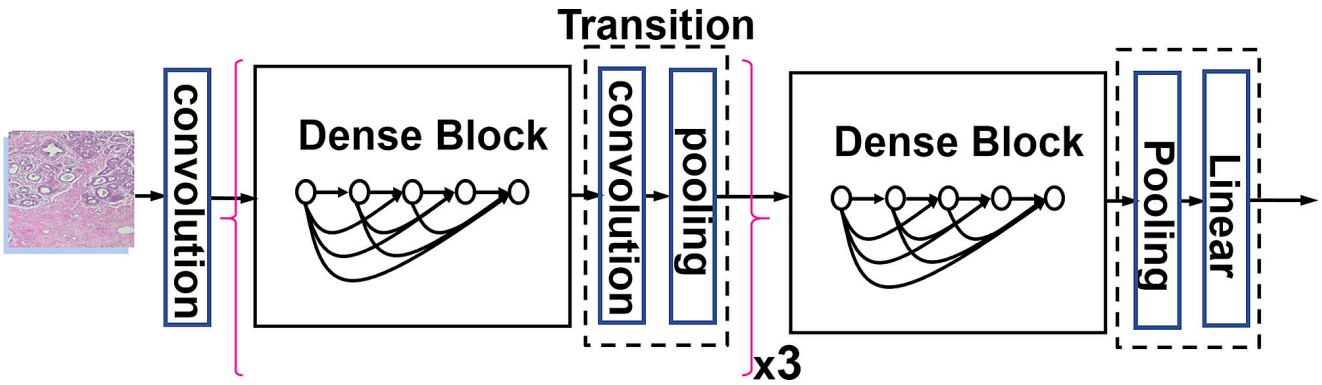

Figure 8. DenseNet-121 architecture

3.8. Using EfficientNetB1

EfficientNetB1 is a convolutional neural network (CNN) architecture introduced by Tan and Le in 2019. It is designed for exceptional performance in image classification tasks. EfficientNetB1 represents the baseline model. In the EfficientNet series, there are also many larger versions (e.g., EfficientNetB2, EfficientNetB3, etc.). Let's explore the architecture and layers of EfficientNetB1:

1. Stem Convolutional Layer:
EfficientNetB1 begins with a stem convolutional layer, which serves as the initial feature extraction module. This layer has many convolutional layers, batch normalization, and activation functions, mostly Swish. The stem layer extracts low-level features from the input image and prepares it for further processing.

$$z = \sum_{m=0}^{F-1} \sum_{n=0}^{F-1} \sum_{c=0}^{C_{in}-1} W_{m,n,c,k}^{(l)} \cdot x_{i+m,j+n,c}^{(l-1)} + b_k^{(l)}$$

$$a_{i,j,k}^{(l)} = Swish\left(z_{i,j,k}^{(l)}\right)$$

2. Efficient Blocks:
The core of EfficientNetB1 architecture comprises multiple efficient blocks, also known as MBConv blocks. These blocks are based on mobile inverted bottleneck convolution (MBConv) operations, which optimize both accuracy and efficiency. Each MBConv block consists of depthwise separable convolutions, followed by expansion and squeeze operations, which enhance feature representation while reducing the computational cost.

3. Depthwise Separable Convolutions:
Within each efficient block, depthwise separable convolutions are employed to perform spatial convolutions with significantly fewer parameters compared to traditional convolutional layers. This separable convolutional strategy

reduces both computational complexity and memory footprint, making EfficientNetB1 highly efficient.

$$z_{i,j,k}^{(l)} = \sum_{m=0}^{F-1} \sum_{n=0}^{F-1} W_{m,n,k}^{(l)} \cdot x_{i+m,j+n,k}^{(l-1)} + b_k^{(l)}$$

4. Squeeze-and-Excitation (SE) Blocks:
EfficientNetB1 adds squeeze-and-excitation (SE) blocks at predetermined network intervals in addition to efficient blocks. The goal of SE blocks is to dynamically recalibrate feature maps according to their significance and capture channel-wise dependencies.

5. Global Average Pooling and Softmax Layer:
Towards the end of the network, EfficientNetB1 typically includes global average pooling followed by a softmax layer.
.

Global average pooling aggregates feature maps across spatial dimensions, reducing them to a one-dimensional vector. Subsequently, the softmax layer converts the vector into a probability distribution over the different classes, enabling the network to make predictions.

$$z = GlobalAveragePooling\,(x)$$

$$Soft\max(zi) = \frac{e^z i}{\sum_{j=1}^{K} e^z j}$$

In summary, EfficientNetB1 is characterized by its efficient architecture, leveraging efficient blocks and depthwise separable convolutions. By optimizing both accuracy and efficiency, EfficientNetB1 represents a significant advancement in CNN architectures.

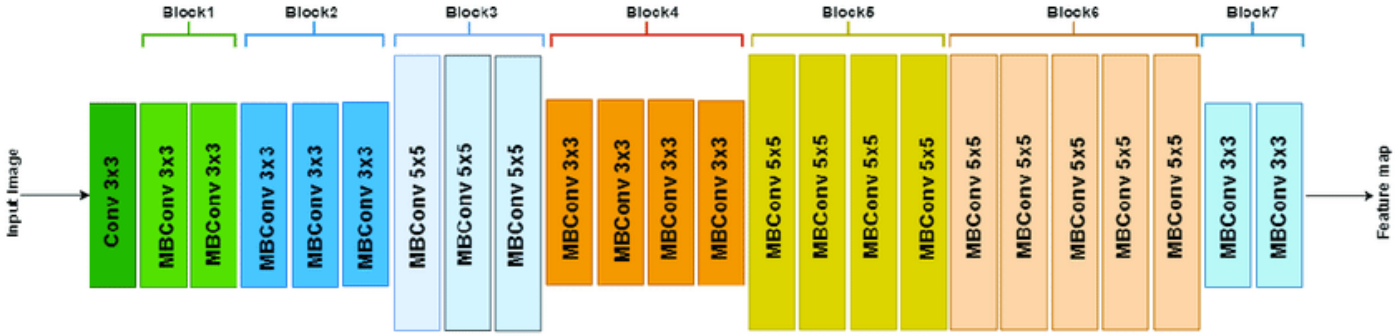

Figure 9. EfficientNetB1 Architecture

4. Dataset

Our dataset serves as a pivotal resource in the domain of dermatology and computer-aided diagnostics, meticulously crafted to facilitate advancements in melanoma detection. Comprising 13,900 images, each pixel within this dataset holds the promise of revolutionizing early detection practices.

The dataset is organized into two primary categories: "benign" and "malignant," reflecting the distinction between non-cancerous and cancerous lesions, respectively. Within each category, subfolders for training and testing are provided, facilitating seamless utilization of the dataset for model development and evaluation.

Figure 10.1. Benign Lesion(1)

Figure 10.2. Benign Lesion(2)

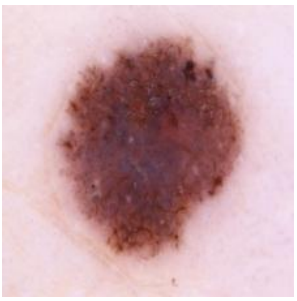

Figure 10.3. Malignant Lesion(1)

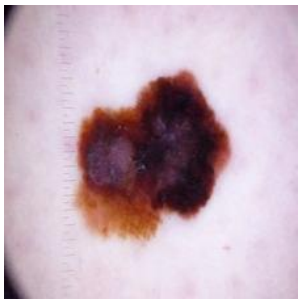

Figure 10.4. Malignant Lesion(2)

Total no. of images for Melanoma Detection	Number of images in the Trained Dataset		Number of images in the Tested Dataset	
	Benign	Malignant	Benign	Malignant
13,900	5950	5950	1000	1000

Figure 11. Image Set Distribution

5. Results

Nine distinct models were evaluated in our comparative study of convolutional neural network (CNN) architectures for melanoma detection: AlexNet, ResNet-50, VGG16, VGG19, Inception-ResNet-v2, ResNet-101, Inception-v3, DenseNet-121, and EfficientNetB1. On a standardized dataset, each model was trained and tested, and its accuracy was noted for performance assessment.

The results of our experiments are summarized as follows:

- AlexNet: Achieved an accuracy of 80.95%, demonstrating superior performance in melanoma detection tasks.

- ResNet-50: Attained an accuracy of 78.8%, showcasing competitive performance with its residual connections.

- VGG16: Recorded an accuracy of 50%, exhibiting moderate performance compared to other models.

- VGG19: Similarly, VGG19 achieved an accuracy of 50%, demonstrating performance similar to VGG16.

- Inception-ResNet-v2: Demonstrated an accuracy of 76.75%, showcasing robust performance with its combination of inception modules and residual connections.

- ResNet-101: Achieved an accuracy of 90.8%, exhibiting competitive performance with its deeper architecture compared to ResNet-50.

- Inception-v3: Registered an impressive accuracy of 84.55%, showcasing outstanding performance with its inception modules and efficient architecture.

- DenseNet-121: Attained an accuracy of 79%, demonstrating commendable performance with its densely connected architecture.

- EfficientNetB1: Registered an accuracy of 51%, exhibiting lower performance compared to other models.

Overall, ResNet-101 emerged as the top-performing model in melanoma detection with an accuracy of 90.8%. ResNet-101 showcased outstanding performance, leveraging its deeper architecture to achieve superior accuracy in classifying melanoma lesions.

These findings underscore the effectiveness of ResNet-101 as a powerful model for enhancing the early detection of melanoma, thereby contributing to improved patient outcomes in dermatological healthcare.

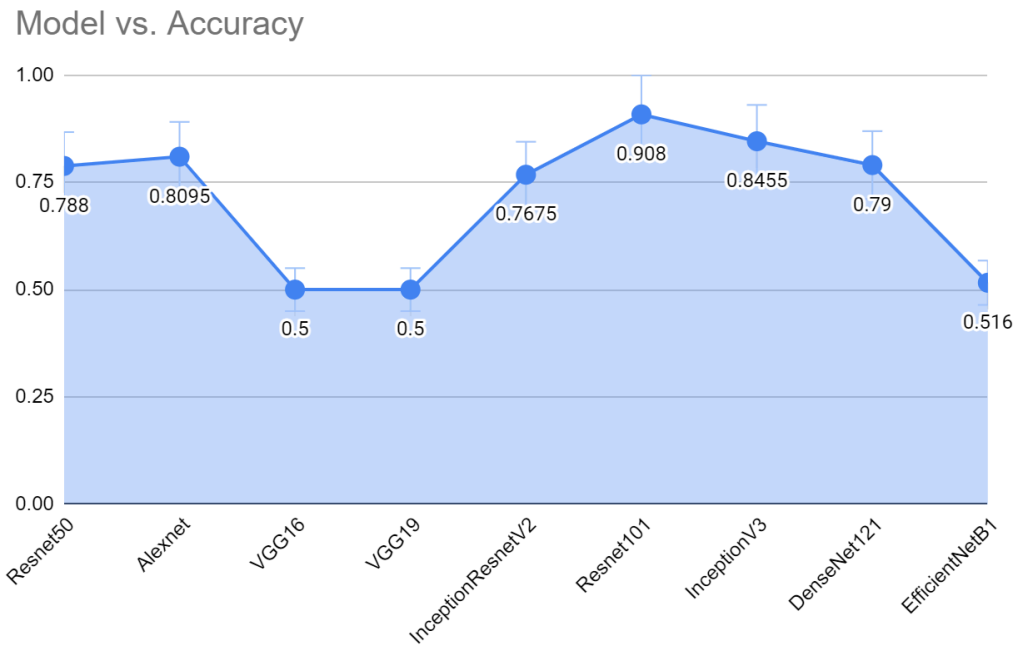

Figure 12. Study Of Accuracy In The Different Models

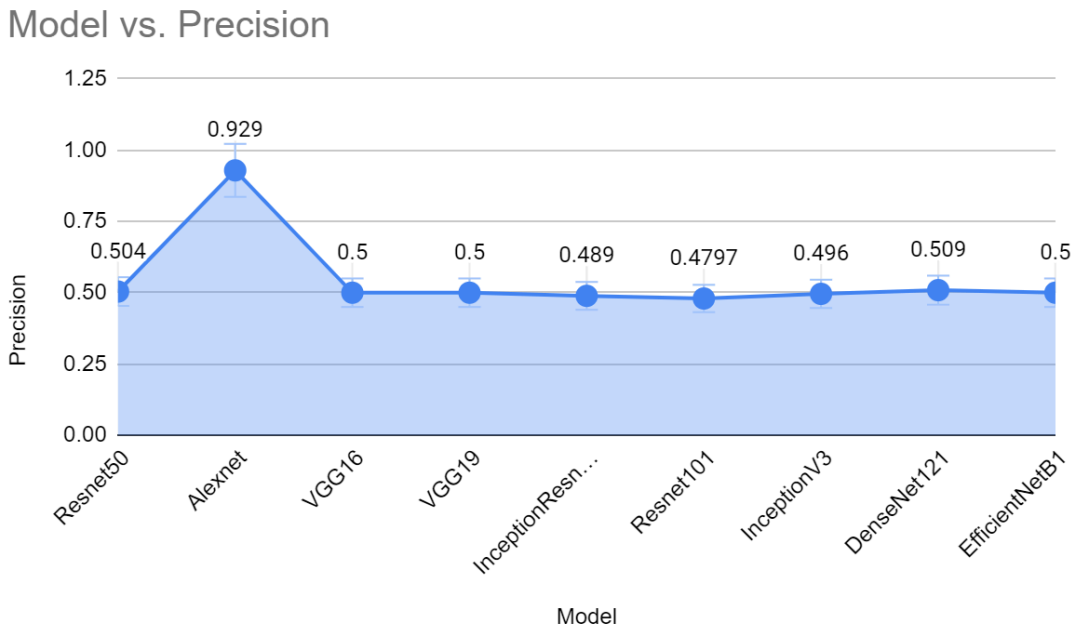

Figure 13. Study Of Precision In The Different Models

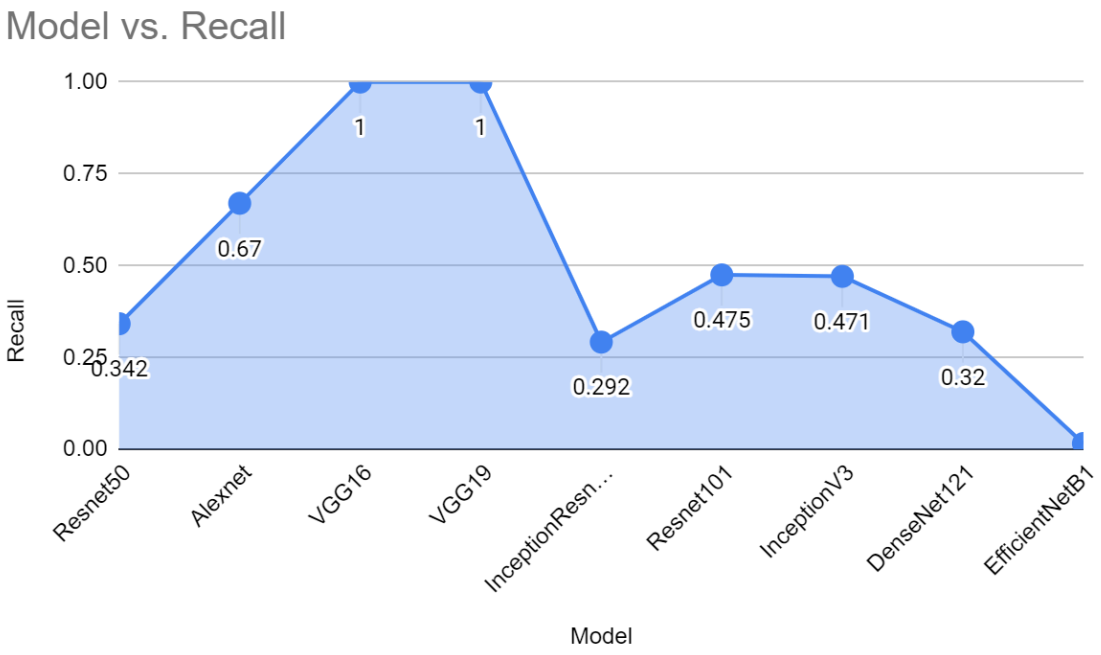

Figure 14. Study Of Recall In The Different Models

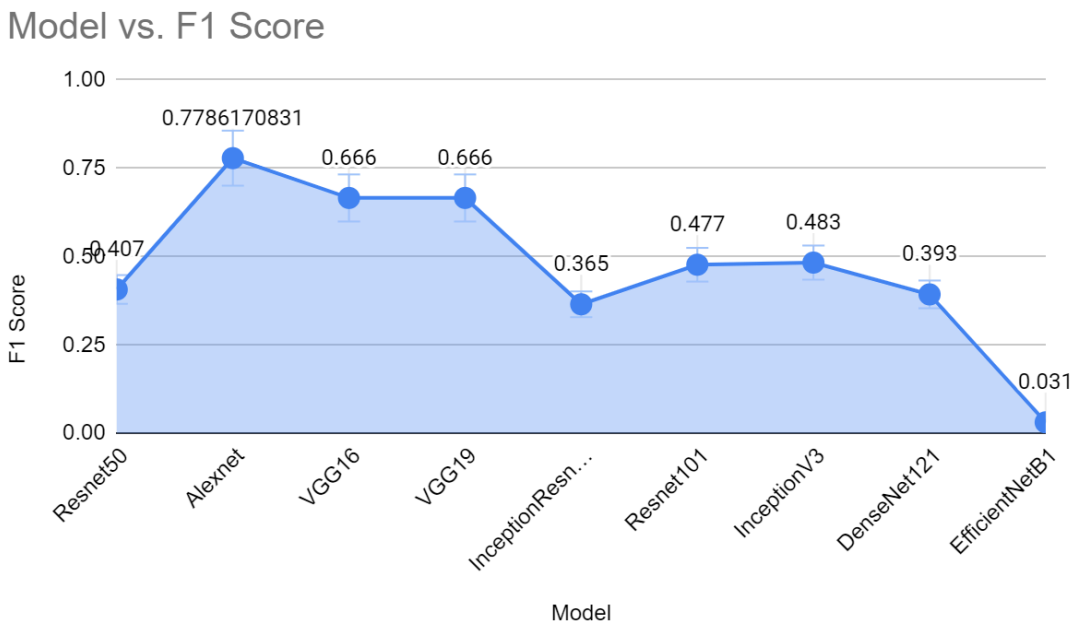

Figure 15. Study Of F1 Score In The Different Models

6. Discussion

The evaluation of various CNN models for the prediction of malignant and benign skin lesions reveals intriguing insights into their performance and suitability for melanoma detection. Among the models assessed, AlexNet emerges as a standout performer across multiple evaluation metrics. Its high accuracy of 0.8095, coupled with a remarkable precision score of 0.929, underscores its efficacy in correctly identifying malignant lesions while minimizing false positives. Moreover, AlexNet exhibits a commendable recall score of 0.67, denoting its capability to capture a significant proportion of malignant lesions within the dataset. The resulting F1 score of 0.778 further solidifies AlexNet's balanced performance, making it a compelling choice for melanoma classification tasks.

On the other hand, models such as VGG16 and VGG19 present a fascinating dichotomy. While they achieve a perfect recall score of 1, suggesting their capability to identify all malignant lesions, their precision scores of 0.5 raise concerns regarding potential false positives. This trade-off between recall and precision highlights the need for careful consideration of the clinical implications associated with misclassifications in melanoma diagnosis. Despite their perfect recall, VGG16 and VGG19 may exhibit a tendency to over-diagnose malignancy, potentially leading to unnecessary patient anxiety and medical interventions.

ResNet models, including ResNet50 and ResNet101, demonstrate balanced performance across precision, recall, and F1 score metrics. Their ability to achieve competitive accuracy rates, coupled with moderate to high precision and recall scores, positions them as robust candidates for melanoma classification. These models achieve a nuanced equilibrium between reducing false positives and false negatives., thereby offering reliable diagnostic outcomes suitable for clinical decision-making.

Conversely, EfficientNetB1 exhibits relatively poor performance compared to other models, with an accuracy of 0.516 and an F1 score of 0.031. These results underscore the importance of selecting appropriate model architectures and optimization strategies tailored to the complexities of melanoma detection tasks. Further exploration and refinement of EfficientNetB1 or exploration of alternative architectures may be warranted to enhance its performance in future iterations.

In conclusion, the evaluation of CNN models for melanoma classification reveals a spectrum of performance ranging from exemplary to suboptimal. While AlexNet demonstrates outstanding accuracy, precision, recall, and F1 score, VGG16 and VGG19 exhibit perfect recall at the expense of precision. ResNet models strike a balance between precision and recall, making them reliable choices for melanoma diagnosis. EfficientNetB1, while underperforming in this evaluation, presents an opportunity for refinement and optimization. Ultimately, the selection of an appropriate model architecture should consider a holistic assessment of performance metrics and clinical implications to ensure accurate and reliable melanoma detection.

7. Conclusion

Our investigation of the different convolutional neural network (CNN) architectures for melanoma detection provides crucial insights into the efficacy of deep learning models in this critical domain of healthcare. Evaluating models such as ResNet-101, Inception-v3, DenseNet-121, AlexNet, VGG16, ResNet-50, Inception-ResNet-v2, VGG19, and EfficientNetB1, we observed varying performance levels in terms of accuracy and computational efficiency.

We can say that the evaluation of convolutional neural network (CNN) models for melanoma classification underscores how minute the difference can be between model architecture, performance metrics, and clinical relevance. While models like AlexNet demonstrate impressive accuracy and precision, others like VGG16 and VGG19 exhibit trade-offs between recall and precision, necessitating careful consideration of their clinical implications. The balanced performance of ResNet models highlights their suitability for real-world applications, while the underperformance of EfficientNetB1 suggests areas for further optimization and refinement.

Moving forward, it is essential to prioritize model interpretability, robustness, and generalization capabilities to ensure the reliable and practical deployment of melanoma detection systems in clinical settings. Ensemble learning approaches and advancements in model interpretability techniques offer promising avenues for enhancing the performance and trustworthiness of CNN models. By addressing these challenges and leveraging the collective expertise of researchers and clinicians, By laying the groundwork for more precise, effective, and easily accessible melanoma diagnosis, we can improve patient outcomes and advance healthcare delivery.

8. Future Scope

In the realm of melanoma detection, the evaluation of CNN models offers a glimpse into current methodologies and their performance metrics. However, there exists a wide array of future research directions that can significantly advance the field. One avenue for exploration involves the integration of multi-modal dermoscopic images and patient demographics, to enrich the information. By merging diverse data sources, we can enhance detection and provide more comprehensive patient assessments. Architectural innovations like attention mechanisms and graph neural networks could enhance feature extraction and capture intricate patterns indicative of melanoma, thereby boosting diagnostic performance.

The utilization of transfer learning and domain adaptation techniques helps improve model generalization. Fine-tuning pre-trained models on diverse datasets can help researchers mitigate data scarcity issues and improve model adaptability to real-world scenarios. Enhancing model interpretability is a very critical aspect of future research. Advancements in interpretability techniques, such as attention maps and feature visualization methods, can improve transparency and trust in CNN-based melanoma detection systems, facilitating their seamless integration into clinical workflows.

Furthermore, rigorous clinical validation studies can be done to evaluate the performance and clinical utility of CNN models in real-world settings. Collaborative efforts between researchers, clinicians, and healthcare stakeholders are necessary to validate models, assess their impact on patient outcomes, and facilitate their adoption into routine clinical practice. User-friendly interfaces, cloud-based deployment options, and integration with electronic health record systems can facilitate the widespread adoption and utilization of melanoma detection technologies.

In summary, the future of melanoma detection is dependent on the ongoing investigation of novel approaches, cooperative alliances, and translational research initiatives targeted at boosting clinical judgment, increasing diagnostic precision, and eventually improving patient outcomes in the battle against melanoma.

9. References

[1] Chandran Kaushik Viknesh,1,* Palanisamy Nirmal Kumar,1 Ramasamy Seetharaman, “Detection and Classification of Melanoma Skin Cancer Using Image Processing Technique.

[2] Codella, N., Nguyen, Q. B., Pankanti, S., Gutman, D. A., Helba, B., Halpern, A., & Smith, J. R. (2017), Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development.

[3] Z. Ur, M. S. Zia, G. Reddy, M. Yaqub, and F. Jinchao, “Texture based localization of a brain tumor from MR-images by using a machine learning approach ,” *Med. Hypotheses, Elsevier*, vol. 141, pp. 1–12, 2020, doi: 10.1016/j.mehy.2020.109705.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks.

[5] Rajesh Godasu, David Zeng, Kruttika Sutrave (2020) Transfer Learning in Medical Image Classification: Challenges and Opportunities. AIS Electronic Library (AISeL)

[6] E. Yilmaz , M. Trocan. A modified version of GoogLeNet for melanoma diagnosis. Journal Of Information and Telecommunication 2021, VOL. 5, NO. 3, 395–405

[6] Sheldon Mascarenha,; Mukul Agarwal.A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)

[7] Sameh Abd El-Ghany and Mai. R. Ibraheem and Madallah Alruwaili and Mohammed Mahfouz Elmogy. Diagnosis of Various Skin Cancer Lesions Based on Fine-Tuned ResNet50 Deep Network. Computers, Materials & Continua 2021, 68(1), 117-135.

[8] N. Jethwa, H. Gabajiwala, A. Mishra, P. Joshi and P. Natu, "Comparative Analysis between InceptionResnetV2 and InceptionV3 for Attention based Image Captioning," 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2021, pp. 1-6, doi: 10.1109/GCAT52182.2021.9587514.

[9] Saumya R. Salian and Sudhir D. Sawarkar. Melanoma Skin Lesion Classification Using Improved Efficient B3. Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 01, March 2022.

[10] Preeti Gupta , Sachin Meshram. AlexNet and DenseNet-121-based Hybrid CNN Architecture for Skin Cancer Prediction from Dermoscopic Images .International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98

[11] Dakhli, R., Barhoumi, W. A skin lesion classification method based on expanding the surrounding lesion-shaped border for an end-to-end Inception-ResNet-v2 classifier. SIViP 17, 3525–3533 (2023).

[12] Mohammad Ali Kadampur, Sulaiman Al Riyaee, Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images, Informatics in Medicine Unlocked, Volume 18, 2020, 100282, ISSN 2352-9148.

[13] Zhang, L., Gao, H.J., Zhang, J. and Badami, B., 2020. Optimization of the convolutional neural networks for automatic detection of skin cancer. Open Medicine, 15(1), pp.27-37.

[14] Hosny, K.M., Kassem, M.A. and Fouad, M.M., 2020. Classification of skin lesions into seven classes using transfer learning with AlexNet. Journal of digital imaging, 33, pp.1325-1334.

[15] Wan, X., Zhang, X. and Liu, L., 2021. An improved VGG19 transfer learning strip steel surface defect recognition deep neural network based on few samples and imbalanced datasets. Applied Sciences, 11(6), p.2606.

[16] Ul Hoque, S., 2022. Performance comparison between VGG16 and VGG19 deep learning method with CNN for brain tumor detection. Int Res J Mod Eng Technol Sci, 4(7).

[17] Kallipolitis, A., Revelos, K. and Maglogiannis, I., 2021. Ensembling EfficientNets for the classification and interpretation of histopathology images. Algorithms, 14(10), p.278.

[18] Kaushik, P., Rathore, R., Kumar, A., Goshi, G. and Sharma, P., 2024, March. Identifying Melanoma Skin Disease Using Convolutional Neural Network DenseNet-121. In 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) (Vol. 2, pp. 1-4). IEEE.

[19] Miglani, V. and Bhatia, M.P.S., 2020, February. Skin lesion classification: A transfer learning approach using efficientnets. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 315-324). Singapore: Springer Singapore.

[20] Hridoy, R.H., Akter, F. and Rakshit, A., 2021, July. Computer vision based skin disorder recognition using EfficientNet: A transfer learning approach. In 2021 International conference on information technology (ICIT) (pp. 482-487). IEEE.

[21] Villa-Pulgarin, J.P., Ruales-Torres, A.A., Arias-Garzon, D., Bravo-Ortiz, M.A., Arteaga-Arteaga, H.B., Mora-Rubio, A., Alzate-Grisales, J.A., Mercado-Ruiz, E., Hassaballah, M., Orozco-Arias, S. and Cardona-Morales, O., 2022. Optimized Convolutional Neural Network Models for Skin Lesion Classification. Computers, Materials & Continua, 70(2).

GitHub Link

https://github.com/snnath/Melanoma-Detection/tree/master