Comparison of deep transfer learning models for cancer diagnosis

by

Nadia Islam Joya
18101105
Tasfia Haque Turna
18301280
Zinia Nawrin Sukhi
18301193
Tania Ferdousey Promy
18101678

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

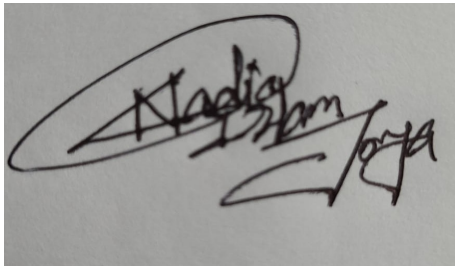Department of Computer Science and Engineering
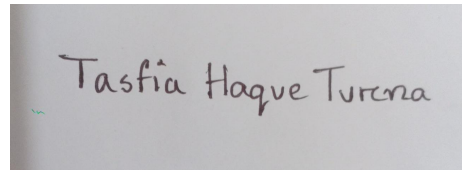Brac University
May 2022

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.
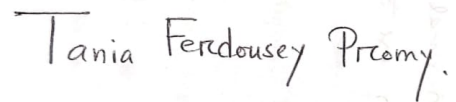
Student's Full Name & Signature:

_____
Nadia Islam Joya
18101105

_____
Tasfia Haque Turna
18301280

_____
Zinia Nawrin Sukhi
18301193

_____
Tania Ferdousey Promy
18101678

Approval

The thesis/project titled "Comparison of deep transfer learning models for cancer diagnosis " submitted by

1. Nadia Islam Joya (18101105)

2. Tasfia Haque Turna (18301280)

3. Zinia Nawrin Sukhi (18301193)

4. Tania Ferdousey Promy (18101678)

Of Spring, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May 19, 2022.

Examining Committee:

Primary Supervisor:
(Member)

<div style="text-align:center">

Dr.Jia Uddin,PhD
Assistant Professor (Research Track)
Technology Studies Department
Woosong University, Daejeon, South Korea

</div>

Co- Supervisor:
(Member)

<div style="text-align:center">

Faisal Bin Ashraf
Lecturer
Department of Computer Science and Engineering
Brac University

</div>

Program Coordinator:
(Member)

_____

Dr. Md. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

_____

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement

We hereby certify that the following conditions are met for the paper Comparison of deep transfer learning models for cancer diagnosis:
The work accurately and completely represents our own research and analysis. The findings are adequately contextualized in light of previous and ongoing research. All resources are correctly credited (correct citation) and The use of quote marks and correct referencing must be used to denote genuine duplication of content. All group members were directly and ensuring accountability in the extensive effort that led to the article, and they will be held accountable for its integrity

Abstract

Cancer is known to be one of the most lethal diseases among all the diseases in the world. It is clinically known as 'Malignant Neoplasm which is a vast group of diseases that encompasses unmonitored cell expansion. It can begin anywhere in the body such as the breast, skin, liver, lungs, brain, and so on. According to GLOBOCAN 2020, approximately 19.3 million new cases were found and around 10.0 million deaths have occurred for cancer globally. As reported by the National Institutes of Health (NIH), the projected growth of new cancer cases is forecast at 29.5 million and cancer-related deaths at 16.4 million through 2040. Breast, colorectal, endometrial, lung, oral, skin, and ovarian cancers are some of the most common malignancies that people develop. There are many medical procedures to identify the cancer cell such as mammography, MRI, CT scan which are common methods for cancer diagnosis. The methods used above have been found to be ineffective and necessitate the development of new and smarter cancer diagnostic technologies. Persuaded by the phenomena of deep learning in medical image classification tasks, the recommended initiative targets to analyze the performance of deep transfer learning for cancer cell classification. Transfer learning is used in visual categorization to solve cross-domain learning issues by transferring useful data from the source domain to the task domain. Cancer, also known as tumor, must be discovered early and accurately in order to determine what treatment alternatives are available. Even if each modality has its own set of problems, such as a convoluted medical history, incorrect diagnosis, and therapy, all of which are major causes of death. Artificial Intelligence-based medical diagnosis is a novel strategy in medicine that eliminates the need for pathologists to work with material in favor of pixels to diagnose illness (imaging in medical sector). Therefore, in our paper, we want to offer a narrative of four different deep transfer learning techniques Vgg16, InceptionV3, MobilenetV2 and Resnet50 to examine the accuracy, compare and discuss for the detection of breast, lung, and melanoma (skin) cancer.

Keywords: Convolutional Neural Network(CNN), Deep Transfer learning, Cancer Detection, Image Processing,

Dedication

This study effort is dedicated to all those who have died as a result of cancer, to every one of those people who have struggling for this fatal disease and fighting for their lives, to all the parents who have lost their child for cancer, to all the children who got separated from their parents by this disease, to all those young people who are cut off from the rest of the world. Lastly, we hope our effort will assist cancer detection research progress and improvement and a small contribution such as this will have a significant effect in future.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.

Secondly, we would like to extend our heartfelt gratitude to our respected to our supervisor Dr.Jia Uddin and co-supervisor Faisal Bin Ashraf for his kind support and advice in our work. they helped us whenever we needed help.

And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

API Application Programming Interface

CAD Computer Aided Diag

CNN Convolutional Neural Network

$DDSM$ Digital Database of Screening Mammography

$GLOBOCAN$ Global Cancer Observatory

$ILSVRC$ Industrial Visual Recognition

$LIDC - IDRI$ Lung Image Database Consortium Image Collection

$MIAS$ Mammographic Image Analysis Society

MRI Magnetic Resonance Imaging

NIH National Institues of Health

$OpenCV$ Open Source Computer Vision

PIL Python Imaging Library

$RESNET50$ Residual Networks

RGB Red Green Blue

$VGG16$ Visual Geometry Group from Oxford

Chapter 1

Introduction

1.1 Cancer - a life threatening illness

In the recent time, cancer is one of the most life-challenging diseases in the world. It is the biggest cause of death in the United States, affecting people of all ages and backgrounds. Early detection is the key to cancer treatment, but it's often not that easy. Cancer is a complicated disease, and there are a lot of ways it can show up in the body. Sometimes the symptoms don't appear until the cancer is quite advanced. On other times there are no symptoms at all. While certain cancers, such as breast cancer, are comparatively easy to identify, on the other hand cancers like lung, kidney or brain are very hard to detect. The majority of cancers are now discovered only after they have progressed beyond the organs in which they began. Nevertheless, because of recent breakthroughs in deep learning (an artificial intelligence approach that allows you to spot patterns in images, audio, or text, among other things), we can now construct the finest artificial intelligence techniques and algorithms for cancer-detection. Much more can be developed to improve cancer detection and quality of life for patients. This is also why cancer researchers are making quite remarkable progress.

1.2 Aims and Objectives

Cancer is recognized as one of the deadliest diseases in the world. It can spread very quickly to the whole body and takes decades in development thus can be preventable if diagnosed at a primary stage. According to GLOBOCAN 2020 [34] , with an expected 2.3 million new cancer cases (11.7 percent), female breast cancer has overtaken lung cancer as the most frequently diagnosed malignancy, followed by lung (11.4 percent), colorectal (10.0 percent), prostate (7.3 percent), and stomach (5.6 percent) cancers. With an expected 1.8 million deaths (18 percent), lung cancer remained the leading cause of cancer death, followed by colorectal (9.4%), liver (8.3%), stomach (7.7%), and female breast (6.9%) cancers. The worldwide cancer burden is forecasted to reach 28.4 million cases in 2040, up 47% from 2020, with a greater increase in transitional (64%) than transitioned (32%) nations owing to demographic shifts, however, this might be worsened further by increasing risk. The usage of automated diagnosis tools, which may assist doctors and even non - professionals in determining certain cancer, has ushered in significant advancement in the medical industry [31]. Hence, early detection is very much needed for this

disease to stop from spreading and turning it into the form of cancer.In this case, an automated diagnosis system can be a great help to detect this disease at the primary stage.

Different forms of cancer detection and classification with deep learning have opened up a new area of research for early cancer diagnosis, demonstrating the possibility to eliminate manual system limitations. Our main purpose is to present a brief analysis, comparisons on four deep transfer learning models (VGG16, InceptionV3, Mobilenet, and Resnet50) utilizing dynamic datasets for breast cancer, lung cancer, and skin cancer (melanoma) diagnosis from accuracy and precision perspectives.

1.2.1 Research Problem

Early [42] identification of the cancer cell is one of the most significant problems in cancer treatment. Cancer is frequently discovered in its advanced stages when it has disrupted the function of one or more important organ systems and has spread throughout the body. Approaches for early cancer diagnosis are vital, and they are a prominent focus of discussion right now. A precise diagnosis and stage of the disease are required for the development of a treatment strategy, following the first treatment of a cancerous cell. Clinical tests and physician inspections are both required for this approach of treatment. It is essential for cancer patients and their families to interpret the consequences of cancer so that they can partake in the treatment protocol's planning. Some of the cancer diagnosis systems are contained in this part. It has also addressed some relevant tests which are still being examined. There is also a part that discusses the cancer staging procedure, as well as information on the results provided in pathology reports. It is indeed crucial to acknowledge the limitations of medical testing since no test is 100 percent accurate. The sensitivity, specificity, and false-positive/false-negative rates of any medical test explain its limits.

The terms sensitivity and specificity are being used by statisticians to describe the accuracy of a medical test. The proportion of times a test generates true positives is referred to as sensitivity. A positive result is more likely to imply that the patient has a condition if the sensitivity is near to 100 percent. The fraction of the time that a test produces genuine negatives is referred to as specificity. A negative result is more likely to imply that the patient is genuinely disease-free if the specificity is near to 100 percent.The preliminary blood test may resort to several persons having needless medical treatments if the second test is challenging or dangerous to conduct.

If the first diagnostic is flawed, it may falsely show that the patients have cancer when they are not. When the condition is minor and the patient's condition is unlikely to be jeopardized, a false positive blood test causes less risk. The blood test may impede people from receiving critical therapies if the condition is severe. Any blood test's result is determined by balancing between the test's sensitivity and specificity, as well as the depth of the condition identified. It's critical for patients to review the sensitivity and specificity of testing with their doctors. [63] There are some significant factors for the development of cancer. Of them, one of the riskiest factors is environmental influences. Cancers such as melanoma depend on the environment the most[59]. Research has shown that Europeans or White people have

more chances of getting melanoma than the black because of the lack of Melanin. [63] Again pollution, contagious agents, radiation, and chemical exposure are also some factors of getting cancer. Sometimes, it can be affected because of some human habits. Excessive smoking, even secondhand smoking can be the reason behind the mouth, pharynx, larynx, esophagus, pancreas, bladder, and 90% of lung cancers. Additionally, diet can sometimes be the reason for colorectal, prostate, and breast cancer. Reproductive history can cause breast, prostate, cervical cancer, and some others too. There are some other factors like environment, genetics, age, and sex that can cause cancers.

Cancer is a deadly disease that, if not treated appropriately, can lead to death. As a result, immediate therapy is necessary to cure it. Specialists can detect the location of cancer cells in the body as soon as possible by utilizing traditional methods such as screening, dermoscopy, mammography, MRI along with others. However, during deep learning retraining, the model is given a vast amount of data and learns model weight and bias. For testing, these weights are assigned to multiple network models. Pre-trained values can be used to start the new network model.In the same space, a pre-trained model already exists.

Procedures cannot guarantee an accurate result in terms of specifics. They might also be unable to detect some particular problems. However, even with the complexities, deep transfer learning approaches with CNN algorithms have a high accuracy in detecting cancer cells' states and circumstances.

1.2.2 Research Objective

As most people fail to recognize cancer at an early stage for various reasons, detecting cancer at an early stage is the main objective of our research. Therefore, we want to utilize Convolutional Neural Network (CNN) with transfer learning, [because using these together can detect objects in an image, store the information, and use it to find additional similar cases [9]. The study's goals are as follows:

1. Understanding CNN and transfer learning

2. Collecting cancer image dataset, splitting individual cancer datasets (breast, lung, skin (melanoma) train it

3. Image classifying by using CNN with Transfer learning

4. Evaluating model which will take image input from the dataset and give output for cancer detection

5. Finally selecting deep transfer learning models (VGG16, InceptionV3, Mobilenet Resnet50) combine, analyze and compare the accuracy results

1.3 About CNN Models

Deep learning in the medical field has gained popularity as a means of developing more reliable diagnostic methods in the recent years. It is a category of machine learning algorithms, has prompted a lot of interest in its application to medical

imaging difficulties due to its fast progress.In Deep Learning Systems, Convolutional Neural Network (ConvNet/CNN) is a system that can take an input picture, allocate importance (learnable weights and biases) to several characteristics in the image, and distinguish between them. The fundamental benefit of employing CNNs in image processing is the increased output given the huge parallelism of the framework, combined with the analog signal processing approach that CNNs are known for[53]. Among the classification approaches, the quantity of pre-refining needed by a ConvNet is comparatively less. While basic techniques require hand-engineering of filters, ConvNets can gain knowledge of these filters/features with sufficient training. Furthermore, as the amount and quality of publicly available datasets increased, data-driven techniques, such as the Convolutional Neural Network (CNN), we are able to learn how to extract representative features from raw data. Nevertheless, a substantial quantity of data is required to properly train such deep models, which remains a difficulty in the medical sector. To address this, a method known as transfer learning proposes that characteristics learned to solve one task may be applied to challenges in other domains.

1.4   Thesis Orientation

This chapter primarily refers to cancer and gives a general summary of the research problem statement and objectives. The rest of the paper is laid out as follows: Chapter 2 gives an outline of the prior information for this work, presents the notion of CNN, and explains CNN's application in this study. Chapter 3 offers a discussion of several previously published studies that used neural networks and deep learning to distinguish cancer cases. The dataset, its analysis, and preparation for use in the study are all covered in Chapter 4 and Chapter 5 discusses the models that were utilized in this study . The results are discussed in Chapter 6 and Chapter 7 summarizes the paper. In the end, The Bibliography at last of the paper identifies all of the websites and publications that have been mentioned.

Chapter 2

Background Models

2.1 Convolutional Neural Network

CNNs are special kind of Deep Neural Network that is able to identify and label specific characteristics in images, and that are commonly utilized for examining images. Recognition of mage and video, classification of image, analysis of medical images, computer vision, and natural language processing are some significant uses of them.

Despite being invented by LeCun in the late 1980s[1], CNN's only became popular in 2012, because of the emergence of huge publicly accessible databases like the ImageNet Large scale industrial Visual Recognition Challenge (ILSVRC). ImageNet has become a key benchmark in deep-learning methods, with 1.2 million natural pictures and over 1000 classifications. CNN offers a variety of pre-trained architectural models, including AlexNet, LeNet, , Xception V3, VGGNet, ResNet50, Inception V3, and others.[3]

CNN's feature several hierarchies, which means that the distribution of inputs alters as the training progresses. Inputs which are preprocessed, or earned by the whitening process are very recommendable for attaining superficial results in diversity tasks[55]. However, significant efforts have been made to beat their achievements since they won the ImageNet 2012 competition, including ResNet50, VGGnet, VGG19, GoogleNet, Xception, InceptionV3, Inception-ResNet-V2, and some others.

Yet, the use of these deep learning models in medicine is limited since there are just not enough large datasets to train them from scratch. To address this problem, Transfer Learning has been intensively studied for Computer-Aided Diagnosis(CAD) techniques.[16], [12],[8],[7],[18].

The Transfer Learning method claims that the spatial transformations used to improve the discriminating power of a generic dataset, such as the LIDC-IDRI (Lung), MIAS, and DDSM dataset(Breast), HAM10000 (Skin) may be applied to other situations.[2],[11],[15].

2.2 CNN Building Blocks

For the method of image processing, there are three kinds of neural network layers that are comprised in a ideal CNN architecture. They are pooling layers, fully connected layers and convolutional layers . Batch normalization layers, feature pooling layers, dropout layers, filter banks and dense layers all work unitedly to construct patterns for several object detection works including identification, classification and segmentation in CNNs.

2.2.1 Convolutional layer

CNN's main building block is the convolution layer. It bears the majority of the processing burden on the network. Convolutions occur in the convolution layer, which is the foundation of CNN. This layer is often comprised of

$$\text{Input Image} * \text{Feature Detector} = \text{Feature Map} \tag{2.1}$$

This layer computes the dot product of two matrices, one of which is the set of learnable parameters known as a kernel and the other is the limited section of the perceptron. The kernel is narrower in space than a picture but more detailed. This implies that if the picture has three (RGB) channels, the kernel height and width will be minimal, but the thickness will span all three channels. The convolutional layer consists of a set of convolution filters that may learn features from training data using different kernels and create different feature maps. Convolutional layers are frequently used in place of fully connected layers to speed up the training process.
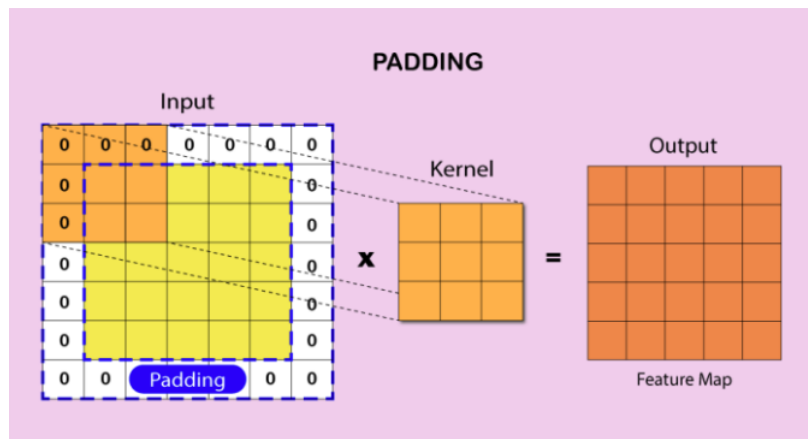
Figure 2.1: Calculating Convolutional Layer (With zero Padding)[50]

The kernel slides over the picture's height and breadth during the forward pass, providing an image representation of that receptive region. This generates a two-dimensional representation of the picture known as an activation map, which contains the kernel's reaction at each geographic place in the visual. A stride is the sliding size of the kernel

If we have an input of size W x W x D and a Dout number of kernels with a spatial dimension of F, stride S, and padding amount P, the size of the output volume may be calculated using the following formula:

$$W_{(\text{out})} = \frac{W - F + 2P}{S} + 1 \qquad (2.2)$$

2.2.2   Padding and Stride

Padding is utilized to make the output dimension identical to the input by adding zeros to the input frame of the matrix. Padding provides extra room for the kernel to cover the picture and is accurate for image analysis. It ensures that information on the edges of pictures is maintained in the same way that it is in the center.

The number of pixels that shift over the input matrix is controlled by the stride of the filter. When stride is set to 1, the filter goes over one pixel at a time, and when stride is set to 2, the filter moves across two pixels at a time. The smaller the stride value, the smaller the output, and vice versa.

2.2.3   Rectified Linear Unit

ReLU is calculated after convolution. It is the more often used activation function that allows the neural network to account for non-linear interactions. In a given matrix (x), ReLU sets all negative values to zero while keeping all other variables fixed.
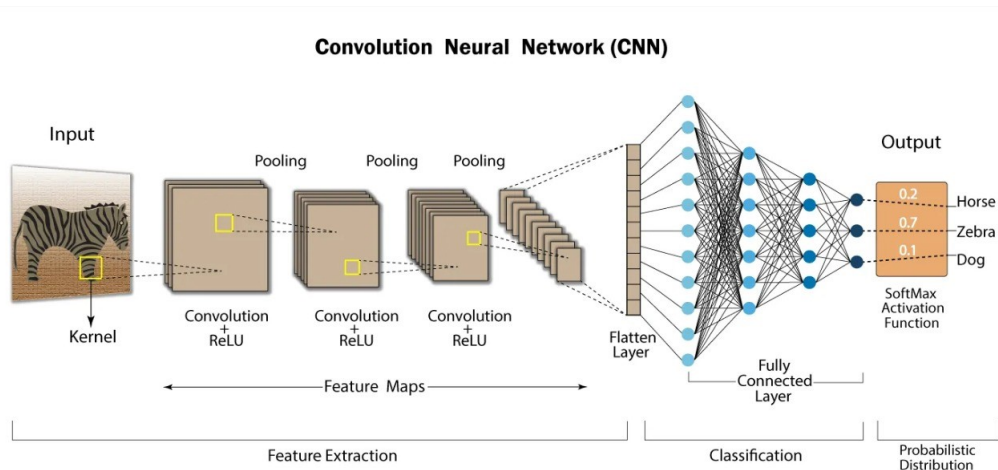
Figure 2.2: ReLU Layer in CNN[50]

2.2.4 Pooling Layer and Max pooling

In most cases, a pooling layer is used to lower the dimension of feature maps and network param It functions individually on each feature map. This decreases the resolution of the feature map by decreasing the length and breadth of the feature maps, but keeps the map's components essential for categorization. This is known as Down - sampling. Max-pooling chooses the most elements from the feature map. The resultant max-pooled layer contains critical characteristics of the feature map. This is the simplest often used method since it produces the best outcomes.

2.2.5 Batch normalization

Batch normalization is a technique for converting inputs to zero-mean and unit variance. A high level knowledge is useful for making data relevant among characteristics. This is often assumed to result in quicker learning rates (although there are other interpretations to its effectiveness).

2.2.6 Fully Connected Layer

Fully connected layer is used to merge the feature maps into a feature vector for classification. It resembles a typical neural network, connecting all neurons and forming the network's last few levels. This fully-connected layer receives the output from the flatten layer.

Following training, the feature vector from the fully connected layer is utilized to categorize pictures into distinct categories. All of this layer's inputs are linked to every activation unit in the following layer. Overfitting occurs because all of the parameters are occupied in the fully-connected layer. Dropout is one strategy for reducing overfitting. Fully connected layers take a lot of computing work during the training process [22].

2.2.7 Dropout

Overfitting in the training dataset is common when all characteristics are linked to the FC layer. Overfitting happens when a model performs so well on training data that it has a detrimental influence on the model's quality when applied to fresh data. To address this issue, a dropout layer is used, in which a handful of neurons are removed from the neural network during the learning phase, leading to a smaller model. When a dropout of 0.3 is reached, 30% of the nodes in the neural network are dropped out arbitrarily[48].

2.2.8 Last Layer (Activation Function)

The activation function is a critical element in the CNN model. It is used to develop and estimate any type of ongoing and complicated connection involving network variables. In other words, it determines which model data should be fired ahead and which should not at the network's end. It introduces nonlinearity into the network. There are various popular activation functions, including the ReLU, Softmax, tanH, and Sigmoid functions. Every one of those has a special use. The sigmoid and softmax functions are recommended for a binary classification CNN model, while softmax is often utilized for a multi-class classification.

To implement a pre-trained CNN model, remove the output layer (last layer) from the model and downsize the remaining model's output to a one-dimensional vector. As a consequence, the new model simply conducts a transformation (characteristic extraction) on the input data, rather than categorization[6],[10],[36].

2.3 Training the convolutional neural network

All Neurons in a particular Layer provide an Output,[62]however they do not have the same Weight as the Neurons in the following Layer. This implies that if a Neuron on a layer notices a particular pattern, it may have less significance in the overall image and will be partly or entirely muffled. This is known as weighting: a large weight indicates that the input is significant, whereas a lower weight indicates that it should be ignored. Each Neural Connection between Neurons will have a Weight. Weights will be modified during the program to meet the goals we've established. To put it simply: Training a Neural Network entails determining the right Weights of Neural Connections through a feedback loop known as Gradient Backward Propagation.

2.3.1 Loss Function

The loss function evaluates the difference between the algorithm's present result and its predicted result.[61] This is a tool for assessing how well an algorithm models data[66]. It converts a conceptual assertion into a realistic statement. Building a highly accurate predictor necessitates continuous issue reinvention by asking, modeling the issue with the selected technique, and evaluation. The only criterion used to evaluate a predictive method is its performance - how accurate the model's judgments are. This necessitates the development of a method for determining how

distant a specific iteration of the model is from the real data. This is when loss functions enter the picture. Loss functions calculate how far an estimated value deviates from its real amount. A loss function connects choices to their effects. Loss functions fluctuate according to the work at hand and the aim to be achieved. The loss function in our paper is a binary cross entropy loss function which analyzes every projected probability to the true class result, which might be 0 or 1. The score is then calculated, which marginalizes the probabilities depending on their difference from the predicted result. That is, how similar or distant the result is to the real result.

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \quad CE = -\sum_i^C t_i log(f(s)_i)$$
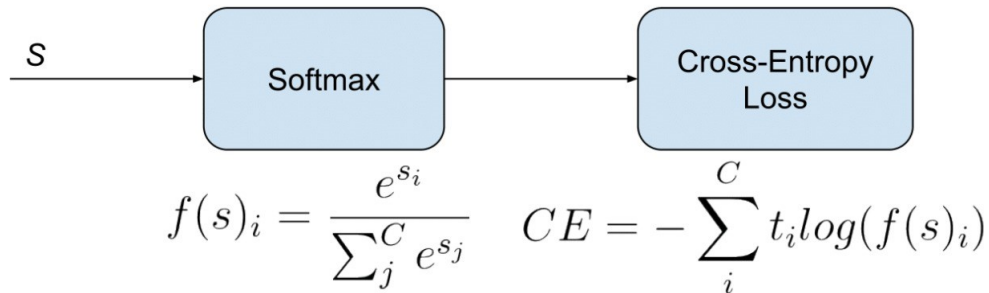
Figure 2.3: Softmax activation plus a Cross-Entropy loss [65]

2.3.2 Gradient Descent

Gradient descent is an optimum approach used in Neural Network model training. It is built on a convex function and continuously changes its variables to reduce a given function to its local minimum.It is now the widely used optimization approach in machine learning and deep learning. It is used for data model training, may be integrated with any technique, and is simple to comprehend and apply[52].

2.3.3 Adam Optimizer

Adaptive Moment Estimation is a method for optimizing gradient descent algorithms. When dealing with huge problems with a lot of data or variables, the approach is quite efficient. It is quick and takes minimal storage. It appears to be a hybrid of the 'gradient descent with momentum' and the 'RMSP' algorithms[54]. Momentum is the approach used to speed up the gradient descent process by considering the gradients' 'exponentially weighted average.' Using averages accelerates the algorithm's convergence to the minima.

$$w_{\text{t+1}} = w_t - \alpha m_t \qquad (2.3)$$

Here,

$$m_t = \beta m_{\text{t-1}} + (1 - \beta)[\frac{\partial L}{\partial w_t}] \qquad (2.4)$$

m_t = time t [current] aggregation of slopes ($m_t = 0$ at first)
$m_{\text{t-1}}$ = time t-1 aggregation of gradients [prior]
w_t = weights at time t
$w_{\text{t+1}}$ = weights at time t+1
α_t = rate of training at time t
∂L = is the Loss Function's derivative
∂w_t =weights' derivative at time t
β = parameter for the moving average (const, 0.9)

On the other hand, RMSprop, or root mean square prop, is an adaptive learning approach for improving AdaGrad. It uses the 'exponential moving average' rather than the cumulative sum of squared gradients, like AdaGrad does.

$$w_{\text{t+1}} = w_t + \frac{\alpha_t}{(v_t + \epsilon)^{1/2} * [\frac{\partial L}{\partial w_t}]} \qquad (2.5)$$

Here, w_t= Weight of time at t
$w_{\text{t+1}}$= Current weights
αt = pace of learning over time t
∂L = loss Function Derivative
∂w_t = weights' derivative at time t v_t = s the sum of the squares of previous gradients. ($v_t = 0$ at first)[e.g sum $\frac{\partial L}{\partial w_t}$]
β = parameter for the moving average (const, 0.9)
ϵ = A little positive number(10-8)

Adam Optimizer takes the best features of the previous two techniques and improves on these to provide a better efficient gradient descent.

2.3.4 Overfitting

At the time when a multivariate model positioned completely in opposition to it's training data, it is known as overfitting. The algorithm can not be implemented in correct way in opposition of an unseen input which lessens it's objective. The capability to employ machine learning algorithms everyday to develop prediction and classify data is ultimately due to the generalization of a model to new data. When a model is trained on sample data for too long or becomes too complicated, this might learn the "noise," or irrelevancies, inside the dataset. The model is "overfitted" when it memorizes the noise and fits too closely to the training set, and it is unable to generalize adequately to new data. If a model is unable to generalize adequately to new data, it will be unable to accomplish the classification or prediction tasks for which it was designed.

Overfitting is indicated by low error rates and a high variance. To avoid this, a portion of the training dataset is usually put aside as the "test set," which is used to check for overfitting. Overfitting occurs when the training data has a low error rate and the test data has a high error rate.[69]

We may now attempt to address the overfitting. There are several ways to accomplish this.

1.**Early stopping** : This strategy has the danger of terminating the training process, resulting in the opposite problem of underfitting. The ultimate objective is to find the "sweet spot" between underfitting and overfitting.

2.**Train with more data** :If clean, relevant data is fed into the model, this strategy becomes more successful. Otherwise, the model may overfit if we continue to add complexity.

3.**Data Augmentation** : While it is preferable to insert clean, relevant data into our training data, noisy input is occasionally used to make a model more stable. This strategy, however, should only be used on a limited basis.

4.**Feature selection** : The process of selecting the most significant features within the training data and then deleting the unnecessary or redundant ones is known as feature selection. Remove layers or reduce the amount of items in hidden layers to increase network capacity.

5.**Using regularization**: which entails costing the loss function for high weights. Applying Dropout layers will delete specific characteristics at random by setting them to zero.

2.3.5   Transfer learning

Transfer learning is the process of improving update learning in a new activity by transferring knowledge from a previously acquired related task. While giving the vast resources necessary to train deep learning models or the huge and challenging datasets on which deep learning models are trained, transfer learning is popular in deep learning. Transfer learning involves initially training a base network on a base dataset and task, and then repurposing or transferring the acquired features to a second target network to be trained on a target dataset and task. This approach is more likely to succeed if the features are universal, i.e. applicable to both the base and target tasks, rather than specific to the base task[49].

2.3.6   Reasons to use CNN in this research

Convolutional neural network (CNN) is a popular and widely used DL network [19, 20]. Nowadays, DL is quite popular due to CNN. The key benefit of CNN over its predecessors is that it automatically recognizes significant features without any human supervision, making it the most widely utilized. Because of the tremendous expansion and evolution of the field of big data, DL has become an extremely popular sort of ML technique in recent years. The following are the advantages of implementing CNNs over other standard neural networks in the computer vision
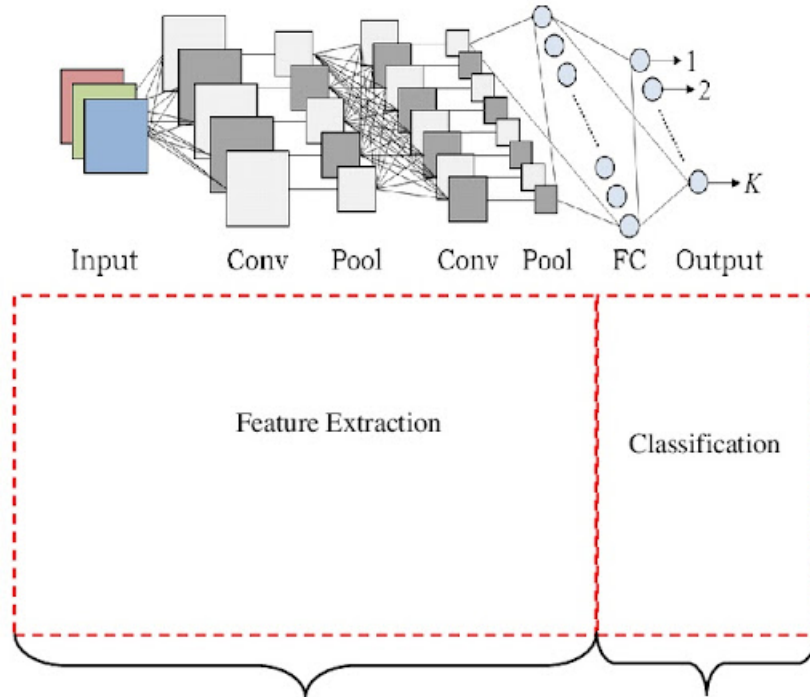
Figure 2.4: Feature Extraction in CNN layers[24]

environment [39]:

- The weight sharing feature, which decreases the amount of trainable network parameters and so allows the network to improve generalization and avoid overfitting, is the primary reason to consider CNN.

- Learning the feature extraction layers and the classification layer concurrently results in a model output that is both highly organized and depends heavily on the extracted features.

- CNN makes large-scale network deployment much easier than other neural networks.

Chapter 3

Literature Review

This segment seeks to critically examine prior important work in the field of cancer detection methods using Convolutional Neural Networks, with a focus on deep transfer learning architectures. Deep learning has been one of the most popular disciplines for categorization and recognition in recent years[29]. One of the most well-known deep neural networks is CNN. The input layer, concealed layers, and output layer make up the network. The convolutional layer,dropout layer, fully connected layers and others are all found in the hidden layer.The Convolutional Network was created by combining these layers.During deep learning retraining, the model is given a large amount of data and acquire knowledge about the weight and bias of a model. For testing, these weights are assigned to multiple network models. Convolutional Neural Networks (CNNs) are deep transfer learning systems that can learn invariant characteristics.Pre-trained values can be used to start the new network model.In the same space, a pre-trained model already exists[2]. Now researchers have been using CNN in many fields including cancer detection. For detecting cancer by CNN, researchers collect data and use them on pre-trained CNN models to classify normal and abnormal cancer cells. Breast cancer is frequently discovered following mammographic images , a type of dataset with an unique sort of X-ray that is then examined by radiologists for a more accurate detection. Another type of dataset that has been used in breast cancer detection is histopathological images which obtains the images usually involving a sample of the afflicted area. Traditionally, it is recommended to use Computed Tomography screening (CT scan) in lung cancer detection[70].So, in CNN, researchers use the image dataset of CT scans to detect lung cancer. Moreover, to detect skin cancer, researchers use dermatoscopic image dataset are the collected images done by dermoscopy, a straightforward tool, in vivo method that is most commonly used to examine colored skin conditions[67]. A good number of researches have been done with above mentioned datasets. The details of the datasets and review of related studies have been summarized below.

3.1 Research Using Mammograms

Mammogram is the X-ray image of breast. It is a method which used by doctors to detect initial indications of breast cancer. Mammograms are divided into two types: digital and traditional.[20] Traditional mammography is scanned and saved on film, whereas digital mammograms are scanned and recorded on a computer so that the data may be improved, enlarged, or altered for future study.[64] Mammography is a commonly used diagnosis system in breast cancer[26]. In mammography, the low-energy X-ray uses two types of angles[5] which are the craniocaudal view(CC) and the mediolateral-oblique view(MLO). Earlier mammographic[15] diagnosis using computer-aided detection (CAD) approaches can increase clinical results and patient survival times for breast cancer. Traditional CAD systems rely on manually extracted features, but they have a number of limitations. For example, hand-crafted features are often domain specialized, and the feature creation process can be time-consuming, complicated, and non-generalizable.An alternate way for feature extraction is to use a Convolutional Neural Network (CNN) to detect faces from whole pictures.

Different machine learning approaches have been used in recent research to identify breast cancer/tumors using mammography. The most regularly used public mammography datasets are the MIAS and the Digital Database for Screening Mammography (DDSM), and 10-fold cross validation is widely used to validate trained models. Some research employed typical automatic feature extraction (not manual extraction) approaches to collect features, such as the Gabor filter, fractional Fourier transform, and Gray Level Co-Occurrence Matrix (GLCM), and then classified them using SVM or another classifier. Neural networks were also utilized as classifiers, and CNN was used in certain research to derive characteristics from mammographic pictures. As transfer learning techniques, some of these experiments employed pre-trained CNN. Yet, some earlier research revealed outcomes gained by utilizing CNN for feature creation and classification in mammograms for breast cancer diagnosis.Dataset for the Curated Breast Imaging Subset (CBIS-DDSM). CBIS-DDSM is an upgraded version of the DDSM dataset. The primary goal of this dataset is to improve and update the DDSM's picture classification. The CBIS-DDSM analyzes specialized and classification approaches while also updating the region of interest (ROI) labeling. For training and testing any breast cancer detection algorithm, the dataset comprises over 1000 pictures that are divided into two categories of abnormalities: rigidity and density.The INBREAST dataset was a breast research group's public mammography dataset. The Breast Center at CHSJ Porto Hospital of St. John (CHSJ) provided the data, which was published in 2010. There were 115 DICOM-formatted cases in total, with 90 pictures in two perspectives (CC, MLO) and 115 cases (410 photos), with 90 cases (4 images per case) from women with both breasts and 25 instances (2 images per case) from breast critical care. According to the BI-RADS categorical, the INBREAST dataset comprised bulk, stiffness, and normal pictures. The dataset is no longer available.

Dataset	Mias	DDSM	CBIS-DDSM	INbreast	MIRacle	Magic5	Nijmegon	Trueta	IRAM	MALAGA	LINL
Original	UK	USA	USA	Portugal	Greece	Italian	Netherlands	Spain	Germany	Spain	USA
Year	1994	1990	2017-2018	2010	2009	2002	1998	2008	2008	2008	2008
Number Of Cases	161	2620	6775	115	196	976	21	89	NA	35	50
Number Of Images	322	10,480	10,239	410	204	3369	40	320	10,500	NA	198
Views	MLO	MLO,CC	MLO,CC	MLO,CC	NA	MLO,CC	MLO,CC	MLO,CC	MLO,CC	MLO,CC	MLO, CC
Image Type file	PGM	LJPEG	DICOM	DICOM,XML	NA	DCOM	NA	DCOM	Several	RAW	ICS
BI-RADS	NO	YES	YES	YES	YES	NO	NO	YES	YES	NA	NA
Ground Truth	YES	YES	YES	NO	YES	YES	YES	YES	NO	NO	NO
Patient Information	NO	YES, AGE	YES,AGE	YES	NO	YES,AGE	NA	NA	NA	NA	NO
Dataset Type	Private	Public	Public	Public	Private	Private	Private	Private	Private	Private	Private

Table 3.1: Summary of mammographic datasets[15]

We looked at a few recent studies that were quite similar to ours. In breast cancer, we have seen many studies using CNN with transfer learning.Several researchers used CNN to diagnose breast cancer/abnormality based on mammograms using transfer learning. In this paper[15], They have used MIAS and DDSM for image processing and classifying.The researchers used MIAS PGM formatted images and DDSM Utility to convert DDSM images into PNG.They Run these images into MATLAB and use ROIs to train the model.The normal and abnormal ROIs will be cropped to small rectangular and the size of the abnormal will be different based on the boundaries. Since CNN(VGG16 model) takes only RGB images of a particular size, they changed those images according to it and resample it. This study used pre-trained CNN with hand-crafted features that give average accuracy (benign vs malignant) is 91.02%, AUC is about 0.76. Another study that used pre-trained AlexNet that gives accuracy (malignant vs nonmalignant) is 90.02% and AUC is about 0.85. On the other hand, our study is pre-trained VGG-16 which gives accuracy (abnormal vs normal) is 90.5% and AUC is about 0.96. So, we can say our study has almost an accuracy level compared with the previous studies. However, The classifying reliability level for benign vs malignant is lesser in our study than for abnormal vs normal. Furthermore, the CC view's mammography classifying reliability (0.931) is higher than the MLO view (0.887). In our investigation,[9] we utilized the VGG-16 model, which is shallower (23 layers) than newer models like InceptionV3 (159 layers) and ResNet50 (168 layers).They compared their research with similar research with the same datasets.Another study that proposed a CAD system, DCNN-SVM –AlexNet –classify benign and malignant masses which gives accuracy of 87.2%[30].

Main Method	Validation (# of images)	Accuracy %	AUC
Pre-Trained CNN on LSVRC datasets & Fine-tuning + Two Step decision	2-fold cross (600)	(Ben-Mal) 96.7	-
Pre-Trained CNN with hand crafted features + RF	5-fold cross (410)	(Ben-Mal) 91 ± 0.02	0.76
Pre-Trained AlexNet + Sparse MIL	5-fold cross (410)	(Mal-nonMal) 90.00 ± 0.02	0.85
Pre-Trained VGG16 + one FC layer	10-fold cross (2600)	(Abnorm-Norm) 90.5 ± 3.2	0.96

Table 3.2: Related Research Analysis[15]

3.2 Research Using CT scans Images

Deep learning CNNs have lately demonstrated amazing effectiveness in lung nodule identification.Diagnostic pictures are handled immediately in CNN using numerous convolutional layers that act as spatially localized filters and completely linked layers, most typically without segmenting the lung fields. The regained deep learning feature vector demonstrated a strong dimensions to accurately characterize the data which are trained and distinguish between normal and malignant lung nodules. [28]To conceal the pictures into a dense superior-volumed feature field, many CNN architectures need huge number of weights or boundaries. Generally, Imaging tests are common lung cancer nodule detection. Imaging tests generate images of the inside of the body by using x-rays, magnetic fields, sound waves, or radioactive chemicals are used for internal detection.

In the paper, The National Cancer Institute established the Lung Image Database Consortium (LIDC) to boost research and development operations (NCI).The LIDC database was established with three types of items to be marked by four radiologists: nodules more than or equal to 3 mm in diameter with assumed histology, nodules less than 3 mm in size with an uncertain origin, and non-Nodules smaller than 3 mm in diameter but benign. The (LIDC)[14]dataset is utilized for preprocessing. The computed tomography (CT) scans of 1018 patients are included in this data collection.

It also includes an XML file for each patient that comprises the four radiologists' individual annotations. To work with the dataset, multiple CT slices of each patient are downloaded and stored in a directory depending on the XML file.
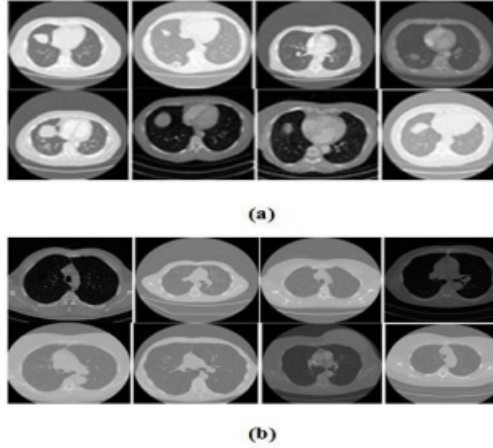
Figure 3.1: Samples of images a) Malignant b) Benign form (LIDC dataset)[58]

A variety of region-growing and morphological techniques are used in the preprocessing stage. To help with feature extraction, it recognizes and distinguishes lung structures and nodules. The malignant features of the annotations provided in the XML file are taken into account for binary categorization. All patient slices are considered malignant if the Malignancy value is more than 3. These slices are deemed benign if the Malignancy grade is less than or equal to 3. The pictures from the LIDC dataset are in DICOM format[58]. The DCM format is used to store these medical pictures.Images are transformed into the.jpg format, along with the same labels, to conduct the experiment for successful picture categorization. The transformed pictures are organized into two directories: benign and malignant. Images are not subjected to any further processing; they are just sent out into the network as it is.

For actual lung cancer diagnosis, The size of the tumor is primitively used. It has appeared that death caused by lung cancer has decreased by screening with Low-Dose Spiral Computed Tomography (LDCT).

Medication for lung cancer is determined by whether the tumor is a small cell (13 percent) or a non-small cell (84 percent).

An Automatic Lung Nodule Detection System Using a Multi-group Patch-Based Deep Learning Network[17], the input is multi-group 2D Lung CT images, which are utilized in the LIDC dataset. There are three steps to it. A slope analysis approach is used to restore lung shapes. The Frangi Filter is then used to remove the vessel-like feature in a CT picture. After that, the CNN structure is tested on two sets of pictures; one set of original images, and the other set of binary images created by sophisticated binarization processing to determine whether or not the nodule is malignant.

The goal of segmenting lungs from a CT scan is to find distinct characteristics that will help the classifier better categorize the candidates. This is also necessary since the CT scan is too large to be sent directly into the classifier. The classifier will take a long time to detect distinguishing features from the massive DICOM pictures. The segmentation of lung structures is a difficult challenge due to the lack of uniformity in the lung area. The pulmonary structures have comparable density, but distinct scanners and scanning techniques. Besides,[23] contemporary lung Computer-Aided Diagnosis (CAD) systems can help medical decision-making by using just chest CT

scans and completing an automated assessment of pulmonary nodules. The ultimate objective of these systems is to distinguish between cancerous and non-cancerous nodules.

Some researchers cropped the original pictures from the datasets into smaller plots based on the centroid position of the malignant nodules and utilized them as cancer cases using Convolution Neural Networks[13]. Furthermore, when it was constructed to Multi-Level CNN and tested it on the LIDC dataset, an accuracy of 84.81 percent was obtained reported by researchers[25]. Thus to identify lung cancer, a respective number of researchers have propositioned contrasting techniques using deep transfer learning.

3.3 Research Using Dermatoscopic Images

Skin cancer (Melanoma) can sometimes be detected through optical observations at the primary stage but most of the people at that time don't care about it and come to consult a doctor at later stages. But despite that most people are only treated by doctors once the illness has progressed. Training in dermoscopy has been shown to improve the early detection of melanoma for both non-expert dermatologists and primary care providers in clinical studies. Dermatologists have created dedicated clinical systems for image collection, management, and retrieval for future follow-up and monitoring[56].

Dermoscopy is a diagnostic process that is used to identify small lesions from a broad area of the body. Other than using dermoscopy, image classification can be used for the same purpose. For the image classification[34], Convolutional Neural Networks pre-trained on ImageNet along with transfer learning can be used so that it can recognize the edges of the object in an image, can store the data when solving the case, and finally can apply it to solve another similar kind of case. Training the model on the HAM10000 dataset which is a vast collection of dermatoscopic images of general pigmented skin lesions from multiple sources so it would automatically get classified as one of the seven classes of that dataset.

Deep learning methods have been implemented to detect and classify malignant cancer in medical pictures, which can assist radiologists in making judgments, particularly in difficult-to-identify instances, while boosting accuracy and efficiency. The pre-trained foundation models to use are ResNet50, InceptionV3, VGG16 and MobileNet with Keras deep learning library as Keras' pre-trained networks are capable of accurately detecting 1000 different object categories which are analogous to objects we see in our daily lives. However, a comparatively recent technique known as Big Transfer might soon take over the place of ResNet50 as the concept of bigger pre-training has been the main emphasis of this technique. Even with a small number of data, Big Transfer operates admirably.

In the paper[34], the researchers have used HAM10000 as their dataset. This dataset has covered all 7 distinct skin cancer cases(Actinic Keratosis, Basal cell carcinoma, Melanoma, Benign Keratosis, Dermatofibroma, Vascular Skin Lesion, Melanocytic Nevi) and is consists of 10015 dermoscopic images with a resolution of 600×450 pixels[27]. So,they have trained their model with this dataset, then they got almost full coverage as they got all kinds of skin cancer-related problems output from it.However, all of these 7 types of distinct skin cancer cases can be covered from only 7470 images of the dataset. So rest of the 2545 images are the copy of existing

images. Also, this dataset can be particularly inconsistent as the number of images of different cancers differs from one another. Identical photos were discovered during an early experimental data analysis. By removing identical photos, a validation set with an 83:17 divide ratio was created. The training set now has 9077 photos, while the validation set contains 938. With the ultimate picture size set to 224x224 pixels, data augmentation in the form of flipping, cropping, and rotating was performed. To preserve the inter-class proportion, stratified sampling was adopted. The pictures from the training set are then fed via a special pre-process function that takes into account the model architecture. Each model needs input in a specific structure, and the preprocess function assists us in transforming their data into that layout. The
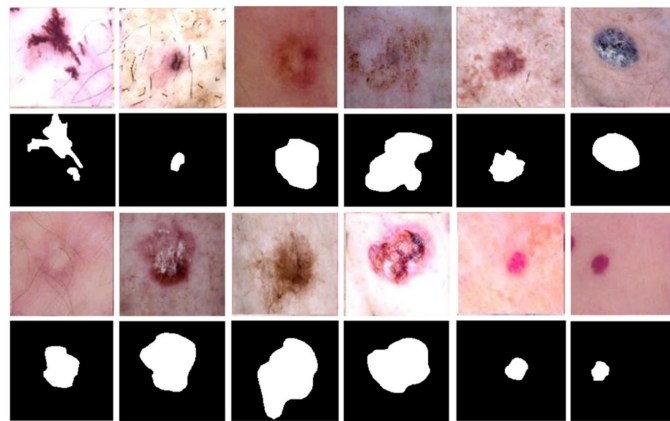
Figure 3.2: Malignant and Benign Samples of Melanoma[41]

researchers tested pre-trained models with the dataset and computed the weighted average recall,accuracy and weighted average precision of the models:

Architecture	Accuracy	Weighted Average Precision	Weighted Average Recall
ResNet50	90	89	90
Mobilenet	87	86	87
Xception	84	84	84
Inception V3	85	85	85

Table 3.3: Tested Architecture of paper [34]

Based on the weighted recalls, they chose ResNet50 as their ultimate base model for their custom model. They prepared their model and compared it with other models. In this table, the ResNet50 is considered as the custom model of the researchers in this paper :

Architecture	Accuracy	Precision	Recall
Mobilenet	92.70	87.00	81.00
Efficient B1	94.00	94.00	94.00
Inception ResNet From Scratch	83.96 78.00	72.00	69.29
ResNet50[34]	90.00	89.00	90.00

Table 3.4: Accuracy of other work [34]

Chapter 4

Dataset

4.1 Data Collection and Analysis

For Breast cancer, we have used multiple datasets which include mammographic images. The datasets we have used are MINI-DDSM [35] and KAU-BCMD each 50%[38].

It is a lighter edition of the now-outdated DDSM (Digital Database for Screening Mammography) data collection. To address the persistent issue of why Mini-DDSM, it's vital to understand that the DDSM database has a website maintained by the University of South Florida for the purpose of maintaining it online. Images, on the other hand, are condensed using a faulty program that uses lossless JPEG (i.e. ".LJPEG") encoding (or at least an outdated tool as described on the DDSM website). Although CBIS-DDSM supports an alternate host for the original DDSM, photos are sadly shorn of its initial title and year tag. To get it in order, it took a lot of work, code, and machine processing capacity, Here are some of the benefits of this new Mini-DDSM version:

1. The goal is to enable the DDSM accessible to everyone (half resolution though)

2. The age/density characteristics, patient files, initial identifying filename, and lesion binary mask are all included in the data set.

3. Pictures extracted/loaded from tfrecords are not complicated. If we want photos, we can get them whether we are working with Python, MATLAB, JAVA, or C++, the pictures are saved as pictures.

4. Because the lesion binary mask is based on the original freeman chain-coding, this set of data saves us the trouble.

5. The data collection includes an excel sheet featuring direct access to all picture features and metadata.

6. Access is public and available, with no long processes or papers to fill out or approve.

7. The MINI-DDSM-Complete in JPEG format (4 GB) data set in response to multiple requests from those with machine/internet capacity limits that prevent them from downloading the 45GB data set.
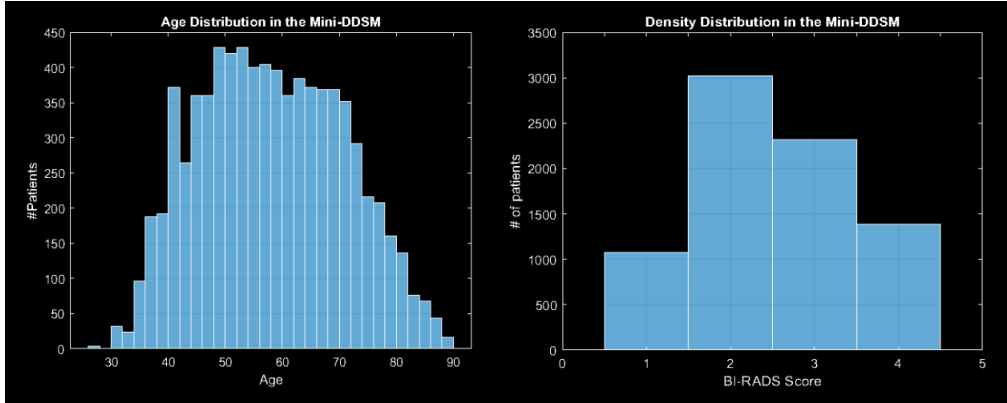
Figure 4.1: Dataset Dispersion by Age and Concentration in the Mini-DDSM[35]

From April 2019 to March 2020, KAU-BCMD, the planned mammography dataset was gathered through the Sheikh Mohammed Hussein Al-Amoudi Center of Excellence in Breast Cancer at King Abdulaziz University in Jeddah, Saudi Arabia. Between April and June 2020, the inscription was made. Breast imaging technology from IMS Giotto, a GMM Group business, was utilized for screening. The camera produces high-quality photos with a low signal-to-noise ratio. The dataset comprises 1416 instances, including pictures from both breasts (right and left) and two types of views (CC and MLO) for a sum of 5662 mammography images. Following the BI-RAD approach, the dataset was divided into six groups.US scans are used to verify the BI-RADS. Using US scans, three researchers independently confirmed the BIRAD method. The ultimate BIRAD categories are then determined using the consensus vote method.

For skin cancer(melanoma), we have used HAM10000 which is dermatoscopic pictures from diverse demographics gathered and preserved using various modalities. The completed dataset contains 10015 dermatoscopic pictures that may be used as a training set for machine learning in academia. Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratosis(bkl), dermatofibroma(df), melanoma(mel), melanocytic nevi(nv), and vascular lesions are among the cases (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc) [27] More than half of tumors are verified by histopathology (histo), with the other instances relying on follow-up examinations(follow-up), expert consensus(consensus), or in-vivo confocal microscopy confirmation (confocal). The lesion id-column in the HAM10000 metadata file may be used to track tumors with many pictures in the dataset.

Like breast cancer, we have used multiple datasets which are IQ-OTH/NCCD [40] and Chest CT-Scan images Dataset[53]. We mostly gathered our data from IQ-OTH/NCCD and only used around 40 files from Chest CT-Scan images Dataset for abnormal (cancerous) cell inadequacy in the main dataset(IQ-OTH/NCCD). The IQ-OTH/NCCD contains CT scans of patients with lung cancer at various stages, as well as healthy subjects. Oncologists and radiologists from both centers identified the IQ-OTH/NCCD slide. In the autumn of 2019, the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) lung cancer dataset

23

was gathered across three months in the above-mentioned specialty hospitals. It covers CT images of lung cancer patients at various phases and also normal people. Oncologists and radiologists from these two hospitals marked IQ-OTH/NCCD slides. There are 1190 photos in total, reflecting CT scan slices from 110 instances in the collection. Normal, benign, and malignant instances are separated into three categories. There are 40 cancer cases, 15 benign cases, and 55 normal cases among them. The CT scans were initially saved in DICOM format.The scanner utilized is a Siemens SOMATOM. The following CT protocol was utilized for reading: 120 kV, 1 mm layer depth, window widths varying from 350 to 1200 HU, and window centers varying from 50 to 600. with complete inhalation and breath hold Prior to evaluation, all photos were de-identified with board granted written approval.

The organizational regulatory boards of the collaborating major hospitals gave their approval to the research.There are multiple layers in every image. The amount of layers varies between 80 and 200, and each one displays a picture of the human chest from various sides and perspectives. Gender, age, educational achievement, residential location, and living situation are all different in the 110 instances. A few work for the Iraqi ministries of transportation and oil, while others are farmers and gainers. The majority of them are from Iraq's central area, specifically the provinces of Baghdad, Wasit, Diyala, Salahuddin, and Babylon. The dataset is available online at Kaggle.

Additionally, Chest CT-Scan images Dataset is a dataset which was made for a project of chest cancer detection with machine learning and deep learning (CNN). The target of this collected data was to make the models classify the images effortlessly.To suit the models, the photos are in jpg or png format instead of dcm. The data includes three forms of chest cancer: large cell carcinoma, and squamous cell as well as adeno carcinoma and one folder for normal cells. The Data folder is the primary folder that contains all of the phase folders, such as test, train, and valid. Valid represents validation set, test represents testing set, and train represents training set. The training set is 80% and the testing set is 20%.

4.2 Preprocessing Data for models

The images must be preprocessed before being utilized in the CNN models to guarantee that the models' predictions are free of bias or inaccuracies caused by the characteristics of the data. For the pre-processing, we followed the following steps.

4.2.1 Gathering Dataset

We used a variety of cancerous cell images, including breast, lung, skin (melanoma) cancers. We utilized around 3000 images of cancer cells designated as cancerous. The dataset is available in a variety of forms and may be used with a number of different deep learning models. This data collection will be used for educational and research purposes. In order to apply our approach, we used certain deep transfer learning models.

4.2.2   Image labeling

Image labeling is the process of recognizing and labeling numerous elements within an image. Using on-device and cloud-based technologies, this technique can recognize features in images automatically. The most prevalent type of image labeling in computer vision is bounding boxes. Bounding boxes are rectangular rectangles that define the location of an item. They may be computed using the x and y-axis coordinates in the rectangle's upper-left corner and the x and y-axis coordinates in the rectangle's lower-right corner. Bounding boxes are commonly used in tasks involving object detection and localization. Bounding boxes are often expressed by two coordinates (x1, y1) and (x2, y2), or a single coordinate (x1, y1) plus the bounding box's width (w) and height (h).

4.2.3   Images to array conversion

Numpy is the most popular and commonly used package in Python for image data processing. The array() method in NumPy turns PIL pictures into NumPy arrays. The np. array function also returns the same result. The class of an image is returned by type. You may reverse the procedure using the Image from array() method. When working with NumPy.ndarray picture data that we subsequently want to export as a PNG or JPEG file, this method comes in handy. The print function returns the value of each pixel in the NumPy array picture (data). In order to load and show a picture, we used Matplotlib, OpenCV, and the Keras API. After that, we used the NumPy libraries to conduct simple image editing and store it to our local system after converting the loaded photos to and from the NumPy array. Images are read as arrays in both the Keras API and OpenCV.

4.2.4   Data normalization

For the convenience of use, we restructure and resize the whole dataset's photos into a single size which is 256*256.

4.2.5   Data Cleaning

Data cleaning is the practice of deleting or cleaning any undesirable data from a dataset so that the remaining data may be utilized to make better decisions. It is a crucial stage in machine learning since the quality of the data can have a big influence on the model's accuracy. Text cleaning, entity recognition, non-speech feature extraction, and image cropping are the most prevalent methods of data cleaning in machine or deep learning. There is some duplication data in our three types of datasets. Since duplicate images might influence the findings and lead to unexpected outcomes, we screened for duplicates and only selected images that were unique.

4.2.6   Data Encoding

Categorical encoding is the conversion of categorical data to numerical data. It employs a series of rules to transform perfectly spelt words that do not follow a number of laws into a number between 0 and 1. OHE (One-Hot Encoding) is a

categorical encoding method that converts all elements on a categorical column into new columns with binary values of 0 or 1 to indicate the existence of the category value. Here we implemented one hot encoding method for categorize our image datasets by converting normal (healthy) image datasets into 1 and cancerous image datasets into 0.

4.2.7 Data train

Data training in deep learning is the practice of employing artificial intelligence (AI) to improve the performance of a Predictive Model. Deep learning algorithms have the ability to give new levels of accuracy and precision.We downloaded pre-trained weights and printed out the InceptionV3, VGG16, MobilenetV2, Resnet50 model after importing the relevant Deep Learning libraries. We started with Inception V3 and subsequently moved on to VGG16, MobileNet V2 and ResNet 50 for training our datasets.

4.2.8 Implementation of Transfer learning

With the recent advances in deep learning, we are now able to train neural networks with huge amounts of datasets. Transfer learning is such a technique of using a neural network that has been pre-trained on a large dataset to solve a related but different problem.We have collected basic input and output for applying transfer learning. The trained model has several pre-trained layers, and we used the very first layer for base input and the fourth final layer for base output. For improved accuracy, we tweaked the models. Maximum layers of them are pre-trained in the model we termed them. As a result, we do not need to retrain them again. We will set different values of the input layers suitable for our datasets.The key advantage of Transfer learning is that it reduces the training time in half and give us output with proper accuracy.

4.2.9 Data Splitting

The technique of labeling portions within an image is known as data splitting or segmentation. It may be used to increase the quality and quantity of data needed to train other machine learning models, such as those that identify and classify objects. Deep learning is one approach to do this. The fundamental concept is to run an image through a sequence of neural networks to get a final output that categorizes the picture into regions. The data segmentation or splitting as well as the preparation were carried out in parallel. Based on an 80:20 ratio, the complete dataset was randomly splitted into 80 percent training data and 20 percent testing data. Furthermore, the divided datasets each have 20 batches, while the goal size is (256, 256).

Here, for each cancer, 1000 images were utilized, with 600 malignant image (cancerous) files and 400 healthy (non-cancerous) image files. We implemented four models for training and testing: Inception-v3,VGG16, MobilenetV2 AND Resnet50. We divided the data set into two sections, with around 80% of the data used for training and 20% for testing.

Data Segmentation	Percentage	Total	Target size	Batch size	Class mode	Subset
Training	80%	800	(256,256)	32	Categorical	Training
Testing	20%	200				

Table 4.1: Data Splitting

Chapter 5

Deep Transfer Learning Models

Transfer learning refers to the application of a previously learnt model to a new challenge. Since it can train deep neural networks with very little data , it's especially popular in deep learning right now . This is especially useful in data science, because most real-world scenarios do not necessitate millions of labeled data points to train complex algorithms. Transfer learning is especially beneficial when the training dataset is small. In this instance, we can give examples like using the pre-trained models' weights to initialize the weights of the new model. For our paper we have used pre-trained models of transfer learning , which have the advantage of being general enough to be used in various real-world applications. For example:

- Text classification needs familiarity with word representations in some vector space. You can train vector representations yourself. The issue here is that you may not have enough data to train the embeddings. Furthermore, training will take a long period. In this instance, you can use a pre-trained word embedding like GloVe to accelerate your development process.

- ImageNet-trained models can be utilized to solve real-world picture classification challenges. This is because the dataset comprises over 1000 classes. We can use these models and fine-tune them to classify images.

[57]ImageNet is a research project that aims to provide a huge image database. It has over 14 million images in over 20,000 categories (or synsets). In addition, they provide bounding box annotations for approximately 1 million images that can be used in Object Localization tasks. When we discuss deep learning and Convolutional Neural Networks, we're usually talking about the ImageNet Large Scale Visual Recognition Challenge, also ILSVRC for short. [19] ImageNet is a massive picture database with about 14 million images. It was created by academics with the goal of doing computer vision research. In terms of scale, it was the first of its sort. A hierarchy of images is sorted and labeled. Machine Learning and Deep Neural Networks are both based on the training of computers using a large dataset of pictures. Machines must be able to extract meaningful information from these training photos. They can utilize these characteristics to identify pictures and perform a variety of other computer vision tasks once they've learned them. Researchers may use ImageNet to compare their models and algorithms against a shared set of pictures. The aim of this image classification challenge is to build a model that can accurately categorize 1,000 different item categories from an input image. Models are trained using 1.2 million training images, 50,000 validation images, and 100,000

testing images. Those 1,000 image classifications depict object classifications encountered in our daily lives, such as dog and cat kinds, numerous home furnishings, transport sorts, and much more. The images are reduced to a fixed resolution of 256 x 256 pixels. A rectangular picture is resized and the middle 256 x 256 patch is clipped out of the final image. The cutting-edge pre-trained networks used for the Keras core library have consistently outperformed Convolutional Neural Networks on the ImageNet classifications that depict object classifications encountered in our daily lives, such as dog and cat kinds, numerous home furnishings, transport sorts, and much more. The images are reduced to a fixed resolution of 256 x 256 pixels. A rectangular picture is resized and the middle 256 x 256 patch is clipped out of the final image. The cutting-edge pre-trained networks used for the Keras core library have consistently outperformed Convolutional Neural Networks on the ImageNet challenge. Such networks also exhibit a massive capacity to apply to images outside of the ImageNet dataset with transfer learning, such as feature extraction as well as fine-tuning.

5.1   Inception V3

Inception v3 is a commonly used image recognition model that has been demonstrated to achieve higher than 78.1 percent accuracy on the ImageNet dataset. The model represents the result of several concepts explored over time by a number of scholars. It is based on the Szegedy, et al original article, "Rethinking the Inception Architecture for Computer Vision." Convolutions, average pooling, max pooling, concatenations, dropouts, and fully linked layers are among the symmetric and asymmetric building components in the model. Batch normalization is done to activation inputs and is utilized significantly within the model. Softmax is used to calculate losses. [44]

5.1.1   Factorized Convolutions

As the set of parameters in a network is reduced, this contributes to improving computational speed. It also monitors the network's performance.[43]

5.1.2   Smaller Convolutions

It is undeniable that substituting larger convolutions with smaller convolutions speeds up learning. A 5 x 5 filter, for example, has 25 parameters; two 3 x 3 filters, in place of a 5 x 5 convolution, have just 18 (3*3 + 3*3) variables.
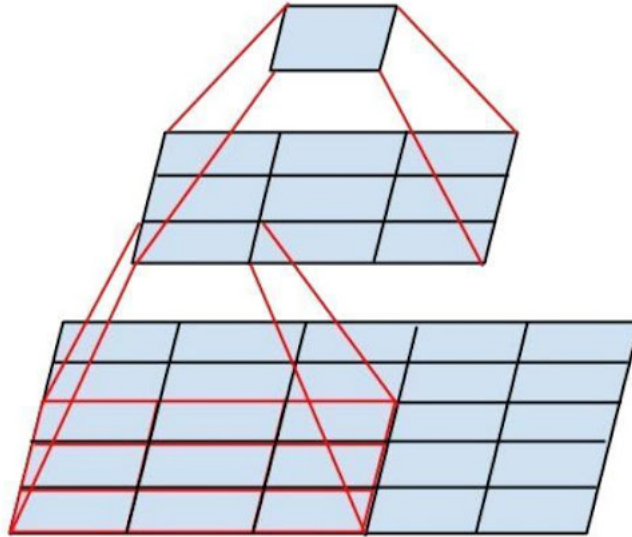
Figure 5.1: Smaller Convolutions[21]

5.1.3 Asymmetric Convolutions

A 1 x 3 convolution followed by a 3 x 1 convolution might be used instead of a 3 x 3 convolution. The number of variables will be significantly greater than the asymmetric convolution described if a 3 x 3 convolution was substituted with a 2 x 2 convolution.

Figure 5.2: Asymmetric Convolutions[45]

5.1.4 Auxiliary Classifier

Auxiliary classifiers are small CNNs that are used to fill in gaps among layers during model training. The accidental loss is often linked to the core network loss. Auxiliary classifiers were used by GoogleNet to generate a larger network, however, Inception-v3 uses them as a regularizer.
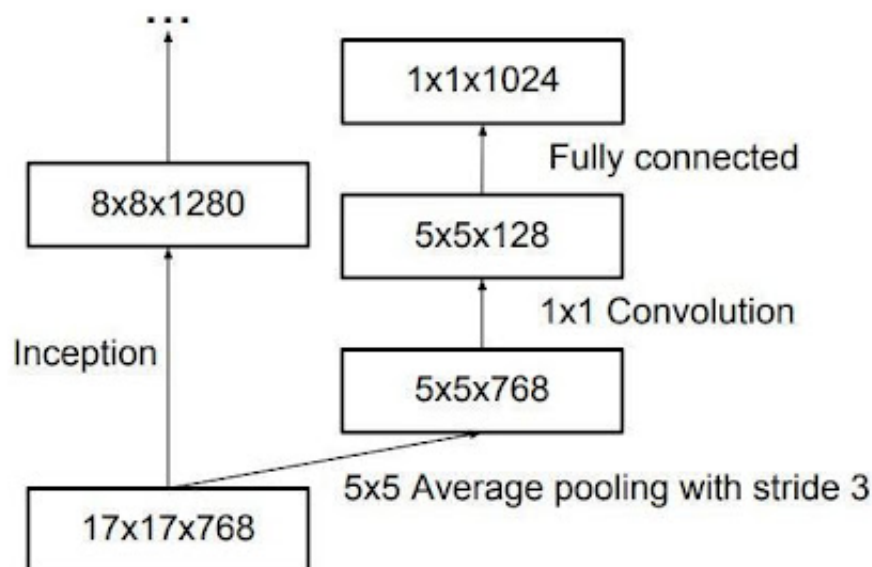

Figure 5.3: Auxiliary Classifier[46]

5.2 Visual Geometry Group(VGG 16)

In the research publication [68] "Very Deep Convolutional Networks for Large-Scale Image Recognition," K. Simonyan and A. Zisserman from the University of Oxford proposed the VGG16 convolutional neural network model. In ImageNet, a dataset collection of above 14 million pictures belonging to 1000 classes, the system achieved a 92.7 percent top-5 accuracy rate. It was a well-known model that was proposed to the ILSVRC-2014. It outperforms AlexNet by sequentially replacing huge kernel-size filters (11 and 5 for the first and subsequent convolutional layers, accordingly) with numerous 3X3 kernel-size filters. VGG16 had been training for weeks on NVIDIA Titan Black GPUs.
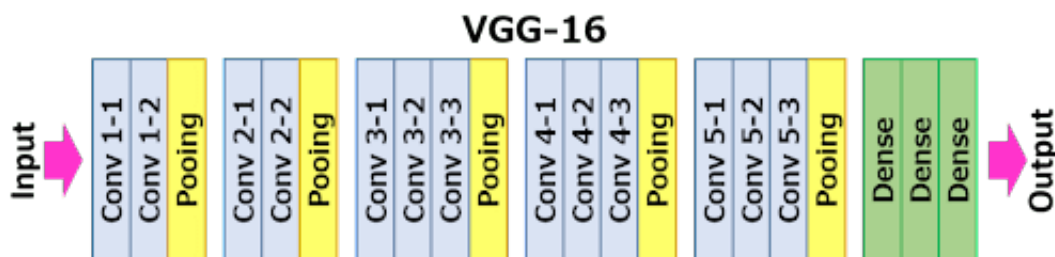

Figure 5.4: VGG-16[47]

A 224 × 224 RGB picture is used as the input to the cov1 layer. The picture is processed through a pile of conv (convolutional) layers with a very narrow receptive field: 3x3 % (the least size to capture the notions of left/right, up/down, and center). It also features 1x1 convolution filters, which are a linear transformation of the input channels in one of the configurations. The convolution step is set to 1 pixel and the spatially padding of convolution layer input is set to 1 pixel for 33 convolution layers so that the spatial resolution is maintained after convolution. Five max-pooling layers, which follow some of the conv., are in charge of spatial pooling. With stride 2, max-pooling is executed across a 2x2-pixel frame.
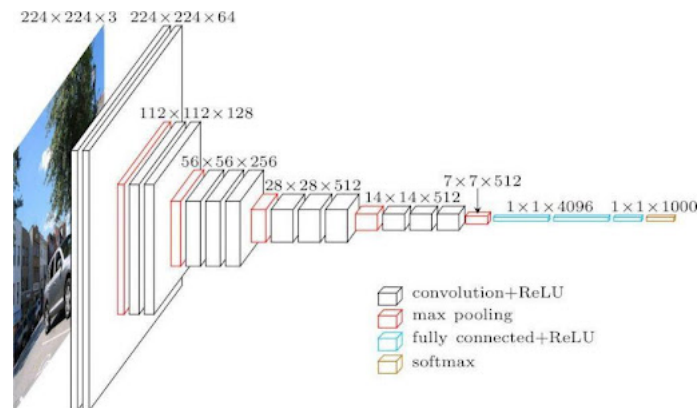

Figure 5.5: A visualization of the VGG architecture[33]

Following a stack of convolutional layers (that have varying depths in various designs), three Fully-Connected (FC) layers are added: the very first two contain 4096 channels individually, while the third executes 1000-way ILSVRC classification and therefore has 1000 channels (1 for each class). The soft-max layer comes last. All networks have the same configuration of completely linked layers. The rectification (ReLU) non-linearity is implemented in all buried layers. Local Response Normalization (LRN), which does not enhance performance on the ILSVRC dataset but increases memory usage and computation time, is also included in none of the networks (except for one).

5.3   Residual Network (RESNET 50)

ResNet50 is a ResNet model variant with 48 Convolution layers, 1 MaxPool layer, and 1 Average Pool layer. ResNet-50 is a 50-layer deep convolutional neural network. This network trained on the ImageNet data set is returned by resnet50. The network accepts 224-by-224 image input. ResNet50's architecture is divided into four stages, as illustrated in the diagram below. The network can receive input images with height and width multiples of 32 and channel width of 3. At the sake of clarity, we'll assume the input size is 224 x 224 x 3. Every ResNet architecture uses 7x7 and 3x3 kernel sizes for initial convolution and max-pooling, respectively.Following that,Stage1 of such a network begins, with three Residual blocks of three layers each. The kernels used to perform the convolution operation in all three layers of the stage 1 block are 64, 64, and 128 in size. The identity connection is represented

by the curved arrows. The dashed connected arrow indicates that the convolution layer in the Residual Block has been performed with stride 2, which means that the size of the input will be split in half in aspects of height and width, but the width of the channel will be doubled. As we progress through the stages, the channel width doubles while the input size is split in half.[51]
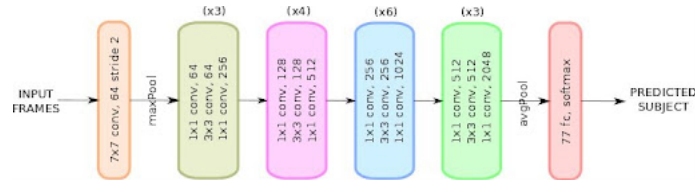

Figure 5.6: ResNet-50 neural network architecture[21]

5.4 MobileNet V2

MobileNetV2 is a convolutional neural network design that aims to be mobile-friendly. It is built on an inverted residual structure, with residual connections between bottleneck levels. As a basis of non-linearity, the intermediate expansion layer filters characteristics with lightweight depthwise convolutions. MobileNetV2's overall design includes a fully convolutional layer with 32 filters, followed by 19 residual bottleneck layers [60].

It is divided into two parts:

- Inverted Residual Block

- Bottleneck Residual Block

In the MobileNet V2 architecture, there are two versions of Convolution layers:

- 1 x 1 Convolution

- 3 x 3 Depthwise Convolution

Every block has three distinct layers:

- Convolution 1x1 with Relu6

- Depthwise Convolution

- Convolution of 1 x 1 without any linearity

Following are the elements for two blocks:

Stride 1 :

- Input

- Convolution 1x1 with Relu6

- Depthwise convolution with Relu6

- Convolution of 1x1 without any linearity

- Add

Stride 2 :

- Input

- Convolution 1x1 with Relu6

- Depthwise convolution to stride 2 and Relu6
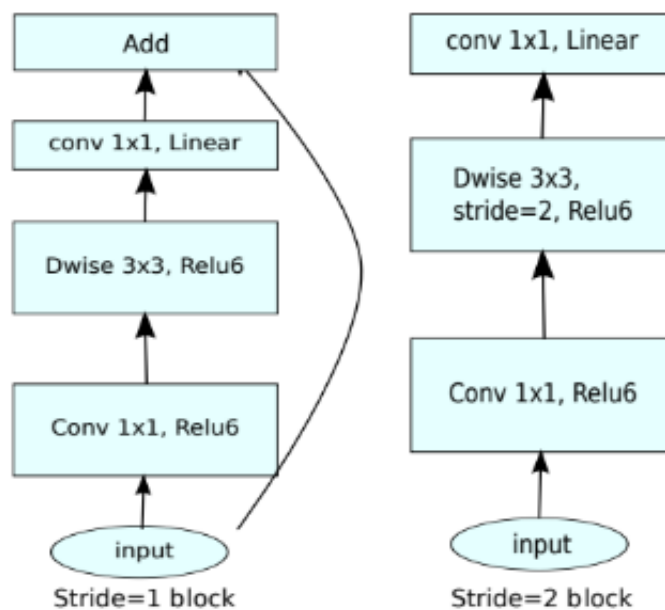
- CConvolution 1x1 with no linearity

Figure 5.7: MobileNet V2 architecture[21]

5.5 Model Evaluation

Model evaluation is the method of analyzing a machine learning model 's output, along with its benefits and limitations, using various assessment criteria. Model evaluation is critical for determining a model's usefulness during the early stages of study, as well as for model maintenance.

5.5.1 Confusion Matrix

The confusion matrix (or confusion table) breaks out the right and wrong categories for every group in further depth. When we need to grasp the difference between groups, a confusion matrix comes in handy, especially when the value of misinterpretation differs between the two classes or we have a lot greater test results on one than the other. Making a false positive or false negative in a cancer detection, for instance, has quite distinct repercussions.

- **True Positive**: when we correctly forecast that an information matches to a classification and it really does.

- **True Negative**: when we correctly forecast that an information matches to a classification and it really does.

- **False Positive**: arise when we assume an information corresponds to a category when it does not.

- **False Negative**: arise when we incorrectly guess that an information does not correspond to a category when it does

The above-mentioned four results are usually presented in a confusion matrix. It demonstrates the binary categorization framework. When the predictions are created on test data and appoint every prediction as one of the four anticipated outcomes indicated above, the matrix would be created

5.5.2 Accuracy

The ratio of the number of correct predictions to the total number of forcasts used to calculate, is known as accuracy. It's easy to understand by simply separating the number of fair calculates by.

$$\text{Accuracy} = \frac{\text{Correct prediction}}{\text{All predictions}} \tag{5.1}$$

5.5.3 Precision

It is expressed as the percentage of applicable occurrences (true positives) included in all the occurrences contemplate to be included in a particular category.

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives + False positives}} \tag{5.2}$$

5.5.4 Recall

The proportion of occurrences estimated to represent a set collated to all of the cases that assuredly fall to the category.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives + False negatives}} \tag{5.3}$$

5.5.5 F1 Score

This is an average Evaluation Metric for calculating a proportion. The Harmonic Mean of the Precision and Recall Evaluation Metrics is also known as the F1 Score. This Evaluation Metric is a measure of our model's overall accuracy in a positive prediction environment—that is, how many of the facts that our model has categorized as positive are truly positive. The formula is:

$$\text{F1 - score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}*\text{Recall}} \tag{5.4}$$

5.5.6 Loss

The error (also known as "the loss") between the result of our algorithms and the supplied goal value is calculated using loss functions.In regression, loss functions are used to determine the optimum fit line by reducing the overall loss of all the points with the line's forecast. During training perceptrons and neural networks, loss functions influence how their weights are updated. The update is proportional to the size of the loss. The accuracy of the model is increased by reducing the loss. However, in these deep transfer learning systems, the tradeoff between update size and low loss must be assessed.

5.5.7 Training and Validation Accuracy

The accuracy estimates obtained after running the model through the training and validation sets are referred to as training accuracy and validation accuracy, accordingly. Training accuracy indicates how successfully the model is trained, while validation accuracy indicates the model's capacity to adjust to fresh datasets, or, in the other words, its reliability. The model exhibits no overfitting although all accuracy are near to identical, but when training accuracy rises above validation accuracy, it displays symptoms of overfitting to training datasets, where it suggests it is fitting to unneeded noise while training.

5.5.8 Training and Validation Loss

The statistic that computes the capability of how efficiently a deep learning model pairs with the training knowledge is known as the training loss. It examines the inaccuracies of a model on the data which were trained. The the dataset that was employed to construct the model originally is a subset of the training set. It is determined statistically by combining all of the errors for every selections in the training set. It needed to be mentioned that after every batch the training loss is counted. The training loss is usually represented by graph drawing.

Validation loss, on the other hand, is a statistic used to evaluate a deep learning model 's efficiency on the validation set. The validation set is a subset of the dataset put aside to test the model 's effectiveness. The validation loss is determined from the total of the mistakes for each sample in the validation set, equal to the training loss. After each epoch, the validation loss is also calculated. This tells us whether the model requires more tuning or tweaks. We normally achieve this by showing a learning curve for the validation loss.

The training and validation losses are frequently graphed simultaneously in most deep learning applications. The goal is to analyze the model's effectiveness and determine which elements need to be tweaked.

5.5.9 K-fold Cross validation

Cross-validation (CV) is a method for evaluating and testing the effectiveness of a deep learning model (or accuracy). It entails conserving a portion of a dataset for which the model has not been trained. The model is then evaluated by testing it on this data.
Cross-validation is a technique for preventing overfitting in models, especially when the quantity of data is small. It's also defined as rotation estimation or out-of-sample testing, and it's most commonly utilized when the model's goal is prediction. There are various types of cross validations. One of the cross validation is K-fold cross validation.
The process includes only one parameter, k, which specifies how many groups a given data sample should be divided into. As a result, the process is frequently referred to as k-fold cross-validation. Whenever a precise value for k is specified, it may be substituted for k in the model's reference, for example, k=10 for 10-fold cross-validation.The k-fold cross-validation method begins by arbitrarily dividing the original dataset into k folds or subgroups. The model is trained on the k-1 subsets of the total dataset in every round. The model is then evaluated on the k-th subset to see how well it performs. This procedure is continued when all k-folds have been used as the evaluation set.The cross-validation accuracy is the mean of the outcomes of every round. Cross-validation accuracy is a statistical indicator that is utilized to evaluate the efficacy of various models.Because each data point in the initial dataset appears in both the training and testing sets, the k-fold cross-validation approach yields less biased models. If we just have a little quantity of data, this strategy is ideal. Nevertheless, because the algorithm must be ran k number of times from the beginning, this procedure may take some time. It implies it takes k-1 times as long to compute as the holdout technique.

Chapter 6

Result

6.1 Accuracy and Loss

6.1.1 Breast Cancer

After all of the models have been trained and tested, we can analyze the findings in order to distinguish between excellent and bad model performance of breast cancer classifiers from raw mammographic image datasets.We ran InceptionV3, VGG16 , MobilenetV2 and Resnet50 models in our chosen dataset MINI-DDSM and KAU-BCMD to obtain the results of our accuracy and loss of these models.

Figure 6.1: Accuracy and Loss graph of VGG16 model on the dataset(Breast Cancer)

The Figure 6.1 shows accuracy and loss in the VGG16 model. In Figure 6.1, the accuracy alternates around 99.9%, and validation accuracy is up to 98.5%. In Figure 6.1, the loss in the training is around 0.0004 and validation loss is around .0860.
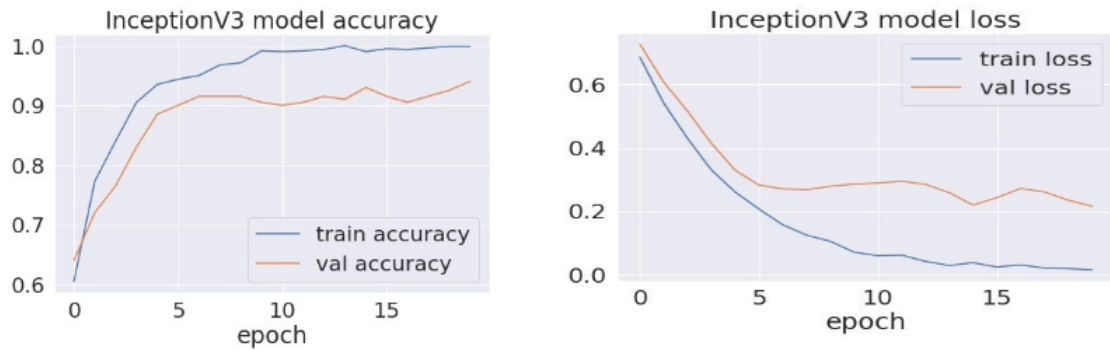
Figure 6.2: Accuracy and loss graph of InceptionV3 model on the dataset(Breast Cancer)

The Figure 6.2 shows accuracy and the loss in the InceptionV3 model. In Figure **??**, the training accuracy is around 99.87% and validation accuracy is around 94%. In Figure 6.2, the loss in the training is about 0.0141 and validation loss is around 0.2154. The Figure 6.3 shows accuracy and the loss in the MobilenetV2 model. In
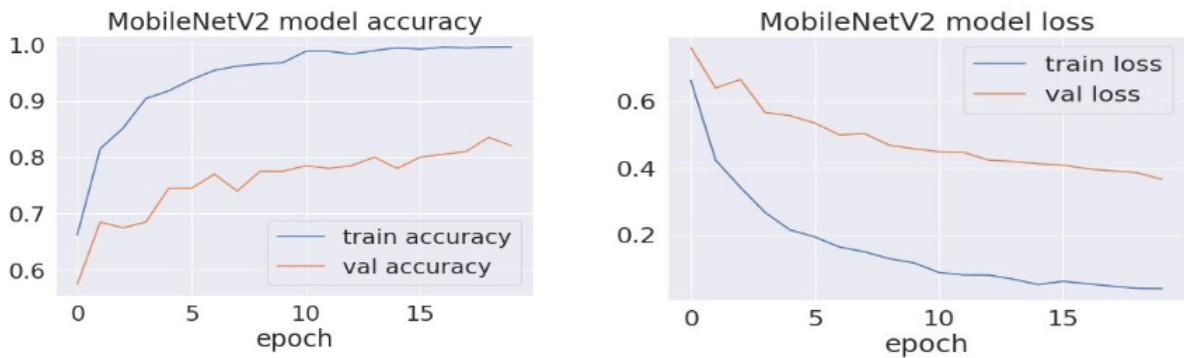

Figure 6.3: Accuracy and loss graph of MobilenetV2 model on the dataset(Breast Cancer)

Figure 6.3, the training accuracy is around 99.9% and validation accuracy is around 69.50%. In Figure 6.3, the loss in the training is about 0.0075 and validation loss is around 0.9186.
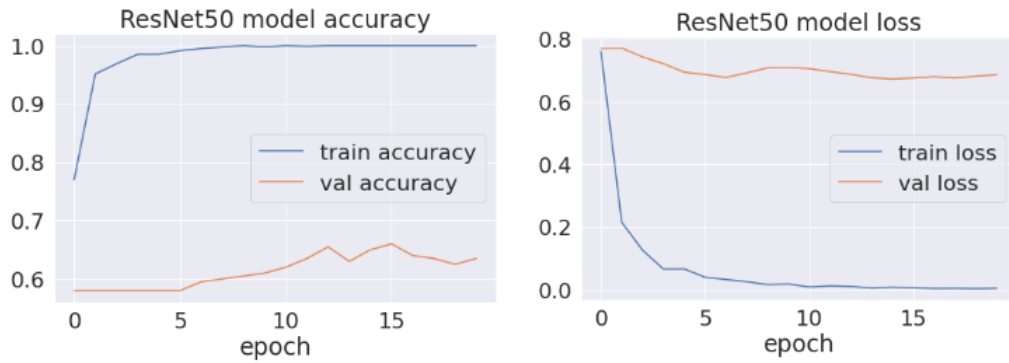
Figure 6.4: Accuracy and loss graph of ResNet50 model on the dataset(Breast Cancer)

The Figure 6.4 shows accuracy and the loss in the Resnet50 model. In Figure 6.4, the training accuracy is around 99.98% and validation accuracy is around 63.50%. In Figure 6.4, the loss in the training is about 0.0038 and validation loss is around 0.6856 .
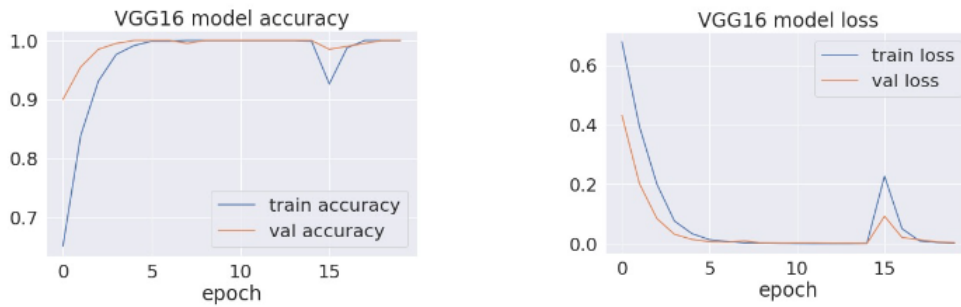
6.1.2 Lung Cancer

Figure 6.5: Accuracy and Loss graph of VGG16 model on the dataset(Lung Cancer)

In VGG16, we got train accuracy around 99.9% and validation accuracy 99.9% shown in the Figure 6.5. In Figure 6.5, the train loss is 0.0026 and the validation loss is 0.0041.
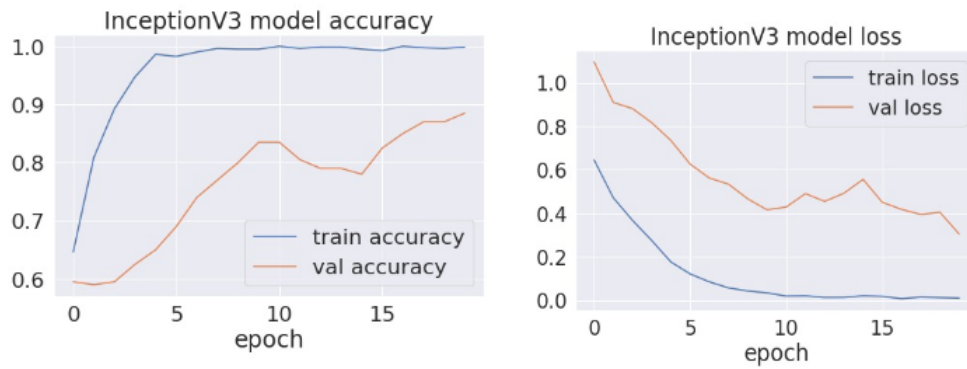
Figure 6.6: Accuracy and Loss graph of Inception V3 model on the dataset(Lung Cancer)

The Figure 6.6, shows Inception V3, we got train accuracy around 99.9% and validation accuracy around 88.5%. Here, the train loss is 0.0105 and validation loss is around 0.304.

Figure 6.7: Accuracy and Loss graph of MobileNet V2 model on the dataset(Lung Cancer)

In Figure 6.7, the train accuracy is around 99.9% and validation accuracy is around 89.5% in MobileNet V2 . Also, the train loss is around 0.002 and validation loss is around 0.267 is shown here.
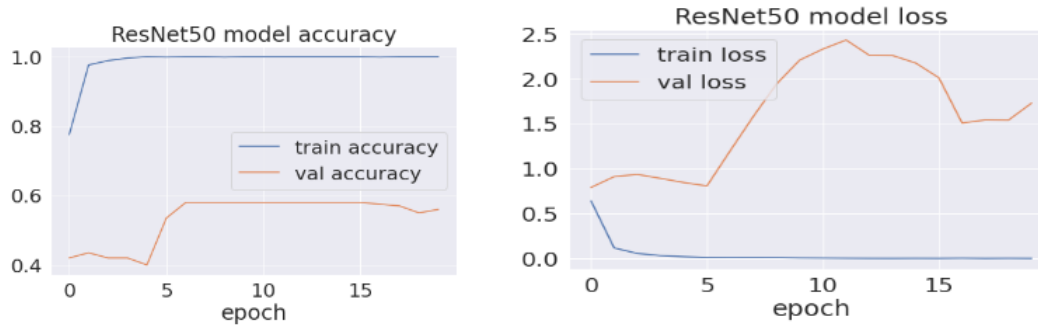
Figure 6.8: Accuracy and Loss graph of ResNet 50 model on the dataset(Lung Cancer)

For ResNet 50, the train accuracy is around 99.9% and validation accuracy is around 56% in the Figure 6.8. The train loss is around 0.0016 and the validation loss is around 1.731 also shown here.

6.1.3 Skin Cancer

We used the 1000 images from the HAM10000 dataset for Skin Cancer . This dataset consists of dermatoscopic images from different populations, acquired and stored by different modalities. VGG16. Resnet50, Inception-V3, Mobilenet-V2 models have been used to run on our chosen dataset .After running all the 4 models on this dataset , we found the following results shown in the graph .
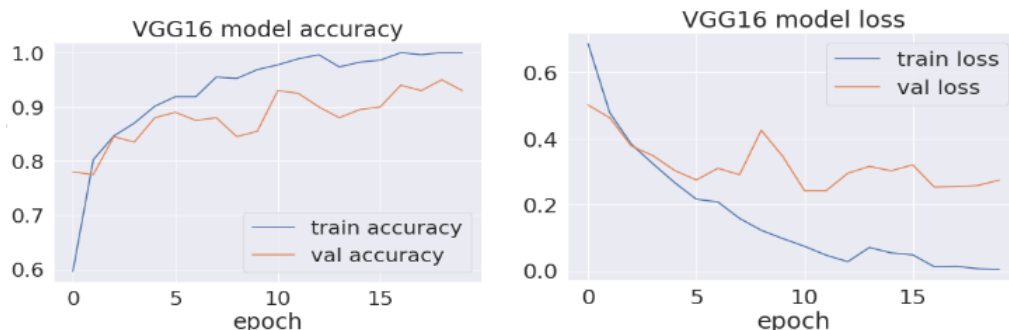

Figure 6.9: Accuracy and Loss graph of VGG16 model on the dataset(Skin Cancer)

Using VGG16, we got train accuracy around 99.9% and validation accuracy 93% shown in the figure 18. In the Figure 6.9, the train loss which is 0.005 while the validation loss 0.27.
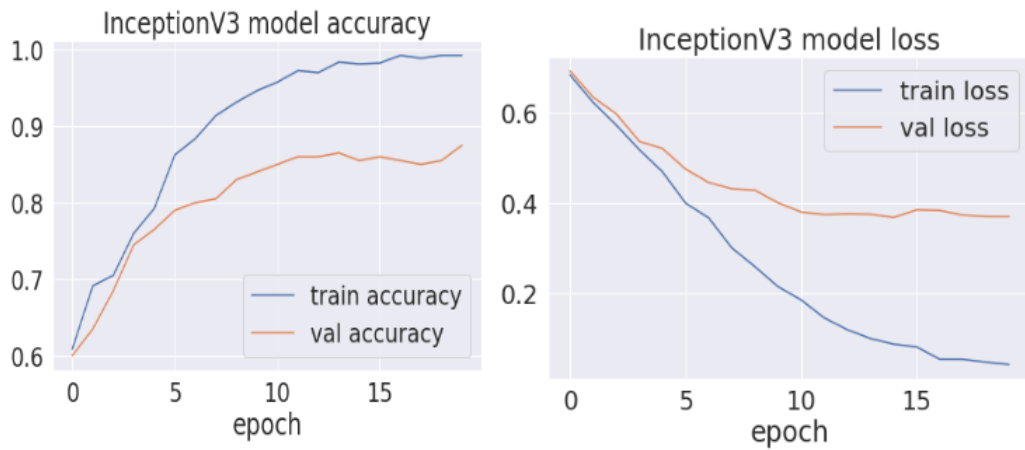
Figure 6.10: Accuracy and Loss graph of Inception V3 model on the dataset(Skin Cancer)

In Inception V3, we got train accuracy around 99.25% and validation accuracy around 87.50%. Here, the train loss is 0.041 and validation loss is 0.37 shown in Figure 6.10.
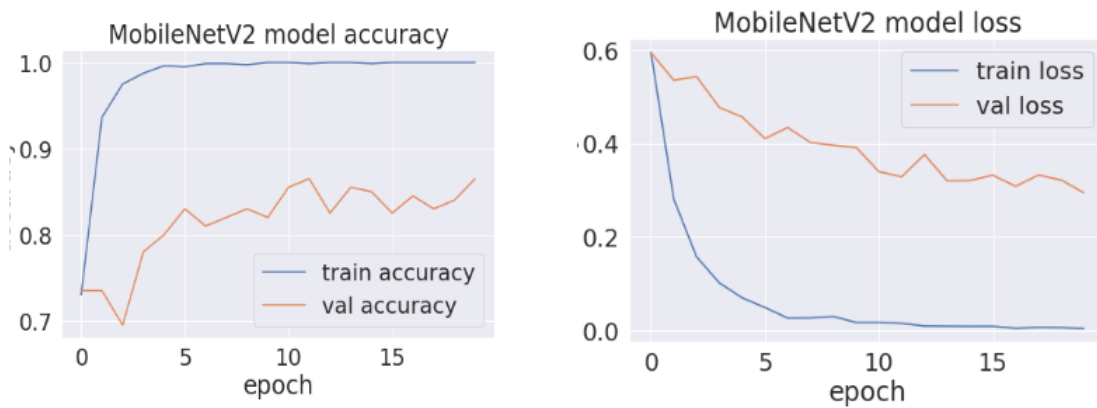

Figure 6.11: Accuracy and Loss graph of MobileNet V2 model on the dataset(Skin Cancer)

In Figure 6.11, the train accuracy is around 99.9% and validation accuracy is around 86.5% in MobileNet V2 . Also, the train loss is around 0.0051 and validation loss is around 0.294 is shown here.
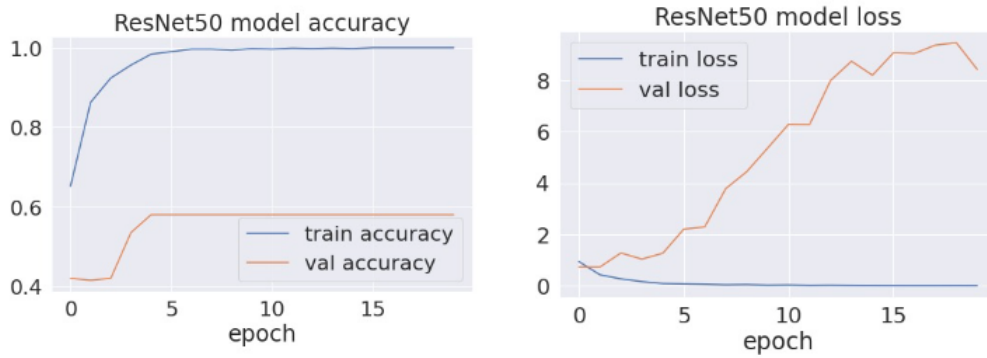
Figure 6.12: Accuracy and Loss graph of ResNet 50 model on the dataset(Skin Cancer)

For ResNet 50, the train accuracy is around 99.9% and validation accuracy is around 58% in the Figure 6.12. The train loss is around 0.0068 and the validation loss is around 8.427 also shown here.

6.2 Confusion Matrix

The confusion matrix is a table that displays a model's performance in each category. The table's rows represent the categories in which the model is trained. Each class reflects the correct and incorrect (True Positive/Negative, False Positive Negative) values correspond to their class. This also summarizes how biased the training process was and how effectively the models can recognize the objective. Here, the following Figure 6.13, Figure 6.14, Figure 6.15 contains the different confusion matrix generated from the different pre-trained models(Vgg16, InceptionV3, MobilenetV2, Resnet50) which we employed for breast, lung and Skin cancer detection in our paper.
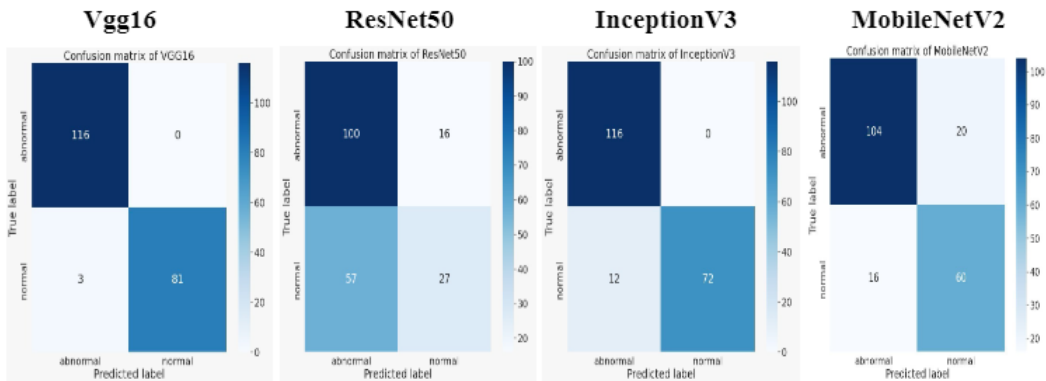
Breast Cancer

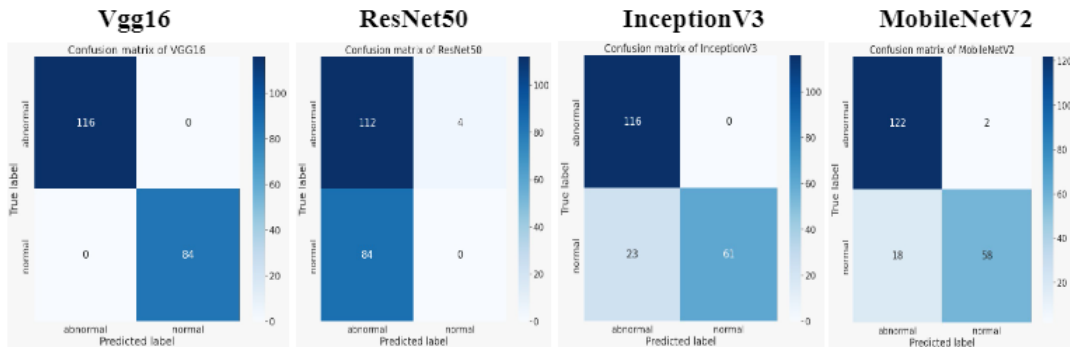Figure 6.13: Confusion Matrix of Breast Cancer

Lung Cancer

Figure 6.14: Confusion Matrix of Lung Cancer

Skin Cancer
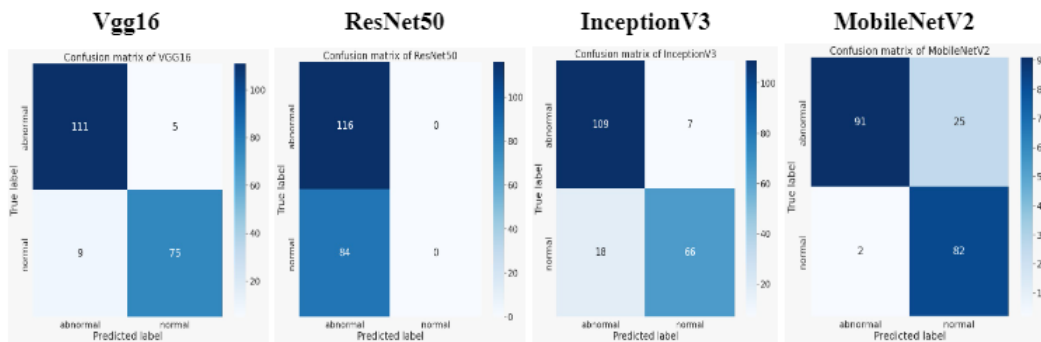
Figure 6.15: Confusion Matrix of Skin Cancer

6.3   Comparison Evalution

6.3.1   Analysis

Our initial purpose was choosing the best deep transfer learning models and examining the performance of the models for small amounts of different cancer datasets. For the small datasets, we have taken a total 1000 images for each cancer. For Breast cancer, we created a mixture of MINI-DDSM and KAU-BCMD datasets and chose 1000 images from it. We again created a composition of IQ-OTH/NCCD and Chest CT-Scan images Dataset to choose 1000 images and make the Lung dataset. From the HAM10000 dataset, we chose 1000 unique images as our Melanoma cancer dataset. Thereafter, we used our newly modified datasets for our selected Transfer Learning models. Hence we are conducting a comparative analysis which presents the models performance. After comparing the accuracy for the cancer detection of predefined models, using the same epoch, Vgg16 performed excellently for the selected separate cancer datasets. At first, for breast cancer, the value of the accuracy

reached 99.8% for training and 98.5% for validation where InceptionV3 delivered almost similar accuracy 99.87% for training and 94% for validation. Although for MobilenetV2 and Resnet50, we got the same training accuracy but the validation accuracy dropped to 69.50% and 63.52%. Therefore, we can consider Vgg16 as the best performing model following InceptionV3. For our lung cancer dataset, we got the training accuracy 99.9% with similar validation accuracy for Vgg16 but for InceptionV3, the training accuracy remained same but validation accuracy delivered 88.5%. Next, for MobilenetV2 and Resnet50 the training accuracy did not change but we got validation accuracy around 89.5% and 56%. So, we can consider Vgg16, MobilenetV2 as suitable models for lung cancer. Though InceptionV3 performed quite well too. Lastly, for skin cancer (melanoma), we got the training accuracy 99.98% for Vgg16 but the validation accuracy delivered 93% in this case. InceptionV3 delivered 87.50% for validation accuracy with the same training accuracy. After that, MobilenetV2 and Resnet50 conveyed 86.5% and 58% respectively.

6.3.2   K-fold cross validation

Since we wanted to do comparative analysis we selected similar amounts of limited datasets. So for the evaluation of unseen data, we applied k cross validation and observed the accuracy for examining it in a better way. As it assures that the result of our model is independent of how we choose our train and test subsets. In this procedure, the data set is divided into k numbers of subsets, and the holdout method is used k times. Here after the application of k fold cross validation we verified our accuracies more precisely. From the following table we can observe all of the accuracy values. Also, the average of the model accuracies gave a better view about the best performing model. Among the values of test accuracies all over, we got best average 96.10% for Vgg16, then 85.46% for InceptionV3. Therefore, it provides the randomized values, it also verifies that Vgg16 is the best suited model for our dataset.

Comparison Table For Test Accuracy								
		With K-Cross Validation (Average)				Without K-Cross Validation		
SI no	Cancer Name	VGG16	RESNET50	INCEPTION V3	MOBILENET V2	VGG16	RESNET50	INCEPTION V3	MOBILENET V2
1	Breast Cancer	97%	43%	91%	73%	98.50%	63.50%	94%	69.50%
2	Lung Cancer	99.40%	52.74%	80.40%	90.79%	99.90%	56%	88.50%	89.50%
3	Melanoma	92%	59.88%	84.99%	90.10%	93%	57.99%	87.50%	86.50%
4	Average	96.10%	51.87%	85%	84.63%	97.13%	59.83%	90%	81.83%

Table 6.1: Table For Test Accuracy Table

6.3.3 Bar Chart

The bar chart below represents the precision, recall and F1-score for breast, lung and skin cancer, shown in Figure 6.16, Figure 6.17, 6.18. It gives us an observation of differences in the separate cancer for separate models.
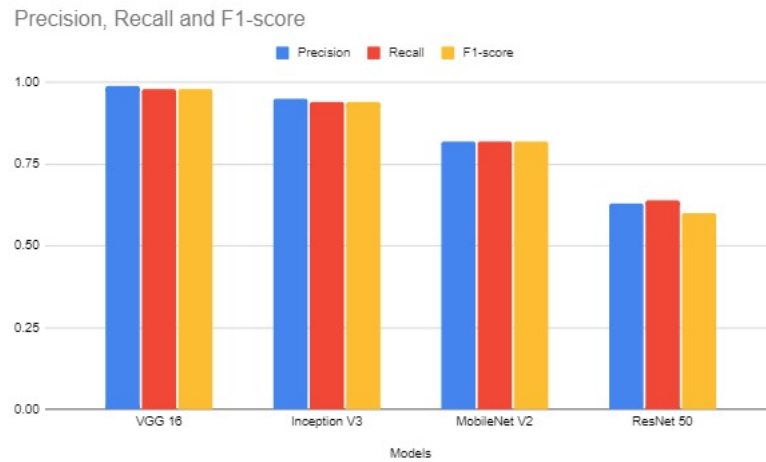

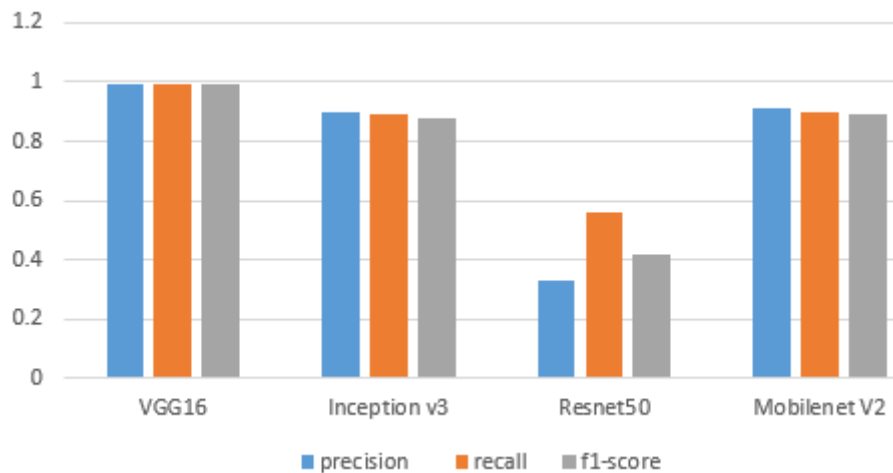Figure 6.16: Precision, recall and F1 score for different models (Breast cancer)

Figure 6.17: Precision, recall and F1 score for different models(Lung cancer)
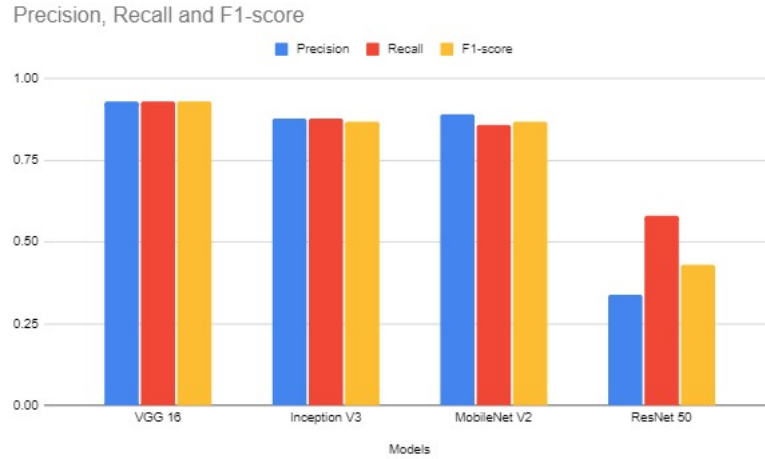
Figure 6.18: Precision, recall and F1 score for different models (Melanoma).

The objective of this specific comparative analysis is to demonstrate that the models' training was done equally for all three classes of cancers with four of our chosen models. We achieved various outcomes in terms of accuracy, recall, and f1-score for different models on different cancers for each case. For all the cases here we can see that Vgg16 works exceptionally well for all of the cases of our selected datasets. For the Vgg16, we got a precision of 0.99 for both Lung and Breast cancers and for the Melanoma, the precision is about 0.93. The highest recall we got for Vgg16, is for Lung cancer which is about 0.99 also. Followed by breast cancer and Melanoma with the recall of 0.98 and 0.93 respectively. Finally, for the F1-score of Vgg16, again the Lung cancer's performance is the best with a result of 0.99. For the Breast and Melanoma, the F1-score we got is about 0.98 and 0.93. As a result, Vgg16 outperformed all of our other models. It moreover employed fewer trainable parameters for our architecture,thus making the Vgg16 model lighter and more adaptable.

6.3.4 Previous work with same dataset

Cancer	Breast Cancer		Lung Cancer		Melanoma	
Used Work Accuracy	[15]	[30]	[37]	[4]	[34]]	[32]
	93.70%	88%	94.61%	90.63%	90%	82%
Our Accuracy	98.50%		99.90%		93%	

Table 6.2: Comparison with Previous Work

In the above chart, we have compared our comparative work with the existing works and can observe different kinds of accuracies. We can also see that our work outperformed the previous works. As we have taken a small amount of data and the existing works have a large amount of data for the models so we can confidently say that our chosen models work best for the small amount of data. In our further approach using these dataset our aim will be using more from these datasets.

Chapter 7

Final Chapter

7.1 Limitations and Future Research

Our technique has demonstrated to be highly effective in terms of multiple datasets. We also offered a thorough overview of existing methods for diagnosing and quick detection of a variety of cancers that have a significant impact on the human body. The purpose of this article is to examine and categorize, compare cancer-related approaches with small datasets, as well as to identify any gaps. Another goal of this study is to provide new researchers with a thorough background in order to begin their research in this sector. However, there are several more areas where we may enhance our performance in this area.

A future development might involve providing additional training data while assuring deep unbiased learning as we employed limited datasets. In terms of augmentation and cross validations, the approach has several significant flaws , which we attempted to adjust with our pre-processing strategy. We expect to see far better results in any other dataset, and we want to continue working on the other datasets in the future to develop a customized model that can be used for multiple cancer detection. Also, one of our future aims is to eliminate the limits and be able to tackle the section where we could not employ data augmentations and k cross validation simultaneously.

To develop a custom model architecture, we may use a clustering algorithm, reorganize the current layers combining the best performances of each of our used existing models, adjust the parameters, and test the epochs using trial and error. K-fold class validation might be improved for better testing validation findings and, if possible, boosting accuracy . For more efficient and accurate results, the pre-processed data can be further enhanced and denoised. We also suggest modifying the loss function and incorporating more up-to-date sequential layers that can enhance our performance.

7.2 Conclusion

The purpose of (TL) transfer learning with convolutional neural networks is to enhance performance on a new project by utilizing prior knowledge of related tasks. It has made a significant impact to medical image evaluation by overcoming the problem of data scarcity while also saving time and physical hardware.

However, the majority of studies have randomly formatted transfer learning. This

research study seeks to offer suggestions and discussions on different transfer learning models and delivers us a proposition on choosing suitable models for specific small datasets from TL methodologies for detecting cancer in an improved way. From previous researches, we have observed some transfer learning models for training neural networks to classify or detect different kinds of cancer. Although cancers may be readily treated if found early, the diagnosis has proven to be difficult. ResNet50, InceptionV3, VGG16, and Mobilenet-V2 were chosen as the final model after we examined them using our assessment metrics. The use of transfer learning in CNN is an effective approach for our selected cancer (Melanoma, Breast and Lung Cancer) detection.The main objective of transfer learning is always to enhance goal learners' achievement by shifting comprehension from all supplemental (but relevant) provevrnience state. Our findings indicate that pre-trained CNN models can automatically extract features from Mammographic, CT scan, and Dermoscopic images, and that a good classifier can really be trained utilizing these features without any need for hand-crafted features. In this analysis, we concentrated on pre-training models, transfer learning approaches,convolutional neural network (CNN) models and data pre-processing as they apply to all the mentioned image recognition and detection. Eventually, various works are compared, and obstacles are discussed.

In conclusion, a thorough evaluation of current state-of-the deep transfer learning aided cancer detection strategies, as well as their comparison of models and limitations. However, there is a detection model which was not suitable enough for our datasets. The accuracy for that model was still far from mature. The majority of the researchers did not utilize benchmark datasets or used tiny sets to evaluate their recommended methods. The present state of the methodologies are evaluated using benchmark datasets using separate models for this purpose, and the limits of existing strategies are emphasized.

Bibliography

[1] Y. LeCun, B. Boser, J. Denker, *et al.*, "Handwritten digit recognition with a back-propagation network," *Advances in neural information processing systems*, vol. 2, 1989.

[2] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[4] E. Dandıl, M. Çakiroğlu, Z. Ekşi, M. Özkan, Ö. K. Kurt, and A. Canan, "Artificial neural network-based classification system for lung nodules on computed tomography scans," in *2014 6th International conference of soft computing and pattern recognition (SoCPaR)*, Ieee, 2014, pp. 382–386.

[5] S. M. Friedewald, E. A. Rafferty, S. L. Rose, *et al.*, "Breast cancer screening using tomosynthesis in combination with digital mammography," *Jama*, vol. 311, no. 24, pp. 2499–2507, 2014.

[6] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[7] C.-K. Shie, C.-H. Chuang, C.-N. Chou, M.-H. Wu, and E. Y. Chang, "Transfer representation learning for medical image analysis," in *2015 37th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 711–714.

[8] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, "Multisource transfer learning with convolutional neural networks for lung pattern analysis," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 76–84, 2016.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning convolutional neural networks for resource efficient transfer learning," *arXiv preprint arXiv:1611.06440*, vol. 3, 2016.

[11] J. J. Näppi, T. Hironaka, D. Regge, and H. Yoshida, "Deep transfer learning of virtual endoluminal views for the detection of polyps in ct colonography," in *Medical imaging 2016: computer-aided diagnosis*, SPIE, vol. 9785, 2016, pp. 590–597.

[12] H.-C. Shin, H. R. Roth, M. Gao, *et al.*, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[13] H. Yang, H. Yu, and G. Wang, "Deep learning for the classification of lung nodules," *arXiv preprint arXiv:1611.06651*, 2016.

[14] L. Fan, Z. Xia, X. Zhang, and X. Feng, "Lung nodule detection based on 3d convolutional neural networks," in *2017 International conference on the frontiers and advances in data science (FADS)*, IEEE, 2017, pp. 7–10.

[15] S. Guan and M. Loew, "Breast cancer detection using transfer learning in convolutional neural networks," in *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, IEEE, 2017, pp. 1–8.

[16] S. Hussein, K. Cao, Q. Song, and U. Bagci, "Risk stratification of lung nodules using 3d cnn-based multi-task learning," in *International conference on information processing in medical imaging*, Springer, 2017, pp. 249–260.

[17] H. Jiang, H. Ma, W. Qian, M. Gao, and Y. Li, "An automatic detection system of lung nodule based on multigroup patch-based deep learning network," *IEEE journal of biomedical and health informatics*, vol. 22, no. 4, pp. 1227–1237, 2017.

[18] G. Litjens, T. Kooi, B. E. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[19] A. Rosebrock, "Imagenet: Vggnet, resnet, inception, and xception with keras," *Mars*, 2017.

[20] L. Shen, "End-to-end training for whole image breast cancer diagnosis using an all convolutional design," *arXiv preprint arXiv:1711.05775*, 2017.

[21] G. Tetteh, M. Rempfler, B. Menze, and C. Zimmer, "Deep-fext: Deep feature extraction for vessel segmentation and centerline prediction," Apr. 2017.

[22] Y. Xue, S. Chen, J. Qin, Y. Liu, B. Huang, and H. Chen, "Application of deep learning in automated analysis of molecular images in cancer: A survey," *Contrast media & molecular imaging*, vol. 2017, 2017.

[23] R. V. M. Da Nóbrega, S. A. Peixoto, S. P. P. da Silva, and P. P. Rebouças Filho, "Lung nodule classification via deep transfer learning in ct lung images," in *2018 IEEE 31st international symposium on computer-based medical systems (CBMS)*, IEEE, 2018, pp. 244–249.

[24] H. Li, P. Ghamisi, U. Soergel, and X. X. Zhu, "Hyperspectral and lidar fusion using deep three-stream convolutional neural networks," *Remote Sensing*, vol. 10, no. 10, p. 1649, 2018.

[25] J. Lyu and S. H. Ling, "Using multi-level convolutional neural network for classification of lung nodules on ct images," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2018, pp. 686–689.

[26] S. J. Mambou, P. Maresova, O. Krejcar, A. Selamat, and K. Kuca, "Breast cancer detection using infrared thermal imaging and a deep learning model," *Sensors*, vol. 18, no. 9, p. 2799, 2018.

[27] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data*, vol. 5, no. 1, pp. 1–9, 2018.

[28] A. Fedorov, M. Hancock, D. Clunie, *et al.*, "Standardized representation of the lidc annotations using dicom," PeerJ Preprints, Tech. Rep., 2019.

[29] S. T. Krishna and H. K. Kalluri, "Deep learning and transfer learning approaches for image classification," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 7, no. 5S4, pp. 427–432, 2019.

[30] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *PeerJ*, vol. 7, e6201, 2019.

[31] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, "Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," in *2020 5th international conference on advanced technologies for signal and image processing (ATSIP)*, IEEE, 2020, pp. 1–5.

[32] H. Gupta, H. Bhatia, D. Giri, R. Saxena, and R. Singh, "Comparison and analysis of skin lesion on pretrained architectures," 2020.

[33] F. Hake, M. Hermann, H. Alkhatib, *et al.*, "Damage detection for port infrastructure by means of machine-learning-algorithms," Jan. 2020.

[34] H. K. Kondaveeti and P. Edupuganti, "Skin cancer classification using transfer learning," in *2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI)*, IEEE, 2020, pp. 1–4.

[35] C. D. Lekamlage, F. Afzal, E. Westerberg, and A. Cheddad, "Mini-ddsm: Mammography-based automatic age estimation," in *2020 3rd International Conference on Digital Medicine and Image Processing*, 2020, pp. 1–6.

[36] P. Wang, J. Wang, Y. Li, L. Li, and H. Zhang, "Adaptive pruning of transfer learned deep convolutional neural network for classification of cervical pap smear images," *IEEE Access*, vol. 8, pp. 50 674–50 683, 2020.

[37] H. F. Al-Yasriy, M. S. AL-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Diagnosis of lung cancer based on ct scans using cnn," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 928, 2020, p. 022 035.

[38] A. S. Alsolami, W. Shalash, W. Alsaggaf, S. Ashoor, H. Refaat, and M. El-mogy, "King abdulaziz university breast cancer mammogram dataset (kau-bcmd)," *Data*, vol. 6, no. 11, p. 111, 2021.

[39] L. Alzubaidi, J. Zhang, A. J. Humaidi, *et al.*, "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, no. 1, pp. 1–74, 2021.

[40] H. F. Kareem, M. S. AL-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Evaluation of svm performance in the detection of lung cancer in marked ct scan dataset," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3, pp. 1731–1738, 2021.

[41] U. A. Usmani, J. Watada, J. Jaafar, I. A. Aziz, and A. Roy, "A reinforcement learning algorithm for automated detection of skin lesions," *Applied Sciences*, vol. 11, no. 20, p. 9367, 2021.

[42] . S.]Ioffe S. "Cancer detection and diagnosis." (), [Online]. Available: Cancerquest. org..

[43] "A guide to resnet, inception v3, and squeezenet. paperspace blog." (), [Online]. Available: https://blog.paperspace.com/popular-deep-learning-architectures-resnet-inceptionv3-squeezenet/. (accessed: (2020, June 5).

[44] "Advanced guide to inception v3. (n.d.)." (), [Online]. Available: https://cloud.google.com/tpu/docs/inception-v3-advanced. (accessed: January 18, 2022).

[45] "Asymmetric convolutions." (), [Online]. Available: https://hackmd.io/@machine-learning/SkD5Xd4DL. (accessed: (27-May-2022).

[46] "Auxiliary classifier." (), [Online]. Available: https://ieee.nitk.ac.in/virtual-expo/image-captioning/. (accessed: (27-May-2022).

[47] "Auxiliary classifier." (), [Online]. Available: https://laptrinhx.com/face-recognition-using-transfer-learning-and-vgg16-3833175636/. (accessed: (27-May-2022).

[48] "Basic cnn architecture: Explaining 5 layers of convolutional neural network." (), [Online]. Available: https://www.upgrad.com/blog/basic-cnn-architecture/. (accessed: 11-May-2022).

[49] J. Brownlee. "A gentle introduction to transfer learning for deep learning." (), [Online]. Available: https://machinelearningmastery.com/transfer-learning-for-deep-learning//. (accessed: (2021, November 20).

[50] "Convolutional neural network." (), [Online]. Available: https://developersbreach.com/convolution-neural-network-deep-learning/.. (accessed: 11-May-2022).

[51] "Detailed guide to understand and implement resnets." (), [Online]. Available: https://cv-tricks.com/keras/understand-implement-resnets/.. (accessed: (12-May-2022).

[52] N. Donges. "Gradient descent: An introduction to 1 of machine learning's most popular algorithms." (), [Online]. Available: https://builtin.com/data-science/gradient-descent. (accessed: 11-May-2022).

[53] M. Hany. "Chest ct-scan images dataset." (), [Online]. Available: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images. (accessed: 2021).

[54] "Intuition of adam optimizer." (), [Online]. Available: https://www.geeksforgeeks. org/intuition-of-adam-optimizer/#:~:text=Adam%5C%20optimizer%5C% 20involves%5C%20a%5C%20combination,minima%5C%20in%5C%20a%5C% 20faster%5C%20pace.. (accessed: 11-May-2022).

[55] S. Ioffe and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." (), [Online]. Available: https: //proceedings.mlr.press/v37/ioffe15.html. (accessed: 01.09.2021).

[56] "Isicarchive.com." (), [Online]. Available: https://www.isicarchive.com.. (accessed: 2-Oct-2021).

[57] "Keras tutorial: Using pre-trained imagenet models." (), [Online]. Available: https://learnopencv.com/keras-tutorial-using-pre-trained-imagenet-models/ .. (Accessed: 12-May-2022).

[58] "Lidc-idri.databas." (), [Online]. Available: https://wiki.cancerimagingarchive. net/display/Public/LIDC-IDRI. (accessed: 11-May-2022).

[59] "Melanoma skin cancer statistics. (n.d.)" (), [Online]. Available: https:// www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html. (accessed: October 2, 2021).

[60] "Mobilenetv2." (), [Online]. Available: https://paperswithcode.com/method/ mobilenetv2.. (accessed: (12-May-2022).

[61] C. Pere. "What are loss functions?," towards data science." (), [Online]. Available: https://towardsdatascience.com/what-is-loss-function-1e2605aeb904. (accessed: 11-May-2022).

[62] J.-L. Queguiner. "What does training neural networks mean?" (), [Online]. Available: https://blog.ovhcloud.com/what-does-training-neural-networks-mean/. (accessed: 11-May-2022).

[63] "Robbins cotran pathologic basis of disease e-book." (), [Online]. Available: www.elsevier.com.. (accessed: 2022).

[64] "The difference between digital film mammography." (), [Online]. Available: https://www.seattlecca.org/blog/2012/10/the-difference-between-digital-film-mammography. (accessed: 13-May-2022).

[65] "Understanding categorical cross-entropy loss, binary cross-entropy loss, softmax loss, logistic loss, focal loss and all those confusing names." (), [Online]. Available: https://gombru.github.io/2018/05/23/cross_entropy_loss/. (accessed: 11-May-2022).

[66] "Understanding loss functions in machine learning." (), [Online]. Available: https://www.section.io/engineering-education/understanding-loss-functions-in-machine-learning/. (accessed: 11-May-2022).

[67] "Update." (), [Online]. Available: https://www.uptodate.com/contents/ overview-of-dermoscopy.. (accessed: 13-May-2022).

[68] "Vgg16 - convolutional network for classification and detection." (), [Online]. Available: https://neurohive.io/en/popular-networks/vgg16/. (accessed: (2018, November 20).

[69] "What is overfitting?" (), [Online]. Available: https://www.ibm.com/cloud/ learn/overfitting. (accessed: 12-May-2022).

[70] "Who should be screened for lung cancer?" (), [Online]. Available: https: //www.cdc.gov/cancer/lung/basic_info/screening.htm.. (accessed: 13-May-2022).