Classification of Melanoma Skin Cancer Based on Image Data Set Using Different Neural Networks

By

Rukhsar Sabir

00000364989

Supervisor

Dr. Tahir Mehmood

School of Natural Sciences (SNS)

A thesis submitted in partial fulfillment of the requirements for the degree
of

In
School of Natural Sciences,
National University of Sciences and Technology (NUST),
Islamabad, Pakistan.

(August 2023)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by **Rukhsar Sabir** (Registration No. **00000364989**), of **School of Natural Sciences** has been vetted by undersigned, found complete in all respects as per NUST statutes/regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/M.Phil degree. It is further certified that necessary amendments as pointed out by GEC members and external examiner of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor: Prof. Tahir Mahmood

Date: _____

Signature (HoD): _____

Date: 04/09/2023

Signature (Dean/Principal): _____

Date: 04.09.2023

National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: **"Rukhsar Sabir"** Regn No. <u>00000364989</u> Titled: **"Classification of Melanoma Skin Cancer Image Data Set Using Different Neural Network"** accepted in partial fulfillment of the requirements for the award of **MS** degree.

Examination Committee Members

1. Name: <u>DR. ZAMIR HUSSAIN</u>   Signature:_____

2. Name: <u>DR. SHAKEEL AHMED</u>   Signature:_____

Supervisor's Name: <u>PROF. TAHIR MAHMOOD</u>   Signature:_____

_____
Head of Department

<u>04/09/2023</u>
Date

COUNTERSINGED

Date: <u>04.09.2023</u>

_____
Dean/Principal

Abstract

Melanoma is an aggressive skin cancer type that is incredibly terrifying because of its tendency to become prevalent across the body if not detected and treated on time. In the field of medical image diagnosis, computer vision can play a significant role, as shown by several existing technologies. In this paper, we present neural network models such as basic CNN, ResNet-18, and EfficientNet-B0 for image processing in melanoma skin cancer detection. This data set has binary classes such as benign and malignant with a 10605 sample size, where 9605 images for training, and 1000 images for testing the model's performance. The process of segmentation, extracting features, classification, and pre-processing process are the processes that were used in this study. A classification with a success of 97% accuracy was produced by EfficientNet-B0, which outperformed 87% from ResNet-18 and 80% from CNN for the classification of malignant and benign. According to other evaluation performances such as sensitivity, specificity, f1-score, precision, error rate, Mathew's correlation coefficient, geometric mean, and bookmaker informedness, EfficientNet-B0 outperforms ResNet-18 and CNN. The findings of this research indicate Neural Network models specifically EfficientNet-B0 show significant potential for accurate and efficient melanoma skin cancer detection to save lives.

Dedication

This thesis is dedicated to:
For the sake of Allah, my Creator and my Master,
My great teacher and messenger, Mohammed (PBUH), who
taught us the purpose of life,
The great martyrs and prisoners, the symbol of sacrifice;
The NUST University; my second magnificent home;
My parents, who never stop giving of themselves in countless
ways,
My dearest Uncle, who leads me through the valley of darkness
with light of hope and support,
My friends who encourage and support me,
All the people in my life who touch my heart,
I dedicate this research.

Certificate of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SNS or at any other educational institute, except where due acknowledgment has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SNS or elsewhere is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics which has been acknowledged.

Author Name: **Rukhsar Sabir**

Signature: _____

Acknowledgment

In the name of Allah, the Most Gracious and the Most Merciful. All praise and thanks are due to Allah alone, who has bestowed His countless blessings upon me and has guided me through this academic journey. His infinite wisdom and mercy have enabled me to overcome challenges, reach this academic milestone, and complete this thesis. I would like to express my deep appreciation to my supervisor, Dr. Tahir Mehmood, whose support, guidance, and encouragement have been invaluable throughout this journey. His patience, dedication, and expertise have helped me navigate the complexities of this research, and I am grateful for his unwavering commitment to my success. To my dear family, thank you for your unwavering love, support, and understanding. Your encouragement, prayers, and sacrifices have been instrumental in helping me overcome challenges and stay focused on my goals. To my cherished friends, thank you for your love, and support, and for keeping me grounded during this journey. Your companionship, humor, and unwavering support have been a source of comfort and strength. Finally, I would like to express my heartfelt appreciation to all those who have contributed in any way to the completion of this thesis. Your feedback, support, and encouragement have been invaluable, and I am grateful for your help and contributions. Once again, I thank Almighty Allah for His blessings and pray that He continues to guide and bless me in all my future endeavors.

Table of Contents

List of Figures

List of Tables

List of Symbols

Nomenclature

Abbreviations

ML	Machine learning
DL	Deep Learning
NN	Neural Network
DNN	Deep Neural Network
CNN	Convolutional Neural network
RNN	Residual Neural Network
ENN	Efficient Neural Network
SOM	Self Organizing Map
MA	Moving Average
SE	Squeeze - Excitation
GAN	Generative Adversarial Network

Chapter 1

Introduction

1.1 Medical Imaging & its Applications

Medical image processing can aid in the early diagnosis, care, and detection of disorders. [1]. Medical imaging utilizes various technologies to visualize internal organs and structures, aiding diagnosis and monitoring. However, as medical image analysis expands, organ segmentation and abnormality identification become more complex. The classification of medical images aids in determining the number of medications and radiation exposure, as well as limiting the growth of conditions like tumors [2]. It is a method and procedure for capturing images of the inside of a body for use in clinical research, medical treatment, and the physiology of certain organs and tissues. It is now crucial for both medical diagnosis and treatment. Due to doctors' interest in studying internal anatomy, these images are crucial in medical applications [3]. Computer vision, pattern recognition, image mining, and machine learning are now all parts of medical image processing [4]. Due to their effectiveness, neural networks provide answers to a number of image identification issues, and they are currently being used in the medical field [5]. As a result of recent developments in computer vision, the field of medical imaging has also seen an increase in interest in Transformers, which can capture global context as opposed to CNNs with local receptive fields [6]. The enhanced outcomes of modern medicine largely depend on medical imaging. X-rays, MRIs, ultrasounds, endoscopies, tactile imaging, computerized tomography (CT scan), and other imaging techniques are some examples of medical imaging methods.
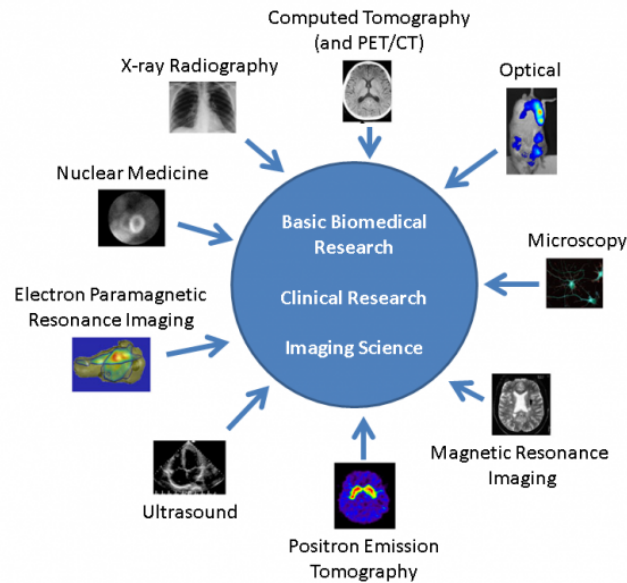
Figure 1.1: Several types of Medical Imaging

Medical imaging refers to the use of various imaging techniques for the purpose of clinical investigation and medical intervention, to produce visual representations of a body's inside. Numerous techniques have been developed, such as those based on cross-sectional and X-ray images (SPECT, PET, or ultrasound), as well as tomographic modalities like computed tomography (CT), magnetic resonance imaging (MRI), and others. The challenging and crucial step of image processing is segmentation. In the area of image interpretation, it has grown in popularity [7]. Methods for segmenting images are currently improving more quickly and precisely. We are discovering a broad segmentation algorithm that can be used to segment different types of photos by merging numerous new theories and technology [8]. To segment tissues and human organs, a variety of image segmentation techniques have been applied in medical applications. Applications include mass detection in mammograms, automated blood cell classification, border detection in coronary angiograms, surgical planning, simulation of surgeries, tumor detection and segmentation, brain development study, functional mapping, border detection in angiograms, image registration, heart segmentation, and analysis of cardiac images, among others. Segmentation is a technique used in the field of medicine to separate different tissues by extracting and categorizing features. Organizing visual pixels into anatomical sections can assist people recognize bones, muscles, and blood vessels. Additionally, by using MRI pictures, this approach can be used to identify the characteristics of breast tumors and extract information about skin cancer from images.

1.1.1 Machine Learning in Medical Imaging

Machine learning models can accurately identify diseases by analyzing medical images, aiding in early detection and diagnosis [9]. It is an effective method for finding patterns in medical photos, but it should be used with caution because it can be abused if the technology's advantages and disadvantages are not understood. [10]. It has made significant strides in medical imaging, revolutionizing how medical professionals analyze and interpret images for diagnostic and treatment purposes. Through the use of machine learning approaches, quantitative characteristics (radiomics) may be extracted from medical pictures, allowing for a more thorough assessment of tissues and lesions [11]. Machine learning assists in aligning and fusing multiple medical images from different modalities, aiding in improved visualization and diagnosis [12]. Many medical imaging modalities produce volumetric data (e.g., CT and MRI), and applying machine learning to 3D data presents unique challenges and opportunities [13]. Integrating information from multiple imaging modalities such as combining MRI and PET scans, using machine learning can offer a more thorough understanding of the patient's situation [14]. Medical images are often subject to noise and artifacts, making it difficult to interpret the underlying information. Machine learning algorithms can be used to de-noise medical images, improving their clarity and facilitating more accurate diagnoses. Ensuring transparency and interpretability of AI models is crucial in medical imaging to gain the trust of healthcare professionals [15].

1.1.2 Neural Networks in Medical Imaging

Neural Networks, particularly due to their capacity to extract subtle patterns from large amounts of complex picture data, convolutional neural networks (CNNs) have become effective tools in Health Imaging. They excel in tasks like image segmentation, disease detection, and feature extraction, contributing to enhanced diagnostic accuracy and improved patient care [16]. More recently, the use of neural networks has begun. This

method has the advantage that characteristics are recognized as part of the learning process rather than having to be calculated and identified as an initial step. Specific pertinent areas of interest in medical imaging are made possible by neural network attention mechanisms, which improves interpretability [17]. Combining information from multiple imaging modalities using neural networks provides a comprehensive understanding of medical conditions [18]. GANs are used to generate realistic medical images, augment datasets, and improve image quality, aiding in data scarcity issues [19].

Three-dimensional convolutional neural networks (3D CNNs) are employed for tasks like brain lesion detection and segmentation in MRI volumes [20]. The accuracy of subsequent studies is increased by using autoencoders to improve the quality and remove noise from medical images [21]. RNNs are used to analyze medical data of a time series, such as electrocardiograms (ECGs) and electroencephalograms (EEGs) [22]. For image segmentation tasks, including locating tumor spots in medical imaging, U-Net topologies are frequently utilized [23]. LSTM networks are applied to predict disease progression and outcomes based on longitudinal medical data [24]. For Automation, CNNs can automate the analysis of medical images, reducing the need for manual interpretation and reducing inter-observer variability. For High-throughput analysis, CNNs can process large amounts of data quickly, enabling the analysis of multiple images in real time. For Multi-modal analysis, CNNs can be trained to simultaneously analyze multiple modalities, such as MRI and CT scans.

1.2   Skin Cancer & its Types

One of the biggest healthcare costs in the world is cancer. According to figures from around the world, there will be roughly 10.0 million cancer-related deaths in 2020 (9.9 million if non-melanoma skin cancer is excluded) [25]. Even though it is typically associated with greater rates of morbidity and mortality, complexion cancer is less common in persons with darker skin tones than it is in light-skinned Caucasians. To increase the likelihood of early diagnosis of malignant cancers, doctors must become knowledgeable about skin cancer in people of color [26]. The fifth most frequently reported disease in the world is now skin cancer, which has a negative impact on both the economic and public health [27]. There are two types of skin cancer melanoma skin cancer (MSC) and non-melanoma skin cancer (NMSC)—can be simply distinguished [28]. The greatest preventable cause of skin cancer is excessive exposure to UV radiation from natural sources like the sun or artificial sources like tanning beds. The World Health Organization (WHO)'s International Agency for Research on Cancer (IARC) estimates that there were 8.2 million cancer-related deaths in 2012 and that there would be 27 million additional instances of the disease by 2030 [29]. The inner dermis layer and the outer epidermis layer make up the two primary layers of skin. Skin cancer first manifests in the epidermis, which is made up of three main cell types. Squamous cells, or thin, flat cells, make up the top layer of the epidermis. Round cells known as basal cells and melanocytes are found underneath the squamous cells. The epidermis's base layer contains the cells that make melanin. Melanin is the name of the pigment that gives skin its color. When the skin is exposed to the sun, melanocytes create more pigment, darkening the skin. The following skin cancer subtypes are.
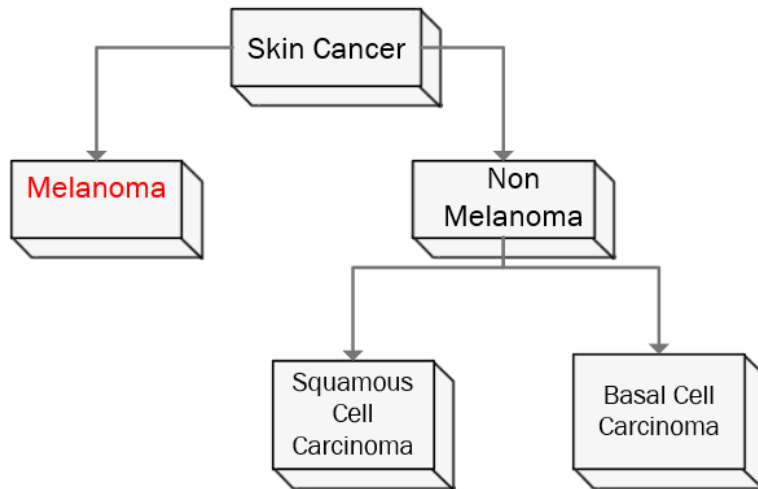
Figure 1.2: Types of Skin Cancer

1.2.1 Non-Melanoma Skin Cancer

The most prevalent cancer in Caucasians is NMSC, and each year, the incidence of this disease rises [30]. A subset of tumors known as non-melanoma skin cancer (NMSC) arise in skin cells other than melanocytes, which are in charge of producing melanin, the pigment that gives skin its color. Basal cell carcinoma (BCC) and squamous cell carcinoma are two examples of the rare malignancies that make up NMSC. The two most frequent types of non-melanoma skin cancer are basal cell carcinoma (BCC) and squamous cell carcinoma (SCC).

Basal cell carcinoma

Skin cancer of the most prevalent kind is basal cell carcinoma (BCC). The face, neck, and scalp are typical sun-exposed locations where it develops. BCC normally grows slowly and seldom metastasizes to other body regions [31]. It develops from basal cells, which are found in the top layer of the skin, the epidermis, in the lower layer. Though it can happen anywhere, BCC generally appears on parts of the body that are frequently exposed to the sun, such as the hands, neck, and face.

Squamous cell carcinoma

Skin cancer squamous cell carcinoma (SCC) is another prevalent kind. It can grow more quickly than BCC and also appears in places that are exposed to light. If SCC is not properly treated, it has a higher chance of spreading to neighboring lymph nodes or other organs [32]. Squamous cells, which are located in the upper layers of the epidermis, the skin's outermost layer, are where it starts.

1.2.2 Melanoma Skin Cancer

Melanoma is the fifth most widespread and deadliest kind of skin cancer in the UK which is most likely either malignant, grow and spread, or benign which is grow but not spread [33]. Melanocytes, the skin's pigment-producing cells, are the source of melanoma, a more severe form of skin cancer. Anywhere on the body, including areas that are not exposed to the sun, might develop it. If melanoma is not found and treated early, it has a high likelihood of spreading to other organs [34]. The Centres for Disease Control and Prevention's cancer statistics show that only 22.1 out of 100,000 people in the US are affected by melanoma, a cancerous tumor that arises from melanocytes. [35]. Despite only accounting for 4% of all skin cancers, it causes 75% of skin cancer mortality. [36]. Usually affecting the back, arms, legs, and face, it spreads quickly and becomes more quickly. UV radiation exposure, numerous moles, fair skin that is prone to burning easily, and a family history of the condition are risk factors. Physicians have a lot of work to do when segmentation is done manually, which could result in prejudice if medical viewpoints are present. After reviewing intricate photos, which takes time, doctors frequently arrive at a joint diagnosis [37]. It gets more challenging to cure and can even be fatal once it has spread deeper into the skin or other parts of the body. Despite having a far lower prevalence than NMSCs, it has been on the rise among populations with fair skin for many years [38].
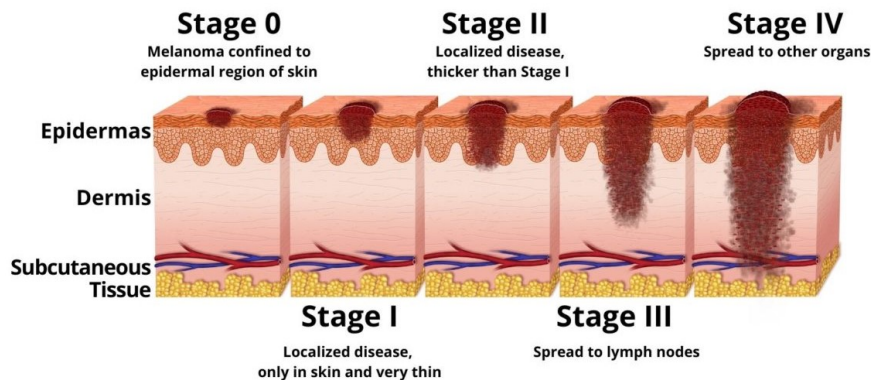
Figure 1.3: Stages of Melanoma Skin cancer

According to estimates, people in the United States with early-detected melanoma have a five-year survival rate of roughly 99%. In the U.S., melanoma is expected to claim the lives of 7,990 people in 2021 (5,420 men and 2,570 women), and 186,680 new cases will be detected there. The top layer of skin, the epidermis, will be the site of 89,070 noninvasive instances, while the dermis, the second layer of skin, will be the site of 97,610 invasive cases. 39,490 of the invasive cases will affect women and 58,120 will affect men. Women (56.7%) were affected by MM of the skin in 1784 incident instances more often than men (43.3%). In comparison to women, men had a mean age at diagnosis that was 17 years higher (566 vs. 549 years, P < 005). For the TNM code, a sizable amount of information that was categorized as "missing" or "unknown" was retrieved. TX led the T category with 44.1 percent, followed by the T1 stage with 30.9 percent [39].

Symptoms of Melanoma Skin Cancer

This form of cancer typically appears on the lower legs in females. On skin that hasn't been exposed to the sun, melanoma can develop in both men and women. Any skin tone can be impacted by melanoma. Melanoma commonly develops on the hands, feet, or under the fingernails or toenails in those with darker skin tones. Symptoms of melanoma include A huge brownish spot with darker flecks, a mole that fluctuates in size, appearance, or feels, or one that bleeds, a little lesion with an unruly border, and areas that are red, pink, white, blue, or blue-black, a hurtful, itchy, or burning lesion, The mucous membranes lining your mouth, nose, vagina, or anus, as well as any dark lesions on your hands, feet, fingertips, or palms. A new mole or a change in an existing mole is frequently how it manifests. Keep an eye out for moles with uneven borders, varying hues within a single mole, or a diameter higher than 6 millimeters. Melanoma lesions can be asymmetrical, indicating that one half is different from the other in terms of size, color, or shape. Shades of brown, black, blue, red, or white, as well as other hues, can all be contained within them. Instead of having a smooth and distinct border, a melanoma lesion may have margins that are ragged, fuzzy, or notched.

A warning indicator could be an uneven distribution of color or the occurrence of different colors within the same mole. Melanomas typically have a diameter bigger than 6 millimeters, or the size of a pencil eraser, compared to regular moles. It's crucial to remember that melanomas can come in smaller sizes as well. A mole or skin lesion should be examined by a medical practitioner if it changes in size, shape, color, elevation, or symptoms (such as itching, bleeding, or crusting).

1.3 Classification

An example of a supervised learning problem is classification, where the algorithm learns from a collection of samples that have already been assigned to the appropriate classes. The objective is to create a model that can extrapolate from the training data to generate precise predictions on fresh, unstudied data. Naive Bayes, Decision Trees, Random Forests, Support Vector Machines (SVM), and Neural Networks are a few common categorization techniques [40]. With a proven promise of ongoing applicability in the real-world environment, it has substantially improved the research paradigm and spectrum. specifically in the areas of healthcare, security, education, gaming, robotics, finance, and autonomous systems [41]. Usually, the integrity of the input-data representation has a significant impact on how well an ML algorithm performs. It has been demonstrated that a good data representation outperforms a bad one in terms of performance. As a result, for many years, feature engineering has been an important area of research in machine learning and has influenced various research investigations. This method attempts to create features out of raw data. Several types of features were proposed and compared in the context of computer vision, such as histogram of oriented gradients (HOG), despite the fact that it is exceedingly field-specific and frequently necessitates significant human work [42].

Machine learning attempts to make it possible for computers to learn from data, spot patterns, and make judgments based on that data. Machine learning comes in a variety of forms, including supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, and deep learning. Machine learning (ML) has recently gained a

lot of traction in research and has been applied to a number of applications, including text mining, spam detection, video recommendation, image categorization, and multi-media idea retrieval [43]. Other technologies that are occasionally utilized include semi-supervised learning and reinforcement learning [44]. The creation of supervised machine learning algorithms allows for the production of broad patterns and hypotheses by using data from external sources to forecast the outcome of future data. The goal of classification algorithms is to categorize data based on past knowledge. Classification is commonly used in data science issues [45]. This type of machine learning uses labeled data to train the algorithm. Applications like image classification, speech recognition, natural language processing, and predictive modeling are frequently utilized in the finance, healthcare, and marketing industries. Because they can make precise predictions and have a proven assessment framework, supervised learning algorithms are a popular choice for many practical applications.
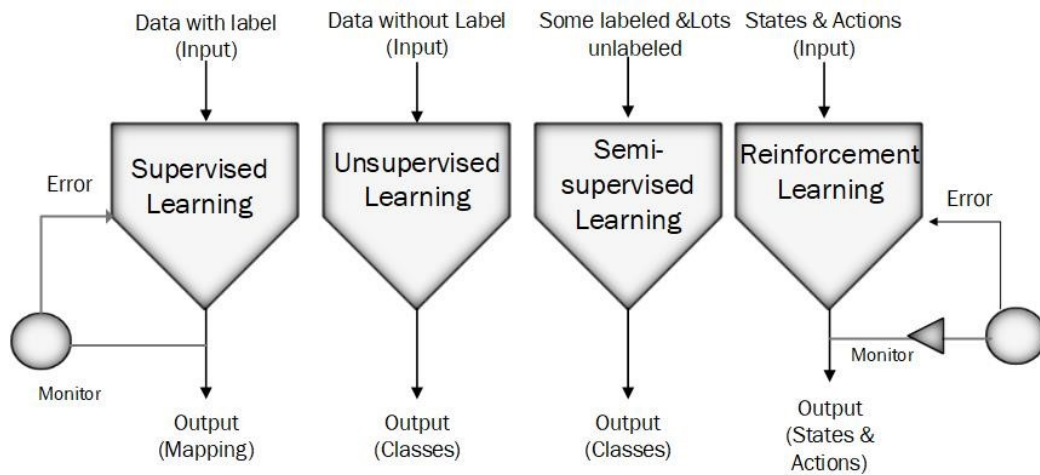
Figure 1.4: Supervised vs Unsupervised vs Semi-Supervised vs Reinforcement Machine Learning.

In medical imaging classification plays a significant influence in diagnosing diseases and conditions, as well as assisting in making judgments that are influenced by healthcare professionals based on image data. These methods involve analyzing and categorizing X-rays, MRI scans, and CT scans are examples of medical imagery, and more, to identify patterns and features indicative of specific diseases or conditions. Deep learning techniques include CNN's architecture which has shown remarkable performance in medical due to its capacity to automatically deduce hierarchical features from images, and classification. CNNs have been widely used for tasks like detecting tumors, identifying specific anatomical structures, and more [46]. For classifying medical images, two common machine-learning methods are Random Forests and Decision Trees. They function by developing a decision tree-based model that can capture intricate connections between target labels and visual attributes [47]. Another well-liked machine learning approach is the use of virtual machines (VMs) for classification tasks in medical imaging. They operate by identifying a hyperplane that effectively divides several classes in the feature space [48]. Ensemble methods combine multiple classifiers to improve classification performance. Bagging and

boosting are common ensemble techniques that have been applied to medical image classification [49]. Transfer learning is a technique for enhancing the performance of medical image classification tasks with less data by utilizing pre-trained models on big datasets. This approach has been particularly effective when labeled medical image datasets are scarce [50].

1.3.1 Types of Classification

In classification tasks, classes can be categorized into different types based on the number and nature of the classes involved.These different types of classes in classification highlight the varying complexities and characteristics of classification tasks. The choice of the appropriate type of classification depends on the nature of the problem, the available data, and the specific goals of the application. The commonly used types of classes in classification are:

Binary Classification

In machine learning, binary classification falls under the category of supervised learning since it is supported by high-quality training data that includes examples from two classes and labels each example as either class 0 or class 1 using a set of feature values. In order to predict the unobserved binary labels of new instances from their seen feature values, a binary decision rule is first built using the training data. A broad family of algorithms that automatically learn prediction rules from training data is represented by binary classification [51]. Classifying incidents into one of two categories—typically expressed as 0 or 1, positive or negative; true or untrue; etc.—is the aim. For applications like email spam detection, sentiment analysis, and medical diagnosis, this kind of categorization is employed.

Multi-class classification

The phrase "multi-class classification" refers to classification issues in machine learning that involve more than two classes. When comparing and evaluating various categorization models or machine learning methods, performance metrics are quite helpful [52]. The objective of this sort of machine learning classification issue is to place instances into one of many classes. For tasks like speech recognition, text classification, and picture classification, this form of classification is employed.

1.4 Neural Networks

A well-known machine learning method called a neural network was developed as an inspiration for biological neural network topologies. Through the creation of an artificial neural network, it imitates the functioning of the human brain [53]. They are made up of linked "neurons," or nodes, arranged in layers. It offers a variety of strong brand-new methods for handling issues with pattern recognition, data analysis, and control. The most fundamental kind of neural network is a feed-forward neural network. A hidden layer or layers, an output layer, and an input layer make them up. For a variety of applications, including image classification, regression, and fundamental pattern recognition, FNNs are employed [54].
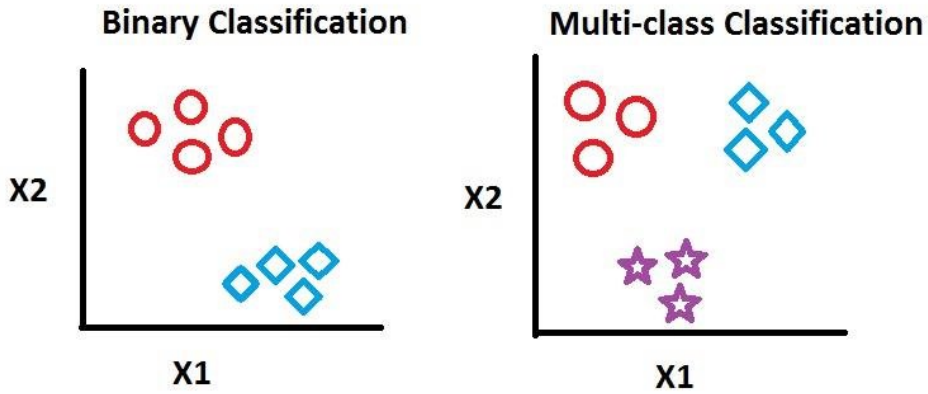
Figure 1.5: Binary classification vs Multi classification
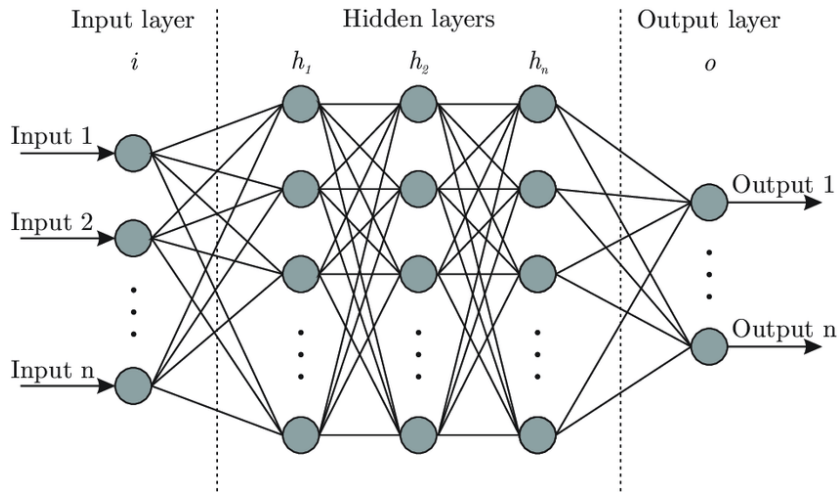
Figure 1.6: Basic structure of Neural Networks

A layer termed bias in a neural network (NN) can contain any number of nodes since the layers are independent of one another. The bias nodes are always initialized to 1. Importantly, a bias value allows for the right or left movement of the activation function, which may be critical for the success of ANN training. When the NN is employed as a classifier, the input and output nodes will correspond to the input features and output classes. The NN typically has an input and an output node when it is used to approximate a function, though. Nevertheless, there are more necessary designed hidden nodes than input nodes.

Images and other grid-like data may be processed using CNNs. In order to automatically learn hierarchical features, they employ convolutional layers, Consequently, they excel in image segmentation, object detection, and classification [55]. Natural language and time series are two examples of sequential data that RNNs are intended to handle. They have a feedback loop that allows information to be passed from one step of the sequence to the

next, making them suitable for tasks like language modeling and speech recognition [56]. Recent research has focused on creating neural networks with fewer parameters to reduce computational requirements and improve efficiency [57]. The most modern DL approaches have demonstrated excellent performance in a number of applications, including natural language processing (NLP), audio and speech processing, visual data processing, and others [58]. Bag of words and scale-invariant feature transform (SIFT) [59]. Once a unique feature is implemented and proven to work successfully, it creates a new area of study that is studied for decades. [60]. Deep learning algorithms can now handle challenging tasks including audio and picture recognition, natural language processing, and autonomous vehicles. TensorFlow, PyTorch, and Keras are a few well-known deep learning frameworks.
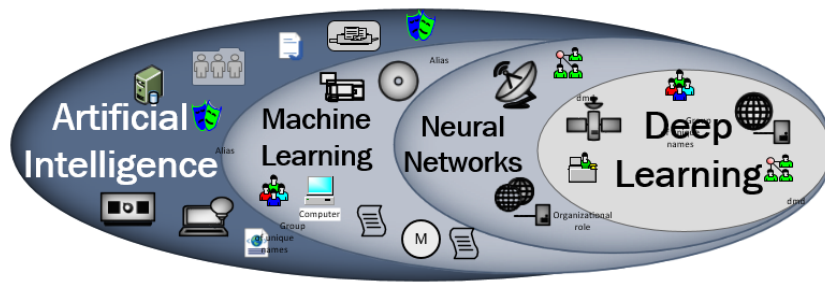
Figure 1.7: Artificial Intelligence vs Machine Learning vs Neural Networks vs Deep Learning

1.4.1 Types of Neural Network

Although there are several neural network types, each with a distinct design and field of use, new architectures are continuously being created for certain tasks and applications. Here are a few examples of the neural network types we utilized for this project.

Convolutional Neural Network

In the field of neural networks, The most renowned and widely used algorithm is CNN [61]. CNN has a number of advantages over its forerunners, but its main benefit is that it does so automatically and without human assistance, recognizing the important components [62]. CNNs have been extensively used in a wide range of fields, including computer vision, audio processing, face recognition, etc. CNNs were created using the neurons present in both human and animal brains and have a structure similar to a regular neural network. More specifically, CNN models the complex cell pattern found in the visual cortex of a cat's brain [63]. Convolutional Neural Networks (CNNs) are a particular kind of neural network made for processing grid-like input, such as images. By enabling automatic feature extraction from images and delivering cutting-edge performance in tasks like image classification, object identification, and more, they have revolutionized a number of disciplines, including computer vision. CNNs are made up of a number of components that enhance their performance.

Convolutional Layers

Convolutional layers are typically used at the end of the CNN architecture for making final predictions. They take the extracted features and transform them into a format suitable for the final classification or regression task [64]. It is the primary component of the convolutional neural network. When specific criteria are met, filtering in the convolutional layer causes the output result to be generated from the input. The neurons in convolutional layers are arranged in either a rectangular grid or a cubic block. Thus, input and output layers with filters may consist of a neuronal block with a rectangular grid or a cubical block shown in Fig 1.8.
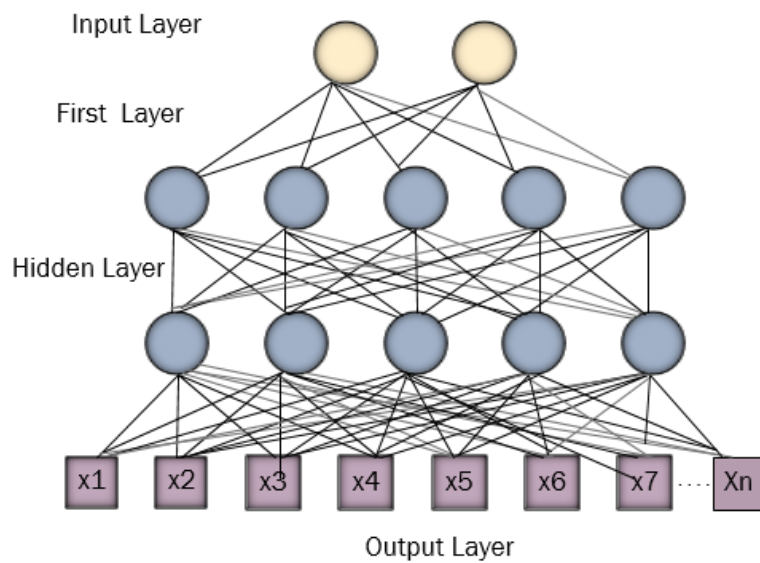
Figure 1.8: Convolutional layer

1.4.2 Building Block of CNN

This process is necessary to make sure that the neural network only obtains the information necessary for classifying an image. By doing this, the network's accuracy is increased while simultaneously ensuring that the least amount of processing power is required to train the network. The result of the convolution technique is referred to as a feature map, convolved feature, or activation map. A feature map is produced by using a feature detector. The kernel or filter are alternate name for the feature detector. The feature map is produced when the kernel is multiplied element by element using the aforementioned picture. These are CNN's pillars, in that order.

Padding

Padding is the technique of enhancing a source image's border pixels before convolutional filters are applied. It helps adjust the output feature's duration map and retain spatial

information. There are two types of padding: "valid" padding (no padding) and "same" padding (adding padding to keep output size the same) [65]. To avoid the issues outlined above, Simply adding layers of zeros to our input photos is the technique of padding. This avoids shrinkage because our $(\mathbf{x*x})$ image becomes $(\mathbf{x + 2p}) \, \mathbf{x} \, (\mathbf{x + 2p})$ after padding if p = the number of layers of zeros added to the image's border shown in Fig 1.9.

Figure 1.9: Padding (Zero Padding)

Kernel

A kernel or filter is a small matrix applied to the input image to perform convolution. It extracts specific features by multiplying its values with the corresponding pixel values of the input and summing the results [66]. Due to the smaller reduction in layer dimensions, it performs better, the resultant image is typical of size $(\mathbf{xk + 1})\mathbf{*}(\mathbf{xk + 1})$ when a $(\mathbf{x*x})$ image is convolved with a $(\mathbf{k*k})$ kernel.
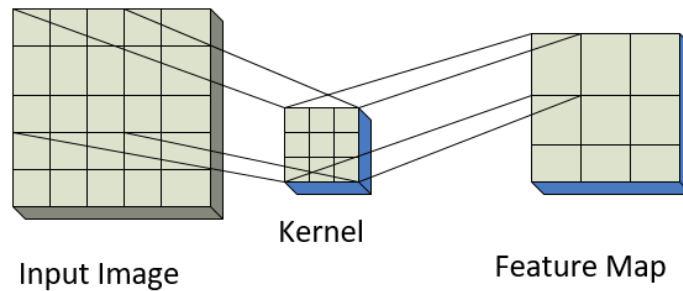
Figure 1.10: The kernel is an iterative matrix that moves through the input data, performs a dot product operation on a section of the data, and then outputs the outcome as a matrix of dot products.

Stride

The convolutional filter's step size is referred to as the stride, and it is measured in the input image's size. A larger stride results in smaller output feature maps, as fewer convolutions

are performed. Smaller strides can help retain more spatial information [64]. Stride is the number of the input matrix's pixels that are shifted. Our output image dimension for padding p, filter size **k*k**, input image size **x*x**,and stride 's' will be:

$$((x + 2p - k + 1)/s + 1) * ((x + 2p - k + 1)/s + 1) \tag{1.1}$$
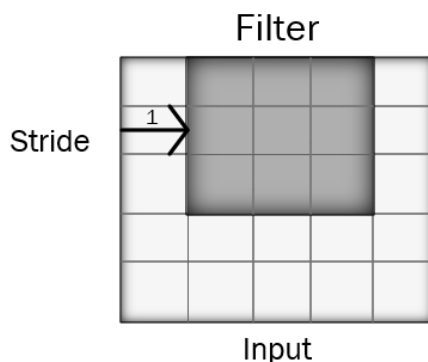
Figure 1.11: The filters are shifted one pixel at a time when the stride is 1. The filters are shifted 2 pixels at a time when the stride is 2, and so on.

Pooling

By sliding a filter of a certain size with a certain stride size and determining the maximum or average of the input, the output of the Convolutional layers is sampled down. The feature maps' spatial dimensions are reduced but vital data is preserved via pooling layers. The feature maps are downsampled using methods like max pooling and average pooling [67].

Types of Pooling

Max pooling is a straightforward method that helps move forward with the image's most crucial attributes by taking the maximum of a region. The brighter pixels in the picture are selected using max-pooling, where average pooling is a technique that uses the average value for feature map patches to build a pooled feature map that has been down-sampled.

A specific form of linear operation called convolution is employed in feature extraction. By applying a tiny array of numbers to the input, which is an array of integers, it extracts the features. An element-wise product between each element of the kernel and the input tensor is calculated at each point of the tensor and summed to produce the output value at the corresponding location of the output tensor. Modern CNN designs often employ zero padding to preserve in-plane dimensions and enable the application of extra layers. Without zero padding, each feature map after the convolution process would get smaller.

Residual Neural Network

He et al. (2015) named the artificial neural network (ANN) known as the Residual Network (ResNet). The idea of residual learning was first proposed by Residual Neural Networks
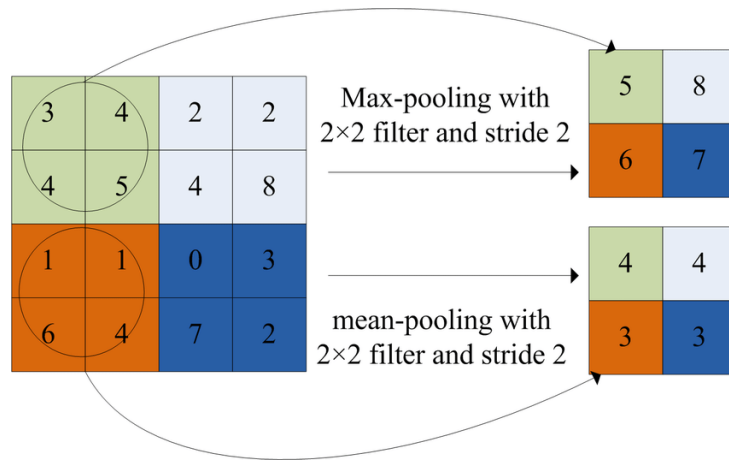
Figure 1.12: Max Pooling and Average Pooling

(ResNets), a form of deep neural network design. They were created to solve the vanishing gradient issue and make it possible to train incredibly deep networks. ResNets are frequently utilized for a variety of computer vision applications and have a substantial effect on the field of deep learning. It is made to handle sequential data, including time series data, audio signals, and text written in natural language.ResNet16, ResNet18, ResNet34, ResNet50, ResNet101, ResNet110, ResNet152, ResNet164, ResNet1202, and so forth are only a few of its many versions.
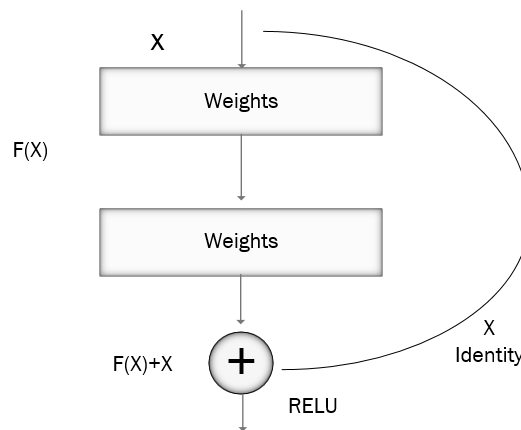
Figure 1.13: Residual unit structure diagram

The core idea of ResNets is the residual block. Instead of learning the desired output directly, The residual (difference) between the input and output is learned by a residual block. A "shortcut" or "skip connection" that connects the input straight to the output of one or more layers is introduced to accomplish this. This allows the network to learn the identity function, making it easier to train deep networks. In Fig 1.13 the first thing that stands out to us is the direct link that bypasses certain triple-layer layers. The decoder's

connection, often known as the core of residual blocks, is this link. Without the skip connection, input 'X is multiplied by the pre-trained layer before being added to a bias term, hence the output is not pre-trained.

ResNet-18

ResNet-18 is the name of a convolutional neural network with 18 layers. A pre-trained version of the network that has been trained on more than a million images is present in the image database. These Residual blocks have made it possible to train incredibly deep networks, and the ResNet model is made up of these blocks.

Building Blocks of ResNet-18

ResNet-18 is made up of a variety of residual blocks. A series of layers, including convolutional layers, batch normalization, activation functions, and a skip connection, make up each residual block. The inclusion of the residual/skip connection, which adds the original input to the output of the convolutional layers, is the main novelty of ResNet. Two convolutional layers make up ResNet-18's fundamental residual block, which is then followed by batch normalization and ReLU activation for each layer.

$$H(x) = f(wx + b) \tag{1.2}$$

Now that a new skip connection method has been developed, the output of H(x) has changed to

$$H(x) = f(x) + x \tag{1.3}$$

Triple-layer of the input may be varying f decoder section output which might happen with a convolutional layer of more-trained layers. The two methods described above can thus be used to solve this problem. Zero is a decoder that uses a skip connect to enhance its dimensions. For the input to fit the dimensions, 11 convolutional layers are added. The result of this situation is:

$$H(x) = f(x) + w_1.x \tag{1.4}$$

When utilizing the first method, no new parameter is supplied; however, an additional parameter, w_1, is introduced here. The skip connections strategy in ResNet addresses the issue of disappearing gradient in deep CNNs by providing an additional shortcut channel for the gradient to flow through. Furthermore, the skip connection feature is helpful since it allows regularization to bypass any layer that negatively affects the architecture's performance.

1.4.3 Efficient Neural Network

A software solution known as an efficient neural network uses machine learning (ML) techniques to "mimic" the functions of the human brain. The neural network architecture used in EfficientNet has been scaled up. Compound scaling is a recently suggested technique that scales all dimensions using a compound coefficient. The three factors—depth, breadth, and resolution—are scaled up in this context in a methodical, rational manner. Each layer's feature maps are expanded by the width scale. Layers are added to the

network by using the depth scale. The supplied photos' resolution is raised using the Resolution scale.

EfficientNet-B0

In order to train EfficientNet-b0, a convolutional neural network, more than a million images from the ImageNet database were used. The network can categorize photos into 1000 different item categories, including various animals, a keyboard, a mouse, and a pencil. Every architecture shares characteristics with its predecessors. The additional parameters caused by the various feature maps are the sole distinction. With the exception of the multiplied block(x2) that grows and covers additional blocks, all the models have the same design as the prior one. This provides a large number of input parameters, making it an extremely robust model. The differences between all the models are easily visible, and they steadily expanded the number of sub-blocks. The compound scaling approach was used to scale up in two phases, starting with EfficientNet-B0:

Step 1

The coefficient, which necessitates a small grid search for the network's depth, breadth, and resolution constants, was set to 1 under the presumption that there would be twice as many sources accessible.

Step 2

To produce the subsequent versions from B1 to B7, the constants are then fixed and scaled up from the baseline network using various coefficients.Due to these restrictions, EfficientNet was given priority in this work as well. The remaining EfficientNet models were also ignored for this primary reason since they calculate a large number of parameters, which uses a lot of processing power and time and results in a subpar result. Fine-tuning has been used in conjunction with baseline versions of each model for determining MVA for EfficientNet models. In Fig 1.14 the flow chart is useful in understanding the structure and workings of an ENN and is often used as a visual aid in explaining the process.
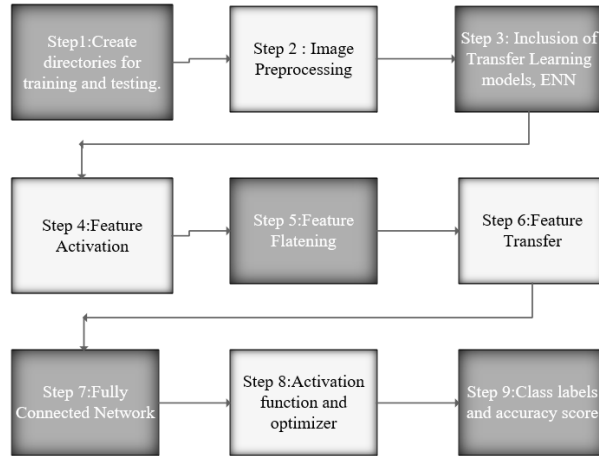
Figure 1.14: This flow chart is a graphical representation of the different stages and operations involved in the ENN architecture.

To enhance EfficientNet models B1–B7, we used the EfficientNet-B0 network and evenly increased the depth, breadth, and resolutions using a simple and efficient compound coefficient. For the common transfer learning problem, EfficientNets are smaller, have fewer parameters, and generalize better to produce higher disappointing results. In order to redesign, the suggested study adjusted EfficientNet models B0-B4. Medical imaging uses pre-trained data sets, hence the data augmentation approach is used to provide extra, distinctive CXR images in order to control overfitting. When we applied the pre-trained EfficientNets to the CXR data set, we enhanced the models by utilizing a global average pooling (GAP) to reduce the number of parameters and handle overfitting. Before a final dense layer that acts as the output and has a SoftMax activation function to calculate the probabilities of the input CXR representing the normal and infected classes, two dense layers with a ReLU activation function and a dropout rate of 0.4 come after the GAP.

Activation Functions

Activation functions are crucial in transforming artificial neural network output into non-linear outputs because without this non linearity, the network's findings would be less accurate. For back-propagation learning to take place, the activation function needs to be differentiable. When utilizing neural networks to code fractal pictures, the nonlinearity of the activation function is crucial since the coefficients of the Iterated Function System vary depending on the different forms of fractals.The sigmoid function, which yields a positive result, is the most often selected activation function. Other functions that can have positive or negative values based on the input to the network, such tans or arctan, have a tendency to train neural networks more quickly [68]. Some of the most popular activation functions are listed below.

ReLU (Rectified Linear Unit)

The activation function used most frequently in deep learning nowadays is ReLU. It leaves all positive inputs unaltered while replacing all negative inputs with zero. In deep learning,

one of the most often used activation functions is the rectified linear unit (ReLU) activation function. It is a non-linear function that transforms all negative inputs to zero while preserving positive inputs. This is how the ReLU function is described:

$$f(x) = max(0, x) \tag{1.5}$$

The ReLU activation function is computationally efficient and easy to implement, which makes it a popular choice for building deep neural networks. IIt has been demonstrated to deliver positive outcomes in a variety of applications, including speech recognition, picture classification, and natural language processing.

Leaky ReLU

The leaky rectified linear unit A different form of the rectified linear unit (ReLU) activation function is known as leaky ReLU. The primary distinction is that the typical ReLU activation function maps all negative inputs to zero, but the leaky ReLU activation function permits a modest negative slope for negative inputs. As specified, the Leaky ReLU function is:

$$f(x) = max(x, x) \tag{1.6}$$

where α is a small positive constant, typically set to 0.01, that determines the slope of the function for negative inputs. The Leaky ReLU activation function can help mitigate the dying ReLU problem, where a traditional ReLU activation function may produce neurons that never activate and become dead. In several deep learning models, it is commonly employed and has been shown to produce good results in many applications, including image classification, speech recognition, and natural language processing.

Sigmoid

Any input may be translated into a probability by using the sigmoid activation function, which transfers it to the range of 0 to 1. In mathematical terms, the sigmoid activation function is:

$$S(x) = \frac{1}{1 + e^{-x}} \tag{1.7}$$

where e^x is the exponential function. The sigmoid function has a smooth and continuous transition, which makes it well-suited for binary classification problems. However, it does have certain drawbacks, such as the vanishing gradient problem, which can make deep neural network training challenging.

Softmax

The softmax activation function addresses multi-class classification problems. It creates a probability distribution over a number of classes from the inputs. The definition of the softmax function is as follows:

$$f(x_i) = \frac{e^{x_i}}{\sum(e^{x_j})} \tag{1.8}$$

where e^{x_i} is the input to the function for class i and the sum is taken over all classes j. Because of the softmax function's guarantee that the output total will never exceed

1, they may be used to represent a probability distribution over several classes. In the output layer of a neural network, it is typically used to produce a probability distribution over a number of classes. The final forecast is then chosen as the class with the highest probability. It is frequently employed in several applications, such as speech recognition, picture classification, and natural language processing.

Tanh (Hyperbolic Tangent)

The input range for the tanh activation function is from -1 to 1. Similar to the sigmoid activation function, but with various values that are squashed. One definition of the tanh function is:

$$f(x) = tanh(x) = \frac{2}{(1 + e^{-2x})} - 1 \tag{1.9}$$

The tanh activation function is similar to the sigmoid activation function, but it maps its inputs to a different range, which can be useful in some cases. It can produce saturation at the extremes of the range, which can make it more difficult to optimize compared to other activation functions, such as the rectified linear unit (ReLU). The tanh activation function is widely used in many applications, including recurrent neural networks, where it is used to model sequences of data. It is also commonly used in CNNs, where it is applied to add nonlinearity to each neuron's output.
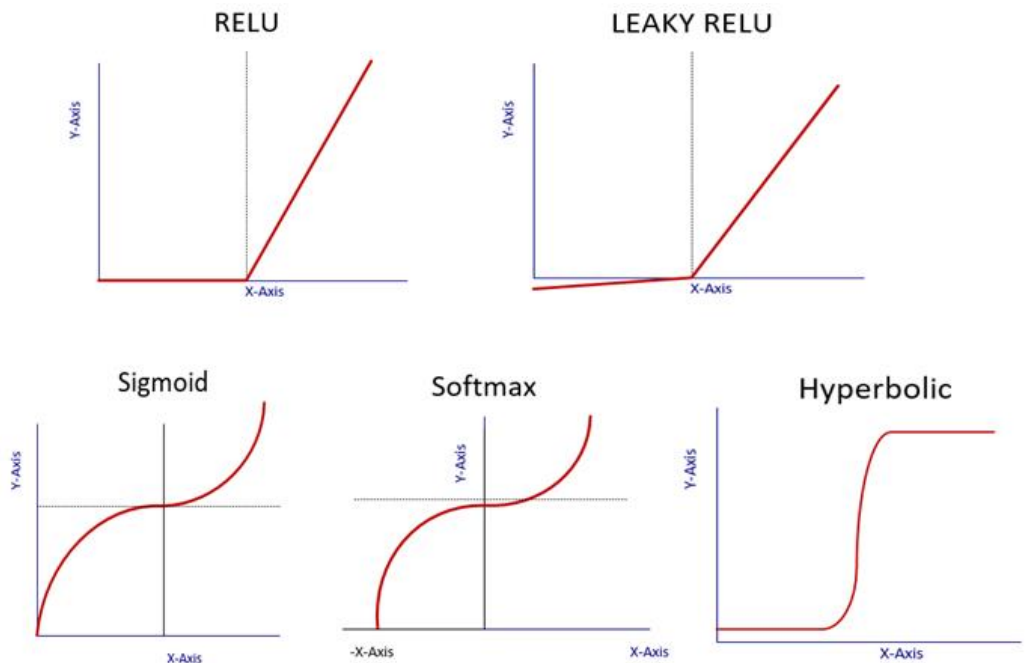
Figure 1.15: Grapical Representation of Activation functions

1.4.4 Application of Neural Networks

Recent studies have extensively explored the use of neural networks for medical image diagnosis, including the detection of diseases like melanoma. CNNs, in particular, have demonstrated remarkable success in analyzing medical images such as X-rays, MRIs, and histopathological slides. Neural networks are being employed in the pharmaceutical industry for drug discovery and development. They aid in predicting the properties of molecules, identifying potential drug candidates, and optimizing chemical structures. In the field of NLP, transformers like BERT and GPT-3 have gained prominence. These models are being utilized for tasks such as sentiment analysis, language translation, and even generating human-like text. Neural networks play a pivotal role in the development of self-driving cars. They enable object detection, lane tracking, and decision-making based on real-time sensor data. Financial institutions are leveraging neural networks for stock market prediction, risk assessment, fraud detection, and algorithmic trading. Neural networks are used for analyzing environmental data, such as satellite imagery and sensor readings, to predict natural disasters, monitor climate changes, and assess environmental impacts. Healthcare providers are using neural networks to analyze patients' medical records, genetic data, and lifestyle information to personalize treatment plans and predict disease risks. Speech Recognition which is used to build speech recognition systems that can transcribe and translate speech into text. Recommended Systems, is used to build recommended systems that suggest items to users based on their preferences and behaviors. Financial Forecasting, is used to predict stock prices, currency exchange rates, and other financial market trends.In order to diagnose illnesses, forecast patient outcomes, and provide individualized treatment approaches, healthcare is used. Robotics, which is used to build intelligent robots that can perform tasks such as object recognition, navigation, and control.

1.5 Goal of the Study

The study aims to enhance the detection of aggressive melanoma skin cancer using computer vision techniques, particularly through the application of neural network models. The goal is to address the challenge of timely diagnosis of melanoma, which can become life-threatening if not detected early. By utilizing various neural network architectures, including EfficientNet-B0, ResNet-18, and basic CNN, the study seeks to develop accurate and efficient models for classifying melanoma skin lesions. The ultimate objective is to improve the accuracy of melanoma detection, potentially leading to early interventions and improved patient outcomes.

Chapter 2

Literature Review

In this chapter, we begin our study with the neural network idea before moving on to discuss neural network methods. The neural network techniques are efficient for fine-tuning the networks, as in Convolutional Neural Networks, Residual Neural Networks, Efficient Neural Networks on image datasets including Skin cancer, etc.

In a recent study, PCA and SVM are used for the classification of MSC image data set [69]. In order to determine how varying the number of training images and epochs could impact accuracy, Rina Refianti and Achmad Benny Mutiara applied CNN to the MSC image data set [70]. An EfficientNet-B0 was utilized in this article to detect network-based malignancies from MRI data [71]. The experiment employs 4,509 TUSP images, with 3,386 training and 1,123 testing images, so the 18-layer CNN model ResNet performs well in categorizing TUSP images [72]. The project focuses on training ResNet-18 and ResNet-50 models to detect colorectal cancer using colon gland images. ResNet-50 surpasses Resnet-18 according to the accuracy, sensitivity, and specificity across all testing data sets [73]. The researcher was able to obtain images of skin lesions from the ISIC 2019 data set making use of the proposed spiking VGG-13 model [74]. MAFCNN-SCD is a unique approach for skin cancer diagnosis in dermoscopic pictures that uses an optimized multi-attention fusion CNN. Extensive simulations show that the technology outperforms competing techniques, showing its potential for precise and efficient skin cancer categorization [75]. Researchers used hybrid CNN models with an SVM classifier to automatically classify dermoscopy pictures as benign or malignant lesions [76]. A dermoscopy-based ML diagnostic system for identifying and categorizing skin lesions with ABCD rules for prepossessing, segmentation, feature extraction, and classification [77]. Bonechi, Simone's manuscript proposes a DL method to classify the cutaneous lesions. Preliminary results from the ISIC data set show promising outcomes, highlighting the significance of information fusion in enhancing classification accuracy [78]. They used DL models they had chosen using a search method such as EfficientNets, SENets, and ResNeXt WSL. Additionally, they incorporate patient metadata with an additional input branch in our dense neural network [79]. The detection and classification of MSC using Support vector machines (SVM) and neural networks (NN) by Mhaske H and Phalke [80]. The system comprises three stages such as input gathering, augmentation, model creation, and prediction by fusing AI techniques that are CNN and SVM with image processing technologies [81]. Results from the benchmarking using 1011 lesion cases demonstrate the effectiveness of the approach in accurately classifying the checklist criteria, performing skin condition diagnosis, and generating feature vectors for image retrieval and region localization [82]. Li Weipeng demonstrates superior performance, particularly for rare disease categories, through comprehensive experiments using various deep-learning model architectures [83]. The researchers applied filter detection using discrete wavelet transformation, dimension reduction using PCA, and classification using supervised ML algorithms [84]. In automated skin lesion diagnosis, a quantitative study concludes that dermoscopic images are better than macroscopical images. Data from 2917 cases, each comprising a dermatoscopic and macroscopic image, and patient metadata, was used to access the effectiveness of the method [85].

In this article, the key contribution is the creation and assessment of a Using additional parameters and the highly accurate neural network-based MSC image categorization system, using CNN, ResNet18, and EfficientNet-B0. In the upcoming sections, we will delve into the Materials and Methods, where we outline the data set and methodology, and the Results, where we present the outcomes of our experiments.

References	Dataset	Methods	Findings
Refianti R, Mutiara AB, [70]	MSC	CNN	Acc=93%
Shah HA, Saeed F. [71]	MRI	EfficientNet-B0	Acc=87%
Guo M, Du Y. [72]	TUSP	ResNet-18	Acc=83%
Sarwinda D, Paradisa RH, [73]	Colorectal cancer	ResNet-18	Acc=80%
Qasim Gilani S, Syed T, [74]	Skin cancer	VGG-13	Acc=89%
Zghal NS,Derbel N [77]	MSC	Image processing	Acc=90%
Bonechi, Simone [78]	ISIC	ResNet-50	Acc= 83%
Gessert, Nils [79]	ISIC HAM10000	EfficientNets	Sens=72%/74%
Mhaske H, Phalke D. [80]	MSC	NN, SVM	Acc:NN=60-70%, SVM=80%-90%
Vijaya lakshmi M. [81]	MSC	CNN, SVM	Acc =85%
Kawahara, Jeremy [82]	7-point dataset	Inception v3	Sens=52%/60% Spec=90%/91%
Li, Weipeng [83]	ISCI	SENet154	Sens=85%/87%
Elgamal M. [84]	Skin cancer	ANN KNN	Acc=95% Spec=Sens=95%
Yap, Jordan [85]	ILSVRCo 2015+Own	ResNet50	Precision=72% Acc=72%
Moldovan D. [86]	HAM(10000)	Deep Learning models	Acc=75%-85%

Table 2.1: Summary of Related Work

Chapter 3

Materials and Methods

In this study, we present a summary of the data set that was implemented in our study in addition to the approach used to include neural networks in the classification task. The methodology discusses the processes needed to create and train the neural networks, while the data set selected for this research serves as the fundamental building block for our investigations.

3.1   Data Acquisition

"Melanoma Skin Cancer" is the name of the data collection that we used in our research. It has 10,605 images with binary classes with the most samples being malignant and the minority class with the fewest samples benign that were gathered from the Kaggle database. Data is collected from different ISIC MSC directories, which is the largest community-based data platform for machine learning in the world.
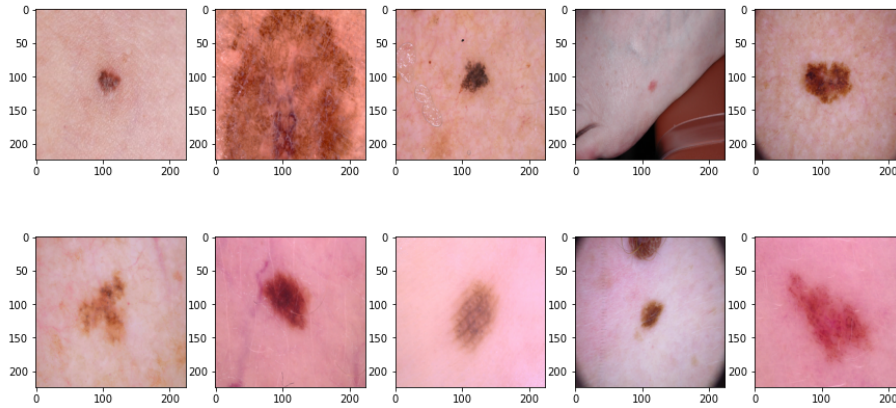
Figure 3.1: The data set consists of 9605 images for training the model and 1000 images for evaluation of the model with image dimensions (224x224) with RGB scale.

In benign which has 5500 images and in malignant, there is 5105. Any data set with an uneven distribution of values across its classifications is deemed unbalanced. However, it is generally accepted in the field that data sets displaying considerable, and in some cases dramatic imbalances, are indicative of unbalanced data. The performance of the majority of popular classifier learning algorithms, which presuppose an evenly distributed class distribution and equal miss-classification costs, suffers significantly when trying to classify data with an uneven class distribution.

3.2   Neural Networks for Image Classification

The methodology discusses the processes needed to create and train the neural networks, while the data set selected for this research serves as the fundamental building block for our investigations.

3.2.1 Convolutional Neural Network (CNN)

It is required to have a sample data set with a variety of melanoma manifestations. The original data set is then used to build subsets for training and testing. Let

$$F = (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) \tag{3.1}$$

where F denotes the data set of MSC images. Each pair (x_i, y_i) represents an input image x_i, and its corresponding label y_i, is the class label (0 for benign and 1 for malignant), and n shows the total number of images. A single output matrix computation is described as follows:

$$F_i = f(\sum_{i=1}^{N} I_i * k + B_j) \tag{3.2}$$

First, a kernel matrix K corresponding to each input matrix I_i is convolved. Then, each member of the resulting matrix is given a bias value B_j, and the sum of all the convoluted matrices is calculated. One output matrix F_i is generated by applying a nonlinear activation function f to every component of the previous matrix.

$$n_o = \frac{n_i + 2p - k}{s} + 1 \tag{3.3}$$

n_o displays the number of output features, n_i reveals the number of input characteristics, p shows padding size, s shows stride size and k is the number of convolution filter sizes. Analyze the effectiveness of the train model using a different testing data set after training. Let F_{test} denote the testing data set, where m is the number of test images.

$$F_{test} = (xtest_1, ytest_1), (xtest_2, ytest_2), ..., (xtest_m, ytest_m) \tag{3.4}$$

Predict the labels for the test images using the trained model

$$\hat{y}test_i = F(xtest_i) \tag{3.5}$$

Compare the predicted labels $\hat{y}test_i$ with the true labels $ytest_i$ to compute the evaluation of the model.

Figure 3.2: The architecture of basic CNN structure, and the various layers that make up the CNN model on the skin cancer image data set

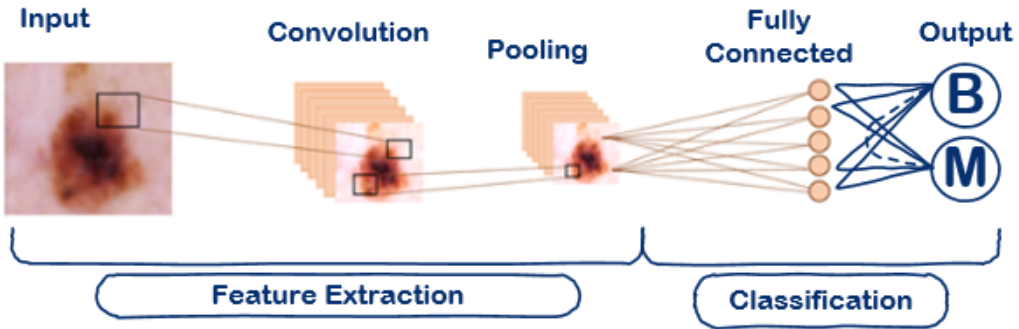The training procedure entails feeding the training images and labels into the model repeatedly. Any required hyperparameter modifications are made once the validation set performance of the model was assessed for over-fitting.

3.2.2   ResNet-18

ResNet-18 is a popular CNN architecture known for its deep residual learning. The architecture has 18 layers, including residual blocks, convolutional layers, batch normalization algorithms, and ReLU activation functions. Propagate an input image x_i through the ResNet-18 model to obtain the predicted output $\hat{y}_i$. For each layer in the ResNet-18 architecture, compute the activation value a_i as follows: For a convolutional layer,

$$a_i = Convolve(x_i, W_i) + b_i \tag{3.6}$$

where W_i represents the weights of the layer and b_i represents the bias term. Implement batch normalization For all activations in a specific channel, BN indicates the same normalization.

$$BN_{b,c,x,y} = \gamma_c * \frac{I_{b,c,x,y} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}} + \beta_c \tag{3.7}$$

In this case, BN subtracts the mean activation μ_c from all input channel c activations, where β is the sum of all channel c activations for every attribute b in the whole mini-batch and for all spatial x, y positions. Then, following a similar strategy, BN divides the centered activation by the standard deviation σ_c (plus ϵ for numerical stability). Testing is conducted using running mean and variance averages, followed by an affine transformation channel-wise that is parametrized by γ_c and β_c, which are discovered during training.

Apply the ReLU activation function to introduce non-linearity and enhance feature representation.

$$f(x) = \left\{ \begin{array}{cc} x, & x > 0 \\ 0, & Otherwise \end{array} \right\} = max(0, x_i) \tag{3.8}$$

For each residual block, compute the output by passing the activation value a_i through a series of convolutional layers and adding the input to the output, creating a skip connection. Apply global average pooling lastly to get a compact representation of the characteristics.

$$GAP = \frac{1}{X_k} (\sum_{x \epsilon X_K} x^{p_k})^{p_k} \tag{3.9}$$

When a parameter is $p > 0$. The pooled feature map's contrast is increased and the image's salient features are brought into sharper focus when this exponent is set to $p > 1$. The expected probability distribution over the class labels can be obtained by running the pooled features through a fully connected layer with softmax activation.

$$p_i = \frac{e^{z_i}}{\sum_j^J e^{z_j}} \tag{3.10}$$

where the softmax input is z_i and the number of categories is j. A probability distribution with a range of (0,1) and a sum of 1 can be created from the output value of a multi-classification using the softmax function p_i. Then, the distance between the actual

and desired output may be calculated using the cross-entropy loss function of Softmax, and the formula is as follows:

$$L = \sum_{j}^{J} y_J p_i \qquad (3.11)$$

where the real tag is represented by y_j. The loss function for the classification issue represents the discrepancy between expected and actual outcomes. Finding a set of optimal solutions in the parameter space to reduce L can be thought of as the network's training process as a parameter optimization process.
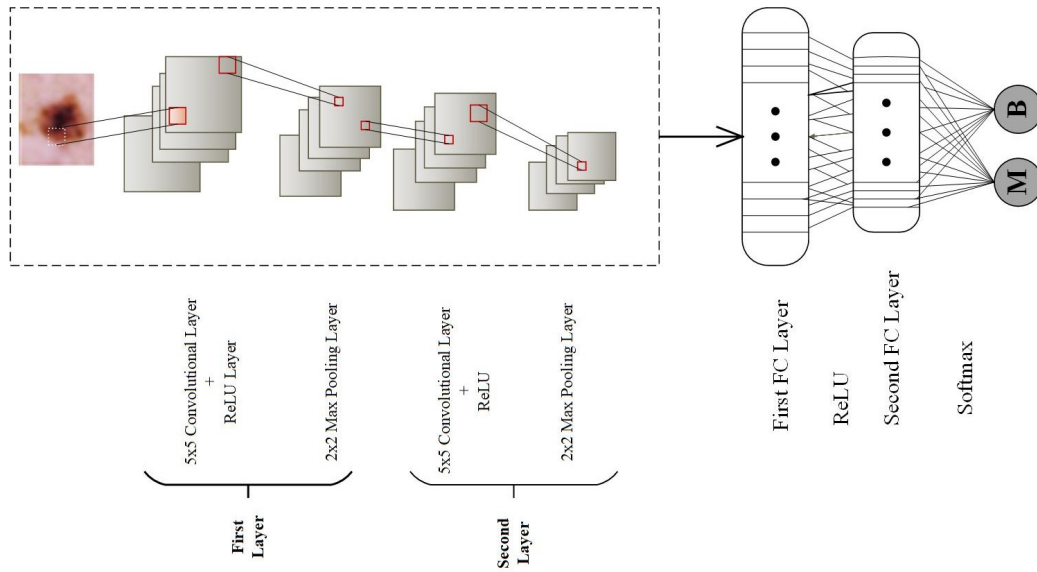
Figure 3.3: The ResNet-18 architecture can be summed up in this visual representation. Convolutional layers, batch normalization, and ReLU activation are applied to the input. The final layers are comprised of global average pooling, fully linked layers with ReLU activation, and a softmax activation for generating class probabilities.

Layer	Filter Size	Resolution	Activation Function
Input	-	224x224	-
Conv	3x3	112x112	Relu
Conv	3x3	112x112	Relu
Conv	3x3	112x112	Relu
Conv	3x3	112x112	Relu
Residual	-	112x112	
Conv	3x3	56x56	Relu
Conv	3x3	56x56	Relu
Conv	3x3	56x56	Relu
Conv	3x3	56x56	Relu
Residual	-	56x56	
Conv	3x3	26x26	Relu
Conv	3x3	26x26	Relu
Conv	3x3	26x26	Relu
Conv	3x3	26x6	Relu
Residual	-	26x26	
Conv	3x3	13x13	Relu
Conv	3x3	13x13	Relu
Conv	3x3	13x13	Relu
Conv	3x3	13x13	Relu
Pooing and FC	-	4096	Relu
Second FC	-	10000	Softmax

Table 3.1: ResNet-18's architecture includes convolutional layers, followed by four levels of residual blocks.
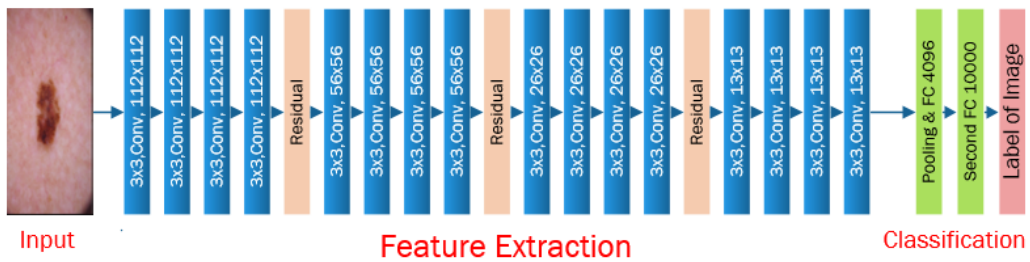
Figure 3.4: Visual architecture of ResNet-18 includes convolutional layers, followed by four levels of residual blocks.

3.2.3   EfficientNet-B0

Efficient-NET defines the design of convolutional neural networks as well as a scaling method that uniformly expanded the network's depth d, width w, and resolution r for the optimal performance. It has several variants inside of it, ranging from b0 to b7; for this study, the b0 version is chosen. The EfficientNet-b0 design has 5.3 million parameters in total.

$$d = \alpha^{\phi} \tag{3.12}$$

$$w = \beta^{\phi} \tag{3.13}$$

$$r = \gamma^{\phi} \tag{3.14}$$

$$s.t : \alpha.\beta^2.\gamma^2 = 2 \tag{3.15}$$

$$where, \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \tag{3.16}$$

According to the equation, FLOPS would rise by $((alpha.beta2.gamma2)phi)$ from the original equation, where phi is the user-defined coefficient. To get the predicted result I, propagate an input image xi via the EfficientNet-B0 model. Iteratively adjusting the parameters based on the obtained gradients will train the EfficientNet-B0 model using the training subset. Measure the trained model's classification accuracy and other performance measures on the testing subset. Fig 3.5 depicts the EfficientNET-b0 network's visual flow chart and Table 3.2 shows the baseline network.

Layer	Filter Size	Resolution
Input	-	224x224
MB Conv	3x3	112x112
MB Conv	3x3	112x112
MB Conv	3x3	56x56
MB Conv	3x3	28x28
MB Conv	3x3	14x14
MB Conv	3x3	14x14
MB Conv	3x3	7x7
Conv	1x1	7x7
Pooing and FC	-	7x7

Table 3.2: This tabular representation summarizes the EfficientNet-B0.Taking into consideration the filter size and resolution, the MB (Melanoma-Biopsy) convolution layer is specifically designed for melanoma skin cancer detection.
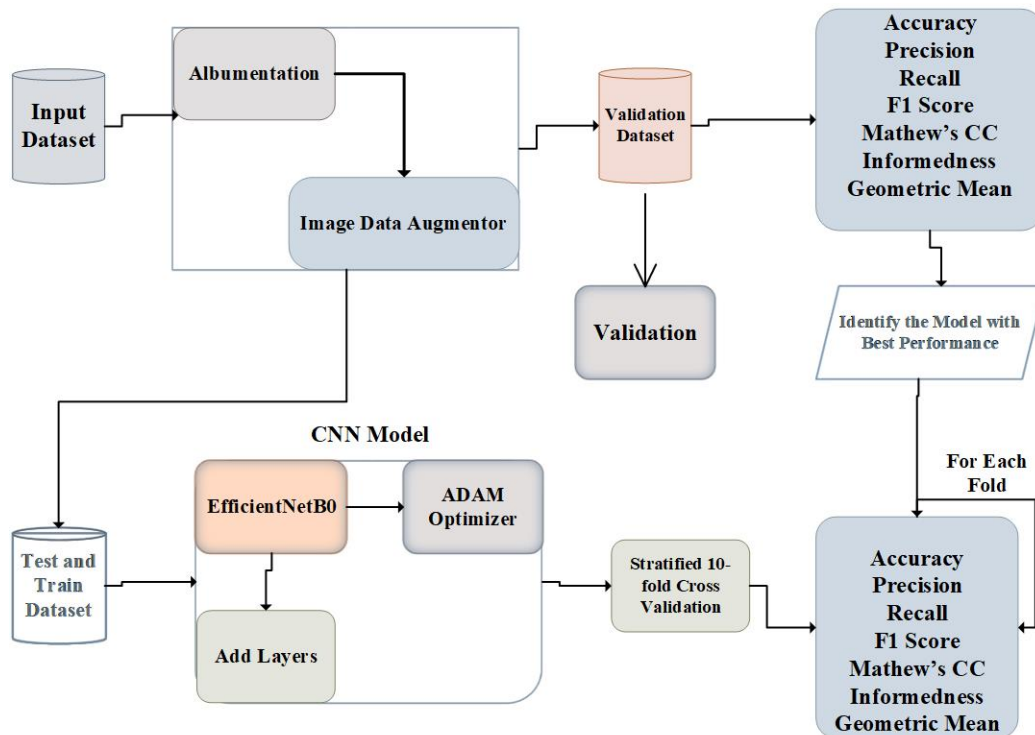
Figure 3.5: EfficientNet-B0 typically consists of multiple layers for feature extraction and prediction. It can be combined with Albumentations, an image augmentation library, to increase data diversity and improve the model's generalization. The Adam optimizer is employed to optimize the model's parameters during training.

Figure 3.6: This representation provides an overview of the EfficientNet-B0 architecture. Taking into consideration the filter size and resolution, the image size, and the MB (Melanoma-Biopsy) convolution layer.

3.3   Evaluation Metrics for Classification

By comparing predictions to actual class labels, measurement To assess the effectiveness and performance of categorization models, measurement m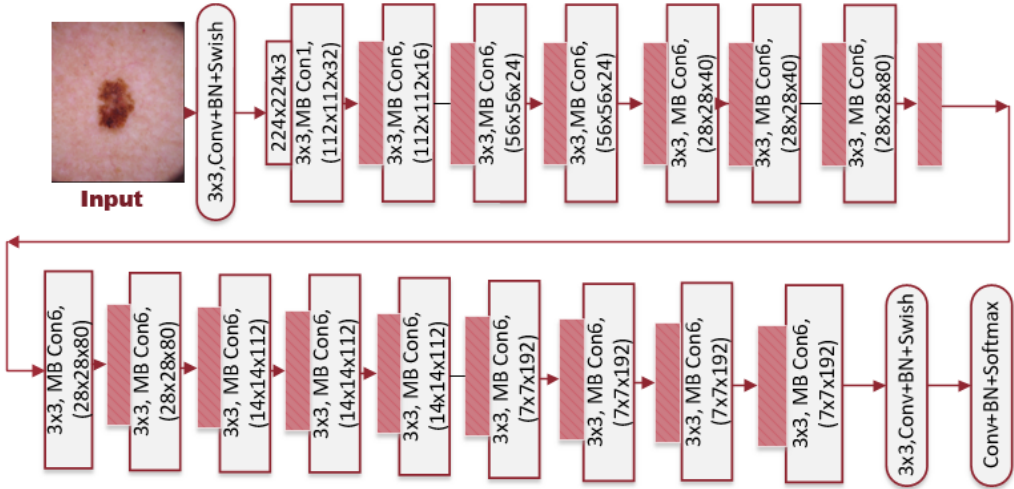easures are utilized. The choice of evaluation metrics depends on the specific requirements of the classification problem and the desired trade-offs between different aspects of performance. With the aid of the Confusion Matrix, a machine-learning concept that provides details regarding actual and anticipated classifications made by a classification system, we may present several often used assessment metrics for classification. A confusion matrix contains two dimensions, one of which is index by the class that an item really belongs to, and the other by the class that serves as the predictive class. We have used the equations 3.17 to 3.25 with the help of a 2x2 confusion matrix showing the predicted and actual classification. The values in the following table are labeled as true positives (TP) for events that were correctly predicted, false positives (FP) for events that were incorrectly predicted, true negatives (TN) for correctly predicted events, and false negatives (FN) for incorrectly predicted events.

		Actual Values		Total
		Positive	Negative	Total
Predicted Values	Positive	TP	FP	$TP + FP$
	Negative	FN	TN	$FN + TN$
	Total	$TP + FN$	$FP + TN$	N

3.3.1   Accuracy

A machine-learning model's performance is assessed using accuracy. The ratio of the model's accurate predictions to all of its prior forecasts is known as the success rate. The accuracy of binary classification issues is determined by dividing the total number of predictions by the proportion of true positive and true negative predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.17}$$

3.3.2   Sensitivity

Sensitivity is a measure of a model's accuracy in identifying positive cases. It is calculated by dividing the total number of true positive data cases by the number of cases in which the model accurately predicted the positive class. A high sensitivity score indicates that the model is good at spotting positive circumstances and has a low rate of false negatives. In some situations when the cost of missing a positive instance is substantial, sensitivity is a valuable statistic. For instance, it is important for medical diagnoses to have high sensitivity to guarantee that all positive instances are found.

$$Sensitivity = \frac{TP}{TP + FP} \tag{3.18}$$

3.3.3   Specificity

Specificity describes the extent to which a model or system can accurately identify a certain class or category based on a set of inputs. In terms of machine learning, Specificity is a

model's capacity to detect negative occurrences, High specificity means that there are few false positive errors, meaning that the model is good at avoiding false alarms.

$$Specificity = \frac{TN}{TN + FP} \tag{3.19}$$

3.3.4 F1 Score

The F1 score, which takes into account both precision and recall, is used to evaluate the accuracy of a model. The harmonic mean of recall and accuracy, it offers a harmony between the two measures. The F1 score is frequently used as a single statistic to assess the effectiveness of a classifier. Its range is 0 to 1, with 1 reflecting perfect accuracy.

$$F1 - Score = \frac{(2 * P) * (SNS)}{P + SNS} \tag{3.20}$$

3.3.5 Precision

In order to assess the accuracy of a classifier or retrieval system, precision is a statistic used in machine learning and information retrieval. The number of true positive instances divided by the total number of true positive and false positive instances is used to calculate the percentage of relevant occurrences among the retrieved instances.

$$Precision = \frac{TP}{TP + FP} \tag{3.21}$$

3.3.6 Error Rate

The ratio of inaccurate predictions to all of the predictions produced is known as the error rate, which serves as a gauge of a model's or system's accuracy. A frequently employed statistic to evaluate the performance of a classifier is the number of errors (false positives plus false negatives) divided by the total number of occurrences.

$$ER = \frac{FP + FN}{TP + TN + FP + FN} \tag{3.22}$$

3.3.7 Mathew's Correlation Coefficient

The effectiveness of binary (two-class) classifications is gauged using Matthew's Correlation Coefficient (MCC). A number between -1 and 1, where 1 is a perfect forecast, 0 is an average random prediction, and -1 is an inverse prediction, is provided after taking into account true and false positives and negatives.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{3.23}$$

MCC = +1: A MCC score of +1 indicates a perfect prediction, meaning the model has accurately classified all instances in the dataset.

MCC > 0: A positive MCC score suggests that the model's predictions are better than random chance. The higher the MCC, the stronger the correlation between predicted and true labels.

MCC = 0: A MCC score of 0 implies that the model's predictions are no better than random chance. The model shows no correlation between predicted and true labels.

-1 < MCC < 0: A negative MCC score indicates an inverse correlation between predicted and true labels. The lower the MCC, the stronger the inverse correlation.

MCC = -1: A MCC score of -1 signifies a completely incorrect prediction. The model's predictions are in total disagreement with the true labels.

3.3.8 Geometric Mean

A geometric mean is a single scalar number that represents the overall effectiveness of a binary classifier. The classifier's true positive, false positive, true negative, and false negative predictions are shown in a table called the confusion matrix. The square root of the product of true positive rate and true negative rate is used to determine the geometric mean of the confusion matrix.

$$GM = \sqrt{SNS * SPC} \tag{3.24}$$

Geometric Mean = 1: A geometric mean of 1 suggests that the model's overall performance is similar to random chance. It means that the model is not able to effectively discriminate between the different classes in the dataset.

Geometric Mean > 1: A geometric mean greater than 1 indicates that the model's performance is better than random chance. The higher the geometric mean, the better the overall classification performance.

Geometric Mean = 0: A geometric mean of 0 typically means that one or more of the classes in the dataset were not predicted correctly by the model. It suggests that the model is unable to capture the patterns or features necessary to classify certain classes accurately.

Geometric Mean < 1: A geometric mean less than 1 suggests that the model's performance is worse than random chance. It means that the model's predictions are inversely correlated with the true labels.

3.3.9 Bookmaker Informedness

Bookmaker Informedness refers to the degree to which the odds set by a bookmaker accurately reflect the true probabilities of an event's outcomes. In other words, it is a measure of how "informed" the bookmaker is about the event they are betting on.

$$BI = SNS + SPC - 1 \tag{3.25}$$

Bookmaker Informedness = 0: A value of 0 indicates that the classifier's performance is equivalent to random chance. It means the classifier is not providing any useful information for making predictions.

0 ¡ Bookmaker Informedness < 1: A value between 0 and 1 suggests that the classifier is providing some useful information but is not highly informative. The closer the value is to 1, the more informative the classifier.

Bookmaker Informedness = 1: A value of 1 signifies perfect informedness, indicating that the classifier is providing complete and accurate information for making predictions.

Bookmaker Informedness > 1: A value greater than 1 suggests that the classifier is more than perfectly informed. However, this scenario is not practically achievable, as it would require the classifier to exceed the limits of information provided by the data.

Chapter 4

Results and Discussion

In the Section "Results and discussion", we present a discussion of the metrics implemented in this tool, by analyzing their properties and relating them to the properties of the segmentation as well as to the requirements of the segmentation algorithms. We draw conclusions from this study on the best metrics to use for certain image data and segmentation jobs. We evaluated a variety of potential Neural Network Framework instantiates in this Chapter, to assess the usefulness of each model. We apply three models to the image classification problem using the Melanoma Skin Cancer based on the Image data set. In the preceding experiments, all models are trained on the training set and tested on the testing set. Our research is based on widely used neural network approaches such as CNN, ResNet-18, and EfficientNET-B0 architecture, which perform well on this image data set and have been demonstrated to generalize well to a variety of other domains. Its capacity to generalize the observed patterns and correlations from the training data to current, unobserved images is what determines how well it can convert colored images to grayscale. This generalization enables the network to efficiently represent the intensity information available in RGB images in gray-scale output. An important area of research in the field of images is the automated chromatic coloring of grayscale images. With the advancement of throughout the previous many years, deep learning, the automatic coloring of grayscale images has progressively become a reality. The deep learning model often outperforms the straightforward coloring approach. We employ 56 example skin lesions to assess the segmentation method's sensitivity to its input parameters as shown in Fig 4.1. It highlights the effectiveness of the gradient- and feature-based adaptive contour model-based melanoma image segmentation approach.

(a) RGB Image   (b) Gray Image
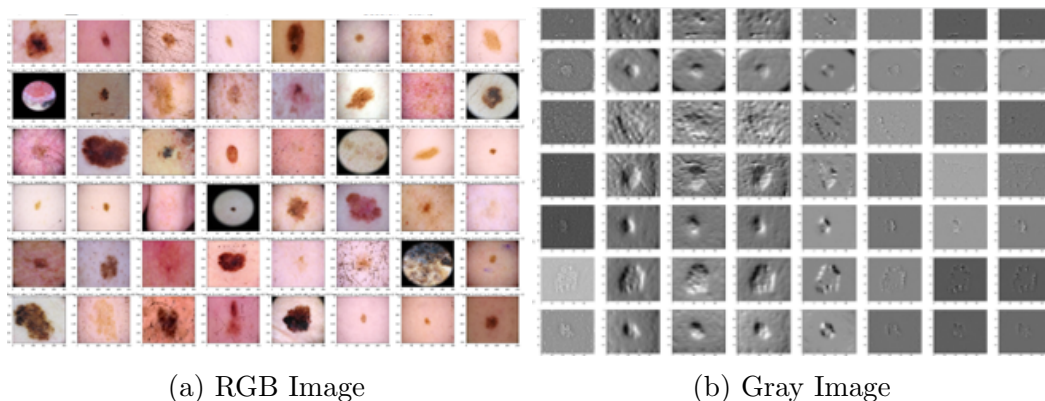
Figure 4.1: Filters extracting from the first layer of a trained Melanoma skin cancer image data set on raw (RGB) input (a) in gray-scale output (b).

To obtain the matrix representation, the image is typically passed through several layers of the neural network, each layer performing a computation to extract features at different levels of abstraction. The network's last layer generates a matrix that effectively and concisely depicts the picture. In order for the neural network employed for this task to become adept at identifying patterns and characteristics in photos, it must be trained on a sizable collection of images. The training process involves adjusting the weights and biases of the network based on the input and output, such that the network is able to accurately classify images based on their content. converting an image into a matrix through NN in Fig 4.2 is a powerful and widely used technique in computer vision and machine learning,

providing a way to represent and process images in a meaningful and computationally efficient way.
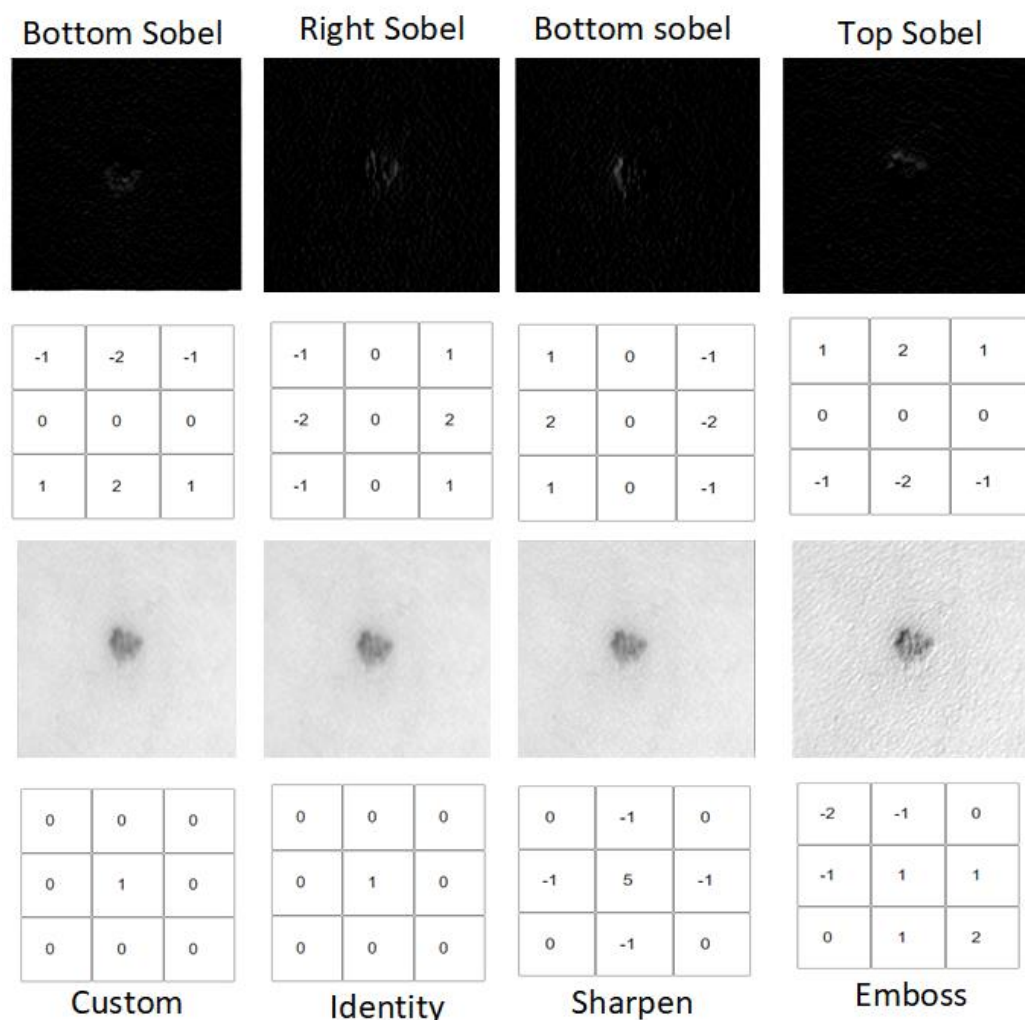
Figure 4.2: Conversion of images in the matrix form with different kernels(Filters).

The mentioned studies explore the application of various machine learning and neural networks techniques for the classification and diagnosis of skin cancer and melanoma. Different models such as PCA and SVM, CNNs (including EfficientNet-B0, ResNet-18, ResNet-50, VGG-13, and hybrid CNNs), and neural networks with additional input branches have been employed to analyze medical image data sets and detect skin cancer with an accuracy range 75%-93% and the other evaluation metrics such as sensitivity, specificity range 52%-95%, and 90%-95%. EfficientNet-B0, ResNet-50, and spiking VGG-13 has shown promising results in detecting malignancies and skin lesions. Hybrid CNNs with SVM classifiers have been used to classify dermoscopy pictures as benign or malignant

lesions, while a dermoscopy-based ML diagnostic system with ABCD rules for prepossessing, segmentation, feature extraction, and classification have been developed. On the MSC image data set, a number of as- assessment metrics, such as accuracy, sensitivity, specificity, f1-score, precision, error rate, Mathew's correlation, geometric mean, and bookmaker informedness was utilized to assess a model's performance is shown in Table 4.1.

Metrics	CNN	ResNet-18	EfficientNet-B0
Accuracy	0.80	0.87	0.97
Sensitivity	0.64	0.82	0.99
Specificity	0.95	0.93	0.93
F1-Score	0.76	0.86	0.97
Precision	0.93	0.91	0.95
Error Rate	0.19	0.12	0.03
Mathew's Correlation	0.75	0.82	0.94
Geometric Mean	0.77	0.84	1.01
Bookmaker Informedness	0.60	0.75	0.92

Table 4.1: Performance of Neural Networks on Melanoma skin cancer image data set.

The estimated Type-I error is the cases incorrectly classified as positive out of all instances incorrectly labeled as negative. the calculated Type-I error from ENetB0 instances 6% actually negative are incorrectly classified as positive by the model, also the results from CNN and RNet-18 are 4% and 6%, respectively. Type II Error from a melanoma skin cancer image data set using the ENet-B0 model incorrectly classifies 1% of actual melanoma-positive images as melanoma-negative, The model demonstrates a moderate tendency to misclassify non-melanoma images. However, it excels with a high sensitivity of 99%, leading to a low Type 2 error rate of 1%, effectively identifying most actual melanoma cases while from Cnn and ResNet-18, the model incorrectly classifies 3.6% and 1.8% shown in Fig 4.3.

	CNN	RNet-18	ENet-B0
Type-I	0.04	0.06	0.06
Type-II	0.36	0.18	0.01

Table 4.2: Showcase the performance of MSC image classification model.

It can be seen that the classification performance of EfficientNet-B0 is significantly better than other mainstream models ResNet-18 and CNN if we focus on accuracy. The calculated accuracy represents the percentage of skin cancer images that were correctly

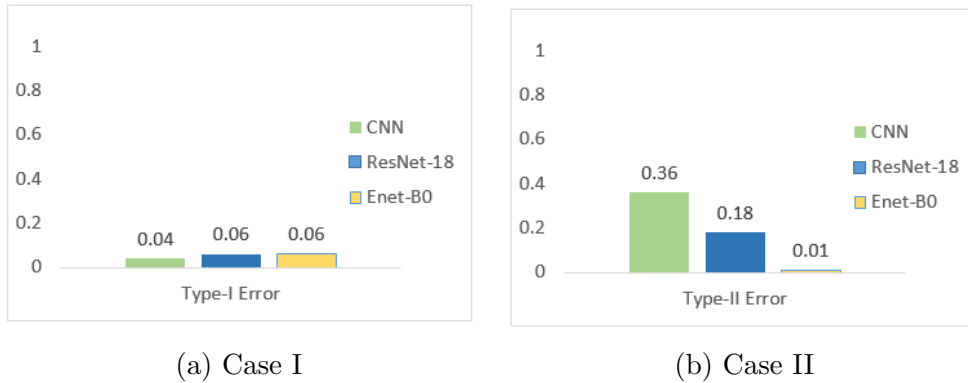(a) Case I   (b) Case II

Figure 4.3: This graph shows the performance of the MSC image classification model. With a Type I and Type II

classified by the different models. It provides an estimation of how well the model can distinguish between benign and malignant cases, so the EfficientnetB0 model accurately classified 97% of the skin cancer images in the evaluation set whereas the ResNet-18 model accurately classified 87% and the basic CNN model classified 80% in terms of accuracy. The percentage of successfully diagnosed malignant instances by the model is represented by the estimated sensitivity. It indicates how well the model is able to detect actual positive cases of skin cancer. In this study, our primary research objective was to investigate the classification of MSC neural networks, with a focus on comparing the performance of three models such as CNN, ResNet-18, and EfficientNet-B0. Our findings shed light on the effectiveness of these models and their potential applications in the field of dermatology. EfficientNet-B0 demonstrated the highest accuracy of 97% in classifying melanoma skin cancer, outperforming both CNN and ResNet-18, which achieved accuracies of 80% and 87%, respectively shown in Fig 4.4 to 4.6b.

EfficientNet-B0 has a 59.9% prevalence of melanoma skin cancer, in this data set, 59% participants have received a melanoma skin cancer malignant diagnosis, and 41% have benign. The model's ability to learn complex representations from image data, combined with its relatively smaller parameter size compared to other models, likely contributed to its superior performance. It is essential to consider the performance metrics, such as sensitivity, specificity, F1-Score, precision, error rate, Mathew's correlation coefficients, geometric mean, and bookmaker informedness in Fig 4.7.
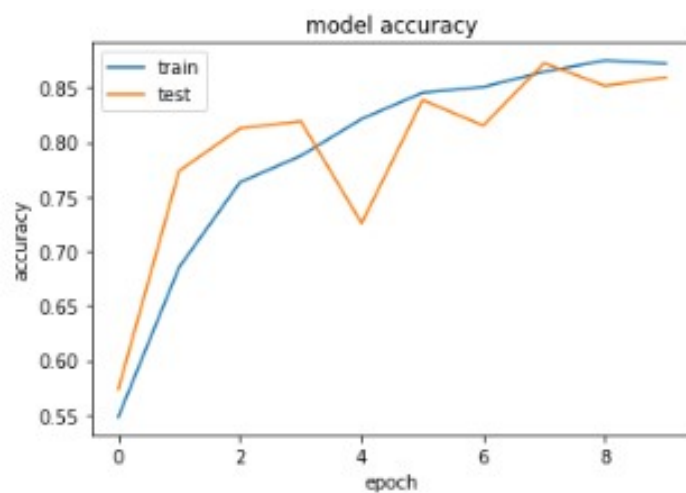
(a) Accuracy
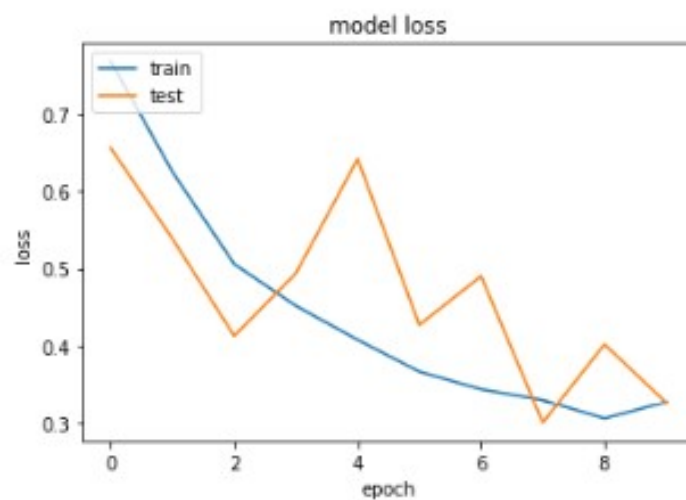
(b) Loss

Figure 4.4: Accuracy and Loss of the Test and Train data set when employing the basic CNN model are shown in this graph. The number of model training epochs is indicated on the x-axis of images (a) and (b) and the y-axis displays the accuracy in the image (a) and loss in the image (b).
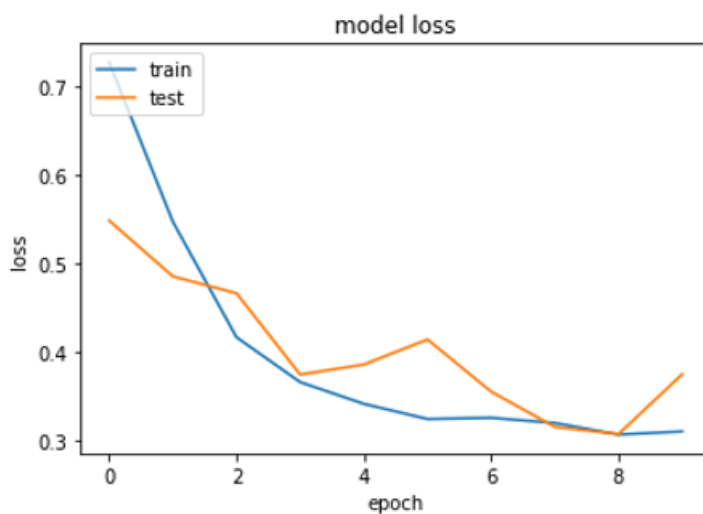
(a) Accuracy

(b) Loss

Figure 4.5: Accuracy and Loss of the Test and Train data set when employing the basic ResNet-18 model are shown in this graph. The number of model training epochs is indicated on the x-axis of images (a) and (b) and the y-axis displays the accuracy in the image (a) and loss in the image (b).
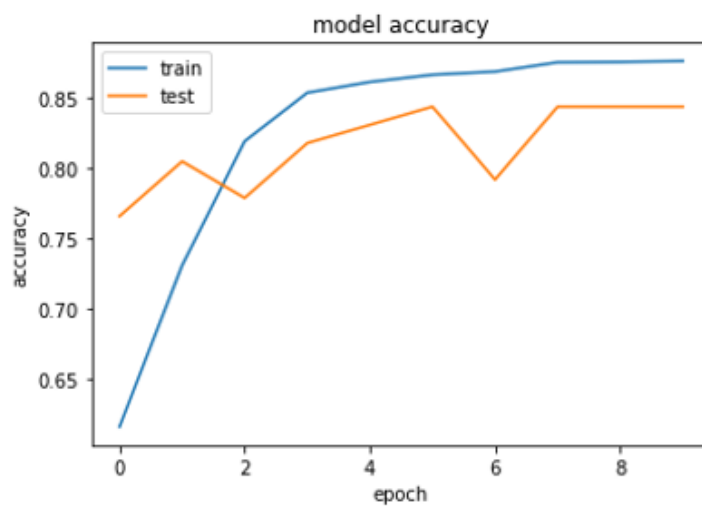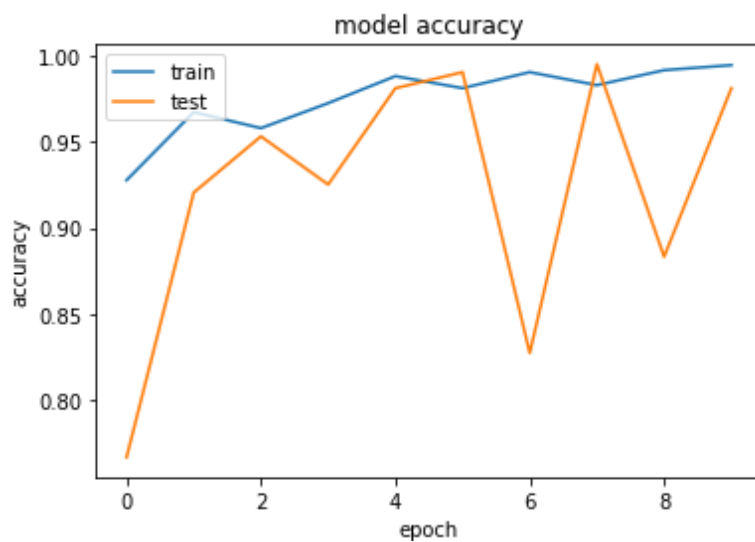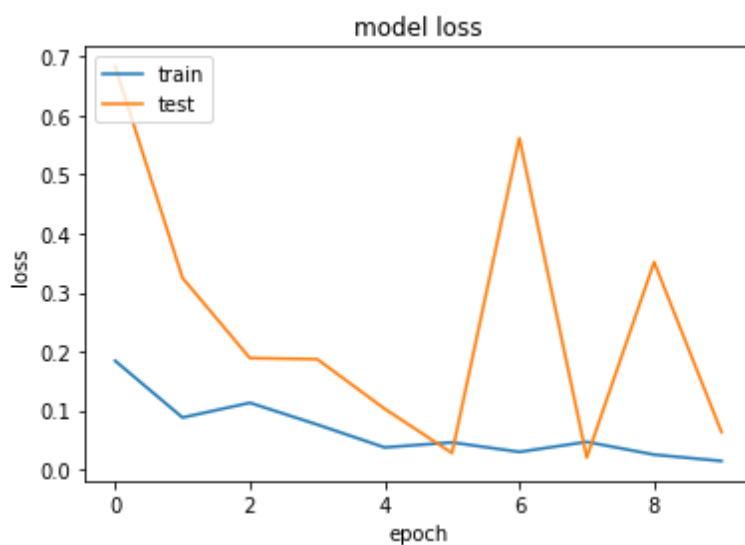
(a) Accuracy

(b) Loss

Figure 4.6: Accuracy and Loss of the Test and Train data set when employing the basic EfficientNet-B0 model are shown in this graph. The number of model training epochs is indicated on the x-axis of images (a) and (b) and the y-axis displays the accuracy in the image (a) and loss in the image (b).

(a) Sensitivity

(b) Specificity

(c) F1 Score

(d) Precision

(e) Error Baar

(f) Mathew's CC

(g) Geometric Mean

(h) Bookmaker Informedness

Figure 4.7: The graph depicts the performance of a melanoma skin cancer classification model over epochs. In the image (a) attains 99% sensitivity and (b) attains 93% specificity. The F1-score and precision are 97% and 95% in images (c) and (d) respectively, with a 3% error rate in (e). The Matthews correlation coefficient is 91%, indicating robust handling of imbalanced data in (f). The model achieves a balanced classification performance with a geometric mean of 1.01 in (g) and effectively captures true positive and true negative rates, as evidenced by a Bookmaker informedness of 92% in (h). The model demonstrates high accuracy, sensitivity, and specificity, making it a valuable tool for early melanoma detection and diagnosis.

	1	2	3	4	5	6	7	8	9	10
CNN	0.845	0.865	0.863	0.85	0.857	0.81	0.842	0.85	0.852	0.847
ResNet-18	0.616	0.731	0.819	0.854	0.861	0.866	0.869	0.875	0.87	0.87
EffNet-B0	0.927	0.967	0.958	0.972	0.988	0.981	0.99	0.983	0.991	0.994

Figure 4.8: This graph uses three different colors to display the model accuracy for each method. The number of epochs is displayed on the horizontal axis, and the vertical axis measures model accuracy. ResNet18 and CNN are represented by the blue and green lines, and the yellow line shows the EfficientNet-B0 method.

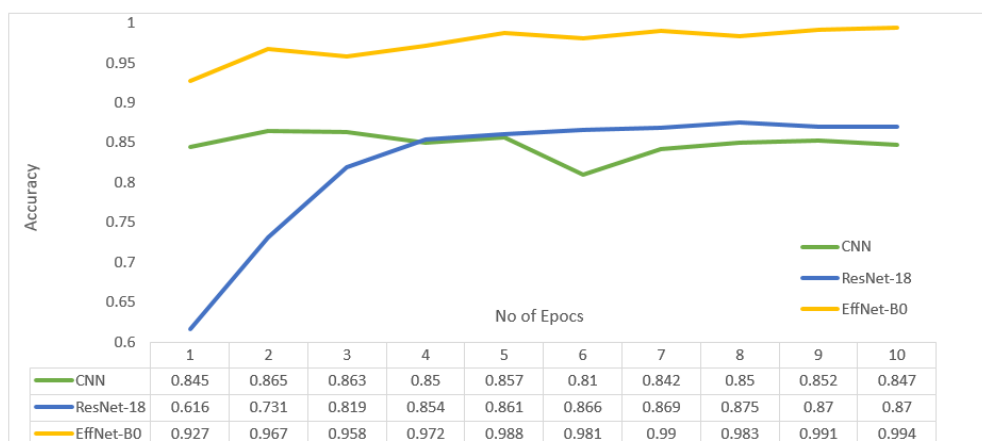# Chapter 5

Conclusion

We successfully applied the CNN basic model, ResNet-18, and EfficientNet-B0 on the Melanoma skin cancer image data set for classification. Our aim was to identify the most effective model among the three for accurately distinguishing between malignant and benign skin lesions. Upon meticulous evaluation and thorough analysis, EfficientNet-B0 emerged as the clear front-runner, demonstrating remarkable performance across various evaluation metrics. Notably, it achieved 97% accuracy, surpassing both CNN and ResNet-18, which achieved 80% and 87%. This strong accuracy implies the model's ability to correctly classify skin lesions with high precision. Moreover, EfficientNet-B0 displayed outstanding sensitivity, indicating its proficiency in detecting 99% true positive cases of melanoma with a 1% error rate. This is a crucial aspect of melanoma diagnosis, as improving patient prognosis and treatment results depends heavily on early identification. Equally important, EfficientNet-B0 exhibited excellent specificity, reducing false positive rates for non-melanoma cases. This implies a lower likelihood of misclassifying benign lesions as malignant, thereby minimizing unnecessary anxiety and medical interventions for patients. Furthermore, the F1-Score, precision, Mathew's correlation coefficient, and other key metrics consistently demonstrated the well-balanced and robust performance of EfficientNet-B0. Its ability to handle imbalanced data and maintain high predictive accuracy across both classes strengthens its suitability for melanoma classification. Based on the conclusive findings of this study, EfficientNet-B0 emerged as the most effective and promising model for melanoma skin cancer classification, surpassing CNN and ResNet-18 in accuracy and other key metrics. Its robust performance and potential for clinical use highlight its significance in improving melanoma diagnosis and patient care.

Bibliography

[1] R. A. Abumalloh, M. Nilashi, M. Y. Ismail, A. Alhargan, A. Alghamdi, A. O. Alzahrani, L. Saraireh, R. Osman, and S. Asadi, "Medical image processing and covid-19: a literature review and bibliometric analysis," *Journal of infection and public health*, vol. 15, no. 1, pp. 75–93, 2022.

[2] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, W. Enbeyle, *et al.*, "Deep neural networks for medical image segmentation," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[3] T. Saba, S. Al-Zahrani, and A. Rehman, "Expert system for offline clinical guidelines and treatment," *Life Science Journal*, vol. 9, no. 4, pp. 2639–2658, 2012.

[4] S. Pang, X. Yang, *et al.*, "Deep convolutional extreme learning machine and its application in handwritten digit classification," *Computational intelligence and neuroscience*, vol. 2016, 2016.

[5] P. Dhruv and S. Naskar, "Image classification using convolutional neural network (cnn) and recurrent neural network (rnn): A review," *Machine Learning and Information Processing: Proceedings of ICMLIP 2019*, pp. 367–381, 2020.

[6] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, "Transformers in medical imaging: A survey," *Medical Image Analysis*, p. 102802, 2023.

[7] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustainability*, vol. 13, no. 3, p. 1224, 2021.

[8] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019.

[9] J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo, and N. Kim, "Deep learning in medical imaging: general overview," *Korean journal of radiology*, vol. 18, no. 4, pp. 570–584, 2017.

[10] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, no. 2, pp. 505–515, 2017.

[11] P. Lambin, R. T. Leijenaar, T. M. Deist, J. Peerlings, E. E. De Jong, J. Van Timmeren, S. Sanduleanu, R. T. Larue, A. J. Even, A. Jochems, *et al.*, "Radiomics: the bridge between medical imaging and personalized medicine," *Nature reviews Clinical oncology*, vol. 14, no. 12, pp. 749–762, 2017.

[12] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *Machine Vision and Applications*, vol. 31, pp. 1–18, 2020.

[13] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, pp. 565–571, Ieee, 2016.

[14] K. Vaidhya, S. Thirunavukkarasu, V. Alex, and G. Krishnamurthi, "Multi-modal brain tumor segmentation using stacked denoising autoencoders," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers 1*, pp. 181–194, Springer, 2016.

[15] H. Yao, H. Derksen, J. R. Golbus, J. Zhang, K. D. Aaronson, J. Gryak, and K. Najarian, "A novel tropical geometry-based interpretable machine learning method: Pilot application to delivery of advanced heart failure therapies," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 1, pp. 239–250, 2022.

[16] Y. Le Lu, "Deep learning for medical image analysis," *Yearbook of Medical Informatics*, vol. 26, no. 01, pp. 08–10, 2017.

[17] G. Langs, X. Li, M. A. Zuluaga, L. Zhang, M. Sonka, S. Ourselin, S. Li, C. M. Nambakhsh, K. Chang, R. Kwong, K. Nikolaou, L. Umutlu, B. Glocker, B. Li, T. Tan, P. Hu, H. Chen, Q. Dou, J. Cheng, and P.-A. Heng, "Deep learning in radiology: An overview of recent advances," *European Journal of Radiology*, vol. 114, pp. 81–87, 2019.

[18] Y. Wang, B. Peng, E. Zheng, X. Li, Y. Jin, and D. Ni, "Deep fusion of medical imaging data: A comprehensive review," *Artificial Intelligence in Medicine*, vol. 115, p. 102052, 2021.

[19] R. Shanmugamani, R. H. Thirunavukarasu, P. Vadakkepat, and U. Raghavendra, "Medical image synthesis with context-aware generative adversarial networks (cagans)," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[20] Y. Zhang, J. Wu, W. Huang, Y. Chen, E. X. Wu, and X. Tang, "Utility of brain parcellation in enhancing brain tumor segmentation and survival prediction," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6*, pp. 391–400, Springer, 2021.

[21] S. Matej and S. D. Metzler, "15th international meeting on fully three-dimensional image reconstruction in radiology and nuclear medicine," *Medicine*, vol. 1107201, p. 28, 2019.

[22] N. Musa, A. Y. Gital, N. Aljojo, H. Chiroma, K. S. Adewole, H. A. Mojeed, N. Faruk, A. Abdulkarim, I. Emmanuel, Y. Y. Folawiyo, *et al.*, "A systematic review and meta-data analysis on the applications of deep learning in electrocardiogram," *Journal of ambient intelligence and humanized computing*, pp. 1–74, 2022.

[23] P. Azad, A. Kumar, and R. Goecke, "U-net plus: A modified encoder-decoder u-net architecture for semantic and instance segmentation of surgical instruments in laparoscopic images," *Computer Methods and Programs in Biomedicine*, vol. 210, p. 106303, 2021.

[24] C. Jung, M. Park, and J. Park, "Predictive modeling using long short-term memory networks for intensive care unit mortality," *Healthcare Informatics Research*, vol. 26, no. 3, pp. 158–165, 2020.

[25] K. Das, C. J. Cockerell, A. Patil, P. Pietkiewicz, M. Giulini, S. Grabbe, and M. Goldust, "Machine learning and its application in skin cancer," *International Journal of Environmental Research and Public Health*, vol. 18, no. 24, p. 13409, 2021.

[26] H. M. Gloster Jr and K. Neal, "Skin cancer in skin of color," *Journal of the American Academy of Dermatology*, vol. 55, no. 5, pp. 741–760, 2006.

[27] L. Zeng, B. J. Gowda, M. G. Ahmed, M. A. Abourehab, Z.-S. Chen, C. Zhang, J. Li, and P. Kesharwani, "Advancements in nanoparticle-based treatment approaches for skin cancer therapy," *Molecular Cancer*, vol. 22, no. 1, p. 10, 2023.

[28] C. Fitzmaurice, D. Abate, N. Abbasi, G. B. of Disease Cancer, *et al.*, "Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: a systematic analysis for the global burden of disease study," *JAMA Oncology*, vol. 5, pp. 1749–1768, 2019.

[29] P. Boyle, B. Levin, *et al.*, *World cancer report 2008.* IARC Press, International Agency for Research on Cancer, 2008.

[30] M. Fransen, A. Karahalios, N. Sharma, D. R. English, G. G. Giles, and R. D. Sinclair, "Non-melanoma skin cancer in australia," *Medical Journal of Australia*, vol. 197, no. 10, pp. 565–568, 2012.

[31] E. de Vries, M. Trakatelli, D. Kalabalikis, *et al.*, "Epidemiology and aetiology of basal cell carcinoma," *British Journal of Dermatology*, vol. 157, no. 2, pp. 47–51, 2007.

[32] E. Perera, N. Gnaneswaran, C. Staines, *et al.*, "Epidemiology and etiology of squamous cell carcinoma," *Current Dermatology Reports*, vol. 1, no. 1, pp. 39–49, 2012.

[33] J. A. Jaleel, S. Salim, and R. B. Aswin, "Computer aided detection of skin cancer," in *2013 International Conference on Circuits, Power and Computing Technologies (ICCPCT)*, pp. 1137–1142, 2013.

[34] S. P. Leong, M. C. Mihm Jr, G. F. Murphy, *et al.*, "Cutaneous melanoma: Etiology and therapy," *Clinical Dermatology*, vol. 24, no. 3, pp. 216–225, 2006.

[35] L. E. Davis, S. C. Shalin, and A. J. Tackett, "Current state of melanoma diagnosis and treatment," *Cancer biology & therapy*, vol. 20, no. 11, pp. 1366–1379, 2019.

[36] S. Jain, N. Pise, *et al.*, "Computer aided melanoma skin cancer detection using image processing," *Procedia Computer Science*, vol. 48, pp. 735–740, 2015.

[37] X.-X. Yin, L. Sun, Y. Fu, R. Lu, Y. Zhang, *et al.*, "U-net-based medical image segmentation," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[38] M. L. Kripke, "Impact of ozone depletion on skin cancers," *The Journal of dermatologic surgery and oncology*, vol. 14, no. 8, pp. 853–857, 1988.

[39] A. Katalinic, U. Kunze, and T. Schäfer, "Epidemiology of cutaneous melanoma and non-melanoma skin cancer in schleswig-holstein, germany: incidence, clinical subtypes, tumour stages and localization (epidemiology of skin cancer)," *British Journal of Dermatology*, vol. 149, no. 6, pp. 1200–1206, 2003.

[40] S. García, A. Fernández, J. Luengo, and F. Herrera, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 31, no. 2, pp. 352–374, 2009.

[41] M. M. Hasan, M. U. Islam, M. J. Sadeq, W.-K. Fung, and J. Uddin, "Review on the evaluation and development of artificial intelligence for covid-19 containment," *Sensors*, vol. 23, no. 1, p. 527, 2023.

[42] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1, pp. 886–893, Ieee, 2005.

[43] M. B. Rozenwald, A. A. Galitsyna, G. V. Sapunov, E. E. Khrameeva, and M. S. Gelfand, "A machine learning framework for the prediction of chromatin folding in drosophila using epigenetic features," *PeerJ Computer Science*, vol. 6, p. e307, 2020.

[44] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," in *2017 15th international conference on ICT and knowledge engineering (ICT&KE)*, pp. 1–6, IEEE, 2017.

[45] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1310–1315, 2016.

[46] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[47] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[48] V. S. Chouhan, S. Saini, and S. K. Saini, "Automated medical image classification and retrieval using hybrid features," *Multimedia Tools and Applications*, vol. 77, no. 23, pp. 31363–31383, 2018.

[49] S. Kumar and I. Kaur, "Ensemble methods for medical image classification: A review," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1104–1119, 2020.

[50] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," *arXiv preprint arXiv:1901.07031*, 2019.

[51] J. J. Li and X. Tong, "Statistical hypothesis testing versus machine learning binary classification: distinctions and guidelines," *Patterns*, vol. 1, no. 7, p. 100115, 2020.

[52] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," *arXiv preprint arXiv:2008.05756*, 2020.

[53] Z. B. Kizilkan, M. S. Sivri, I. Yazici, and O. F. Beyca, "Neural networks and deep learning," in *Business Analytics for Professionals*, pp. 127–151, Springer, 2022.

[54] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[55] 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 (CVPR)*, pp. 770–778, IEEE, 2016.

[56] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[57] J. Frankle and M. Carbin, "The lottery ticket hypothesis: Finding sparse, trainable neural networks," *arXiv preprint arXiv:1803.03635*, 2018.

[58] A. Adeel, M. Gogate, and A. Hussain, "Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments," *Information Fusion*, vol. 59, pp. 163–170, 2020.

[59] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2, pp. 1150–1157, Ieee, 1999.

[60] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, pp. 1–74, 2021.

[61] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[62] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition*, vol. 77, pp. 354–377, 2018.

[63] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[64] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[65] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2012.

[67] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," in *Proceedings of the 2011 International Joint Conference on Neural Networks (IJCNN)*, pp. 2809–2813, IEEE, 2011.

[68] R. A. Al-Jawfi, "Activation functions effect on fractal coding using neural networks.," *Intelligent Automation & Soft Computing*, vol. 36, no. 1, 2023.

[69] H. Alquran, I. A. Qasmieh, A. M. Alqudah, S. Alhammouri, E. Alawneh, A. Abughazaleh, and F. Hasayen, "The melanoma skin cancer detection and classification using support vector machine," in *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1–5, 2017.

[70] R. Refianti, A. B. Mutiara, and R. P. Priyandini, "Classification of melanoma skin cancer using convolutional neural network," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 3, 2019.

[71] H. A. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul, and J.-M. Kang, "A robust approach for brain tumor detection in magnetic resonance images using finetuned efficientnet," *IEEE Access*, vol. 10, pp. 65426–65438, 2022.

[72] M. Guo and Y. Du, "Classification of thyroid ultrasound standard plane images using resnet-18 networks," in *2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, pp. 324–328, 2019.

[73] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer," *Procedia Computer Science*, vol. 179, pp. 423–431, 2021.

[74] S. Qasim Gilani, T. Syed, M. Umair, and O. Marques, "Skin cancer classification using deep spiking neural network," *Journal of Digital Imaging*, pp. 1–11, 2023.

[75] M. Obayya, A. Alhebri, M. Maashi, A. S. Salama, A. Mustafa Hilal, M. I. Alsaid, A. E. Osman, and A. A. Alneil, "Henry gas solubility optimization algorithm based feature extraction in dermoscopic images analysis of skin cancer," *Cancers*, vol. 15, no. 7, p. 2146, 2023.

[76] D. Keerthana, V. Venugopal, M. K. Nath, and M. Mishra, "Hybrid convolutional neural networks with svm classifier for classification of skin cancer," *Biomedical Engineering Advances*, vol. 5, p. 100069, 2023.

[77] N. S. Zghal and N. Derbel, "Melanoma skin cancer detection based on image processing," *Current Medical Imaging*, vol. 16, no. 1, pp. 50–58, 2020.

[78] S. Bonechi, M. Bianchini, P. Bongini, G. Ciano, G. Giacomini, R. Rosai, L. Tognetti, A. Rossi, and P. Andreini, "Fusion of visual and anamnestic data for the classification of skin lesions with deep learning," in *New Trends in Image Analysis and Processing–ICIAP 2019: ICIAP International Workshops, BioFor, PatReCH, e-BADLE, Deep-Retail, and Industrial Session, Trento, Italy, September 9–10, 2019, Revised Selected Papers 20*, pp. 211–219, Springer, 2019.

[79] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution efficientnets with meta data," *MethodsX*, vol. 7, p. 100864, 2020.

[80] H. Mhaske and D. Phalke, "Melanoma skin cancer detection and classification based on supervised and unsupervised learning," in *2013 international conference on Circuits, Controls and Communications (CCUBE)*, pp. 1–5, IEEE, 2013.

[81] M. Vijayalakshmi, "Melanoma skin cancer detection using image processing and machine learning," *International Journal of Trend in Scientific Research and Development (IJTSRD)*, vol. 3, no. 4, pp. 780–784, 2019.

[82] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, "Seven-point checklist and skin lesion classification using multitask multimodal neural nets," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 538–546, 2018.

[83] W. Li, J. Zhuang, R. Wang, J. Zhang, and W.-S. Zheng, "Fusing metadata and dermoscopy images for skin disease diagnosis," in *2020 IEEE 17th international symposium on biomedical imaging (ISBI)*, pp. 1996–2000, IEEE, 2020.

[84] M. Elgamal, "Automatic skin cancer images classification," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 3, 2013.

[85] J. Yap, W. Yolland, and P. Tschandl, "Multimodal skin lesion classification using deep learning," *Experimental dermatology*, vol. 27, no. 11, pp. 1261–1267, 2018.

[86] D. Moldovan, "Transfer learning based method for two-step skin cancer images classification," in *2019 E-Health and Bioengineering Conference (EHB)*, pp. 1–4, 2019.