

Skin Cancer Classification Using Deep Learning

Abdalla Suleiman Sufi
012201023

A project/thesis in the Department of Computer Science and Engineering presented
in partial fulfillment of the requirements for the Degree of
Bachelor/Master of Science in Computer Science and Engineering



United International University
Dhaka, Bangladesh

MAY 2022

©Abdalla Suleiman Sufi, 2022

Declaration

I Abdalla Suleiman Sufi declare that this project titled, Skin cancer classification using deep learning and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a MSc degree at United International University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at United International University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Abdalla Suleiman Sufi
012201023

Certificate

I do hereby declare that the research works embodied in this thesis/project entitled “**Skin Cancer Classification Using Deep Learning**” is the outcome of an original work carried out by Abdalla Suleiman Sufi under my supervision.

I further certify that the dissertation meets the requirements and the standard for the degree of MSc in Computer Science and Engineering.

Mr. Mohammad Mamun Elahi
Assistant Professor, UIU

Abstract

Cancer begins when cells in the body begin to multiply uncontrolled. Cancer can develop in any area of the body and spread to other parts. Cancer of the skin is one of the most prevalent forms of the disease found all over the world, and it can cause substantial years of possible life loss and lost productivity. The primary cause of skin cancer is direct exposure to sunshine. Melanoma is a form of skin cancer, and because it is the most dangerous form of skin cancer, it must be discovered early in order to be effectively treated. It is essential to establish a computer-assisted diagnostic method for early detection. In this report, the proposed approach seeks to differentiate between melanoma and non-melanoma types. Using a dataset containing numerous forms of skin cancer, the system achieved an accuracy of 83% percent on the testing dataset.

Acknowledgement

At the beginning, I would like to express my deepest appreciation to the All-Mighty ALLAH (SWT). Then, peace and blessing upon the last and final prophet Mohamed. who is the one who led me to work in the appropriate direction throughout my life. To begin, I would like to express my appreciation to the All-Mighty ALLAH (SWT), who is the one who led me to work in the appropriate direction throughout my life. Words are often inadequate when it comes to revealing one's profound regards. My efforts never result in a comprehension of the work like this being achieved as a result of my efforts. I would want to take this occasion to convey to everyone who assisted me throughout the course of this job my sincere appreciation and admiration for their assistance. This would never have been the reality that it is today if it weren't for his grace. Without my respected boss Mr. Mohammad Mamun Elahi's support, encouragement, and direction throughout this project, it would not have been feasible to complete it. He was easily approachable and provided a consistent source of inspiration, motivation, and experienced-based information. The majority of the insights and strategies contained in this thesis are the product of the numerous thought-provoking conversations I've had. In addition to being quite helpful, his remarks and suggestions on the editing of this thesis were also of great importance.

I would like to take this opportunity to extend my heartfelt gratitude to each and every member of the department's faculty who assisted me in finishing this project in any way, whether it was directly or indirectly.

Table of Contents

LIST OF TABLES.....	VII
LIST OF FIGURES.....	VIII
CHAPTER1.....	1
INTRODUCTION	1
1.1 MELANOMA	1
1.2 LITERATURE REVIEW.....	4
1.3 PROPOSED SYSTEM ARCHITECTURE FOR IMAGE CLASSIFICATION	5
1.4 OBJECTIVES OF THE PROPOSED WORK	6
1.5 METHODOLOGY	6
1.7 DEVELOPMENT TOOLS	7
1.8 PROJECT ORGANIZATION.....	8
CHAPTER 2.....	9
MACHINE LEARNING	9
2.2 DIFFERENCE BETWEEN ML AND TRADITIONAL PROGRAMMING	9
2.3 SUPERVISED LEARNING	11
2.4 UNSUPERVISED LEARNING	13
2.5 SEMI-SUPERVISED MACHINE LEARNING	14
2.6 REINFORCEMENT MACHINE LEARNING.....	14
2.7 NEURAL NETWORK	14
2.8 CNN	14
2.9 UNDER-FITTING AND OVER-FITTING IN MACHINE LEARNING	18
2.10 DROPOUT REGULATION	20
2.11 REGULARIZATION.....	20
2.12 GOOD FIT IN A STATISTICAL MODEL	20
2.13 NORMALIZATION	21
2.14 PERFORMANCE METRICS FOR CLASSIFICATION PROBLEMS IN MACHINE LEARNING.....	21
2.14.1 Classification Accuracy.....	23
2.14.2 Logarithmic Loss	23
2.14.3 Confusion Matrix	24
2.14.4 Area Under Curve.....	24
2.14.5 F1 Score.....	25
2.14.6 Mean Absolute Error.....	1
2.14.7 Mean Squared Error.....	1
CHAPTER 3.....	3
DEVELOPMENT TOOLKITS.....	3
3.1 TENSORFLOW	3
3.2 KERAS	3
3.3 OPENCV	4
3.4 TENSORBOARD	6
3.5 JUPYTER NOTEBOOK.....	6
3.6 KAGGLE DATASETS	6
CHAPTER 4.....	1
METHODOLOGY	2
4.1 DATA COLLECTION	2
4.2 PRE-PROCESSING.....	1
4.3 IMAGE SCALING	1
4.4 GRAYSCALE CONVERSION	1
4.5 MEDIAN FILER	2
4.6 SEGMENTATION	4
4.7 FEATURE EXTRACTION	1
4.8 CLASSIFICATION.....	3

CHAPTER 5	3
EXPERIMENTAL RESULT AND DISCUSSION.....	4
5.1 EXPERIMENTAL RESULT	4
5.3 DISCUSSION.....	1
CHAPTER 6	2
CONCLUSION AND FUTURE WORK	2
6.1 CONCLUSION	2
6.2 FUTURE WORK.....	2
REFERENCES	3

LIST OF TABLES

Table 5. 1 Model accuracy.....	7
Table 5. 2 Sensitivity and Specificity	1

LIST OF FIGURES

Figure 1. Proposed system methodology.....	7
Figure 2. Traditional programming	10
Figure 3. Machine learning.....	11
Figure 4. Deep learning layers.....	16
Figure 5. Typical CNN Architecture	17
Figure 6. Types of pooling layers	18
Figure 7. Under-fitting and Overfitting	19
Figure 8. Area Under Curve graph	25
Figure 9. Grayscale Image.....	2
Figure 10. Median filter.....	3
Figure 11. Asymmetry sign	1
Figure 12. Un-even borders	2
Figure 13 Before & after data augmentation	5
Figure 14. Classification accuracy on training dataset	7
Figure 15. Classification on validation dataset.....	7
Figure 16. Classification on testing accuracy testing	8
Figure 17. Confusion matrix.....	1

Chapter1

Introduction

It is possible that skin cancer is the type of cancer that occurs the most frequently in the entire world., and its incidence is significantly rising. Just in the United States of America, for instance, current research indicates that one in every five people will acquire skin cancer at some point throughout their lifetime [1]. Melanoma and non-melanoma skin cancers are the two primary categories that can be utilized in order to categorize cases of skin cancer. The non-melanoma kind of skin cancer typically starts in the basal cells, which are found at the bottom of the outermost layer of the skin. Only in the United States are there around 1,200,000 people who have been diagnosed with this form of skin cancer. It is widely believed that basal and squamous cell cancers are caused by overexposure of the skin to sunlight, which is considered to be the primary factor in the development of these cancers. Cancers of the skin that are not melanoma can be cured if they are discovered and treated in their early stages. Melanoma is the most aggressive and potentially fatal form of skin cancer [1]. According to estimates provided by the World Health Organization (WHO), over 70 230 individuals pass away every year around the world as a direct result of excessive sun exposure [2].

Using the ABCD approach, dermatologists are able to correctly identify melanoma in approximately 80 percent of instances [3]. The use of digital dermatoscopy has the potential to provide physicians with a more comprehensive view of suspicious skin lesions.

1.1 Melanoma

Melanoma is a type of skin cancer that metastasizes when melanocytes (the cells that give the skin its color) start to grow uncontrollably. There are various types of skin cancers and the following are some of them.:

- Actinic keratosis
- Recurrent basal cell carcinoma

- Squamous cell carcinoma
- Basal cell carcinoma
- Merkel cell carcinoma
- Melanoma
- Kaposi sarcoma (KS)
- Lymphoma of the skin
- Keratoacanthoma

Melanoma is rarely diagnosed and it accounts for only 4 percent of all skin cancers diagnosed, Yet, it's by far the deadliest of all skin cancers as it's responsible for almost 80 percent of all skin cancer-related deaths. Patients diagnosed with metastatic melanoma, usually have less than a 15 percent likelihood of surviving five years after diagnosis. and this is increasing globally day by day thanks to a mixture of genetic and environmental risk elements. Malignant melanoma is a skin cancer that forms in the melanocytes – the cells that produce the skin coloring known as melanin that assists to protect the deeper layers of the skin from the destructive effects of the sun. Though Melanoma accounts for smaller than one percent of skin cancer cases but yields vast majority of skin cancer deaths.

Melanoma is rarely diagnosed and it accounts for only 4 percent of all skin cancers diagnosed, Yet, it's by far the deadliest of all skin cancers as it's responsible for almost 80 percent of all skin cancer-related deaths. Patients diagnosed with metastatic melanoma, usually have less than a 15 percent likelihood of surviving five years after diagnosis. and this is increasing globally day by day thanks to a mixture of genetic and environmental risk elements. Malignant melanoma is a skin cancer that forms in the melanocytes “ the cells that produce the skin coloring known as melanin that assists to protect the deeper layers of the skin from the destructive effects of the sun. While Melanoma accounts for smaller than one percent of skin cancer cases it yields the majority of skin cancer deaths.

Hippocrates was the first to write a definition of melanoma, which he defined as melas, which means black, and oma, which means tumor in Greek, during the 5th century BC.

When John Hunter performed the first successful excision of a melanoma tumor in 1787, he had no idea what it was and described it as a "cancerous fungous swelling." William Norris was the first to recognize the heterogeneity of some melanoma tumors in the 1820s. Robert Carswell coined the phrase "melanoma" in 1838.

Thomas Fawdington remarked in 1826 that the medical profession was completely ignorant of the distant and current causes of melanoma. As a result, it was critical to expose knowledge's inadequacy at the moment, leaving future scholars to discover the condition's controlling principles. Anesthesia and antiseptic chemicals were not available at the time, and there was little awareness of DNA and cancer formation. In 1844, Samuel Cooper was the first to openly recognize that developed or progressing melanoma was incurable and that the only hope for survival was to treat it early. This statement is still valid today, thanks to ongoing research and development.

In 1892, Herbert Snow described the wonders of eliminating the tumor and enclosing glands as a means of prophylaxis, arguing that merely removing the tumor was insufficient treatment and that the lymph glands, which are eager to acquire infective protoplasm, should be cleansed as well. William Handley investigated the lymphatic spread of secondary melanoma on a woman's leg between 1900 and 1905, resulting in a case report for the illness. He emphasized the need of removing the surrounding subcutaneous tissue and lymph nodes, which guided melanoma surgery for the next 50 years.

In 1956, Henry Lancaster was the first to link UV radiation from sunlight exposure to an increased risk of melanoma. Lancaster and Nelson's research supported up this theory, stating that skin texture, color, hair color, and eye color all had an impact on melanoma development. Sun-related reactions Melanoma risk was shown to be much higher among those with white skin who were exposed to high levels of UV light, particularly in Europe, Australia, and New Zealand.

Following this, a classification of distinct phases of disease progression was established, and numerous drugs for the treatment of melanoma were approved. Today

We have a fairly strong understanding of melanoma at this time, in that we know it is triggered by UV light and results from melanocyte malignancy. Melanoma risk is thought

to be increased in some people due to skin features and polymorphisms in the melanocortin receptor-1 (MC1R) or CDKN2A genes. However, the exact mechanism of a melanocyte's growth into melanoma is unknown, and as a result, preventative strategies are limited. Future research in this area would be beneficial.

1.2 Literature Review

Melanoma can afflict everyone, but Caucasians are more prone than other races to develop it. However, because it is less visible in those with dark skin, it is frequently discovered later, when the condition has progressed. This is due to the widespread misconception that individuals of color cannot get skin cancer [4]. It was the cause of the most cancer-related deaths worldwide in 2016, with an estimated 76,380 people diagnosed with invasive melanomas, with about 46,870 men and 29,510 women [5].

If diagnosed early, melanoma is highly curable, but advanced melanoma can spread to the lymph nodes and other organs, which can be fatal. For early and accurate melanoma detection, medical specialists and professional equipment are essential. In contrast, the lack of access to such expert judgments makes it difficult to provide effective care to those who are at risk of contracting the disease.

Typically, patients notice a skin anomaly first. Then medical professionals utilize dermoscopy to get a diagnosis. This is a high-resolution skin-imaging procedure that eliminates skin surface reflections and allows clinicians to view the deeper underlying tissues. They are utilized to examine the colors and microstructures of the epidermis, dermo-epidermal junction, and papillary dermis in vivo non-invasively [6]. This has opened up new possibilities for studying pigmented skin lesions and, in particular, detecting melanoma in its early stages [7], [8]. Specially trained medical specialists have demonstrated diagnostic accuracies of up to 75%–84% using this approach [9]. However, if doctors are not properly trained, their diagnostic performance suffers dramatically.

To address the problems posed by a lack of access to professionals, particularly in underdeveloped countries, extensive effort has been focused on building automated image processing systems that can detect skin illnesses using dermoscopy images. Several recent publications [9], [10], as well as dermoscopy papers proposing diagnostic criteria for early melanoma detection [11]-[12], have reviewed the various techniques

used. However, these criteria still include dermoscopy picture qualities that dermatologists or medical specialists can only assess.

Drs. Julie Ann A. Salido and Conrado Ruiz Jr.[13] proposed a method for detecting and eliminating hair from dermoscopy images, as well as utilizing deep learning to classify skin lesions. The hair detection technique and the classifier are then evaluated using photos from the PH2 dataset [14]. They were 82.5 percent accurate overall.

Also, M Manoj krishna, M Neelima, M Harshali, and M Venu Gopala Rao [15] discuss image classification as a classic problem of image processing, and they strongly recommend doing image classification using deep learning because of the incredible results they obtained when using deep learning in image classification. For this, they deploy AlexNet architecture and convolutional neural networks. For the classification, four test photos were chosen from the ImageNet collection. They experimented with cropping the photos for different portion locations. The results demonstrate the efficacy of deep learning-based picture classification using AlexNet, a ConvNet category.

ConvNet is divided into two categories: LeNet and AlexNet. The LeNet stands for Shallow Convolutional Neural Networks, and it's used to classify handwritten numbers. Two convolutional layers, two subsampling levels, two hidden layers, and one output layer make up the LeNet [5]. The AlexNet is a set of deep convolutional neural networks that classify an input image into one of thousands of categories.

AlexNet is used to tackle a wide range of issues, including indoor scene categorization, which is common in artificial neural networks. It is a powerful approach of understanding visual features with greater differential vision in the computer field for pattern detection. Their research examined the classification of a certain image size of their choice. For better vision, it can successfully classify the training sample of images in the AlexNet. AlexNet is made up of five convolutional layers, three subsampling layers, and three fully linked layers. The type of Feature Extractor is the key distinction between LeNet and AlexNet. In AlexNet's Feature Extractor module, they employ non-linearity, whereas in LeNet, they use Log Sigmoid. Dropout is used by AlexNet, which is not seen in other networking data sets. In [20] The "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge dataset[18-19] was used to develop the U-Net architecture and train it on 2594 dermoscopy images and ground truth response masks. It has been

demonstrated in [21] that the application of transfer learning in conjunction with DL models can achieve a high level of accuracy in the correct classification of various aneurysms that occur in the retinal region of the eye as a result of diabetic retinopathy.

For the purpose of skin cancer classification, [22] proposed two modified models: the modified VGG16 and modified InceptionV3 models. The application of the data augmentation demonstrated that the reduction of the data imbalance can be useful to improve classification performance; however, careful tuning is required, for instance, to make the data perfectly balanced; this does not necessarily result in a better model being created.

1.3 Proposed System Architecture For Image Classification

The proposed system uses CNN due to the advantage of CNN has in pattern recognition . Convolutional Neural Networks are extremely alike to other regular Neural Networks, which are composed of many neurons that have learnable weights. And Every neuron performs as a dot product by receiving some input and using bias. The entire convent still defines a distinct score function, from the raw pixels on one end to class scores on the other end.

The purposed system has two Convolutional layers with each layer following the Max pooling layer and also some loss function like Relu (Rectified Linear Unit) for the first layers and SoftMax in the last layer and that is a fully connected layer, the purposed system has two output layers (Melanoma/non-melanoma).

1.4 Objectives of the Proposed Work

The main goal of the purposed system is to be able to extract the appropriate features from raw images and learn the difference/unique features that each type has then generalize them so that the model is not biased to any particular type, and give each image/feature its appropriate label seen from the past experience (training dataset).

1.5 Methodology

The following Steps has being followed to build and train the purposed system/model:

1. Data Collection

2. Pre-processing
3. Segmentation
4. Feature Extraction
5. Classification

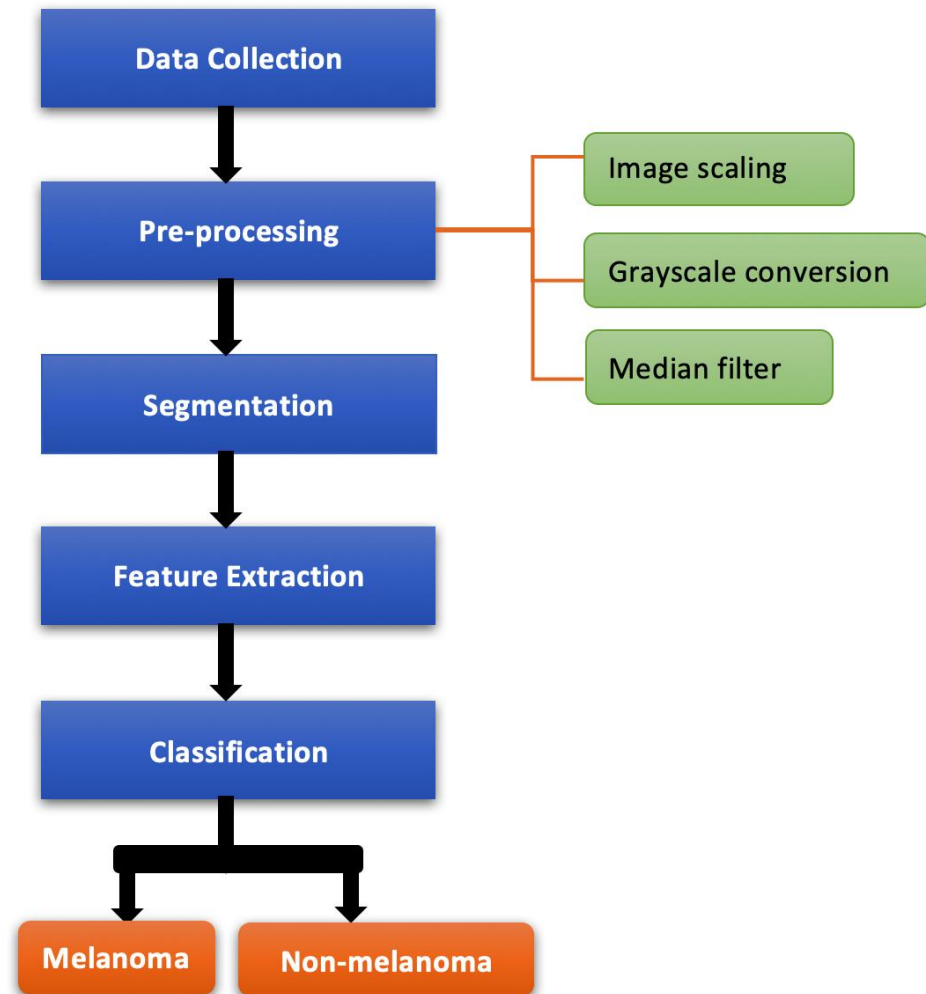


Figure 1. Proposed system methodology

1.7 Development Tools

➤ Software tools:

IDE : ananoda

language : python

Libraries : numpy, pandas, scikit-learn, OpenCV

➤ Hardware tools:

PC with 2.3 GHz processing power

1.8 Project Organization

The project is divided into five chapters. Each of which is made up of a number of sections. Chapter 1: covers the Introduction, Melanoma overview, literature review, proposed system, Objectives, methodology, Development tools in this thesis. Chapter 2: covers the introduction of various types of machine learning algorithms and techniques in AI. Chapter 3: covers the various development toolkits used building the model Chapter4: covers the methodology Chapter 5: provides the experimental result the proposed model generated Chapter 6: concludes and gives summary of what the current system can achieve and what will be the focus for the future work

Chapter 2

Machine Learning

2.1 What is Machine Learning ML?

Arthur Samuel, a leader in the fields of artificial intelligence, is credited with coining the term "machine learning." He defines "machine learning" as a field of study that gives computers the capacity to learn without being explicitly trained to do so. Machine Learning (ML) is a method that automates and refines the learning process of computers based on their experiences without the need for programming, i.e. without human interaction. This method may be characterized as the process of automating and perfecting computer learning. The approach begins with the collecting of high-quality data, which is then used to train our machines (computers) in a variety of ways, including the building of machine learning models based on the data. The types of data we have access to, as well as the actions we wish to automate, all assist us pick which algorithms to use.

Machine learning, also known as ML, is a field of study that entails the investigation of a large number of algorithms and statistical models. These models and algorithms are used by computer systems to carry out computational tasks without being explicitly instructed, relying instead on the patterns and inferences they have observed in the past.

It is possible to divide it into other subcategories, including artificial intelligence. In the process of machine learning, algorithms construct a mathematical model by using representative data, which is then referred to as "training data." The goal of this model is to allow the algorithm to make predictions or judgments without being explicitly programmed to do so. Machine learning algorithms are applied in a variety of applications, including email filtering and computer vision, when it is difficult or impossible to build a standard algorithm for correctly accomplishing the task.

The field of computational statistics, which focuses on using computers to make predictions, is closely connected to the field of machine learning. The discipline of machine learning has advanced thanks to the study of mathematical optimization since

this research provides the foundations, methods, and theory that are necessary for the field to advance. Data mining is another topic that focuses on unsupervised learning for exploratory data analysis. It is an important part of the field.

When it is applied to the resolution of business problems, machine learning can also be referred to as a predictive analytics-based learning system..

2.2 Difference Between ML and Traditional Programming

Traditional computer programming has existed for over a century, with the earliest known computer program dating back to the mid-nineteenth century. Traditional Programming refers to any manually generated program that processes input data and generates output on a computer. But for decades, a sophisticated style of programming has revolutionized industry, especially in intelligence and embedded analytics. In Machine Learning, commonly referred to as augmented analytics, input and output data are fed into an algorithm to generate a program. This produces insightful information that can be utilized to forecast future events. Traditional programming is a manual procedure, in which a human (the programmer) writes the code. In the absence of a programmer, however, rules must be manually formulated or coded.



Figure 2. Traditional programming

In contrast, in machine learning, the algorithm automatically derives the rules from the data. In contrast to conventional programming, machine learning is an automated procedure. It can boost the value of embedded analytics in numerous areas, such as data preparation, natural language interfaces, automatic outlier detection, recommendations, and detection of causation and relevance. All of these characteristics accelerate user insights and eliminate decision bias.



Figure 3. Machine learning

The difference between machine learning and traditional programming can be summarized as follows, Traditional programming expects both data and logic to be fed in order to get your desired output, while machine learning only requires the data and your desired output and it's intelligent enough to figure out the logic to reach the desired output.

2.3 Supervised Learning

Learning that is Supervised implies, by its very name, that there is a teacher or supervisor present. Simply said, supervised learning is a type of machine learning in which we teach or train the machine using data that has been accurately labeled, which implies that some of the data is already tagged with the appropriate response. After then, the machine is given a new collection of examples (data), which is done so that a supervised learning algorithm can analyze the training data (the set of training instances), and then create an accurate result based on labelled data.

Learning algorithms are given known quantities to work with in supervised machine learning systems so that they can make better future decisions. Systems like as chatbots, self-driving cars, facial recognition algorithms, expert systems, and robots are examples of some of the examples of systems that could use either supervised or unsupervised learning. Although retrieval-based artificial intelligence is the most common application

for supervised learning systems, generative learning may also be an option for these kinds of systems.

In a broad sense, supervised learning describes the process by which a system is provided with input and output variables with the goal of teaching itself how those variables are related to one another or mapped together. The goal is to develop a mapping function that is accurate enough so that when new input is given, the algorithm can predict what output will be produced. This is an iterative process, and each time the algorithm produces a prediction, it is either adjusted or provided feedback until it reaches a level of performance that may be considered satisfactory.

Training data for supervised learning consists of a set of examples, each of which pairs an input topic with the intended output . For the purpose of supervised learning in image processing, for instance, a machine learning system might be given photographs of vehicles that have been manually categorized as either automobiles or trucks, depending on the task at hand. The training is considered to be finished whenever the system is able to differentiate between and classify unlabeled images, which should occur after an adequate length of time has passed from the beginning of the process.

Both unsupervised and supervised learning fall under the following kinds of algorithm:

- Classification: Whenever the predicted output is a category, such as White or Black
- Regression: A prediction problem occurs when the outcome variable is a real value, such as weight or dollar.

Examples :

1. Naive Bayes Classification
2. Decision Trees
3. Random forest
4. Support vector machines
5. Ensemble Methods

6. Unsupervised learning
7. Linear regression for regression problems
8. Ordinary Least Squares Regression
9. Logistic Regression

2.4 Unsupervised Learning

Unsupervised learning allows robots to learn from unclassified or unlabeled data without supervision. The system groups unsorted data based on similarities, patterns, and differences without training.

Unsupervised learning uses unclassified data. Unsupervised learning finds data similarities and responds to new data based on their presence or absence. No teacher is provided, therefore the machine isn't trained. The computer can only find hidden structure in unlabeled data.

Unsupervised machine learning algorithms are divided into two categories:

- Clustering: A clustering challenge requires detecting underlying data groups, such as client buying behavior.
- Association: In association rule learning, you must find rules that describe your data.
- Below are some of the most common unsupervised algorithms used:
 1. Apriori algorithm for association rule learning problems
 2. K-means for clustering problems
 3. Independent Component Analysis
 4. Singular Value Decomposition

2.5 Semi-Supervised Machine Learning

Semi-supervised learning issues are those in which a substantial quantity of input data (X) is available, but only a portion of it is labeled (Y). These issues are intermediate between supervised and unsupervised learning. A photo library in which only some of the photographs are named (e.g., dog, cat, person) while the bulk are unlabeled is an excellent example.

Many machine learning challenges in the actual world fall under this category. This is due to the fact that labeling data can be costly or time-consuming, as it may require access to domain specialists. In contrast, unlabeled data is inexpensive and simple to obtain and keep.

Unsupervised learning may uncover the input variable structure. Plus, supervised learning is used to construct best-guess predictions for unlabeled data, then feed that data back into the algorithm as training data and use the model to predict on fresh unusual data.

2.6 Reinforcement Machine Learning

Reinforcement learning (RL) determines how software agents should function to maximize cumulative reward. Reinforcement learning, supervised learning, and unsupervised learning are the three core machine learning paradigms.

Reinforcement machine learning involves choosing appropriate action to maximize reward in a given circumstance. Various software and computers use it to determine the optimal behavior or action to do in a given circumstance.

In supervised learning, the training data contains the solution key, so the model is trained with the right response. In reinforcement learning, there is no answer and the reinforcement agent chooses how to finish the task. It will learn from experience without a training dataset.

2.7 Neural Network

Multiple layers of neurons with trainable weights and biases compose each layer of neural networks. All neurons are totally interconnected with their predecessors and

successors. The first layer, which was seen as a single vector, is the input layer. The final layer is the output layer, whose output view is based on a prediction. Other levels between the input layer and the output layer are referred to as hidden layers, which process and transmit the message from the preceding layer to the post layer. Every neuron will get input from neurons in the layer underneath it. The neuron then executes a dot product of all inputs, followed by an optional nonlinearity function as output.

2.8 CNN

Classification is a methodical organization of things into groups and categories based on their characteristics. Image categorization was developed to bridge the gap between computer vision and human vision by educating the computer with data. On the basis of the content of the vision, the classification of the image is achieved by distinguishing the image into the required category. For decades, image classification has been a significant challenge in computer vision. Image interpretation and classification is a simple operations for people, but for computers, it is a tremendously expensive endeavor. In principle, each image consists of a collection of pixels⁵, each of which is represented by a unique value.

In order to store an image, the computer must now have greater storage space. To classify photos, a greater number of calculations must be performed. For this, higher configuration and more powerful systems are required. It is not possible to make decisions in real-time based on the information since completing so many computations takes too much time.

In general, image classification is the process of collecting information from an image by labeling its pixels with their respective classes. It can be accomplished using either Supervised classification or Unsupervised classification.

Deep learning (DL) is a branch of machine learning that can learn through its own computational approach. A deep learning model is proposed to persistently dissect information with a uniform structure, simulating how humans arrive at conclusions. Deep learning accomplishes this by utilizing a layered framework of multiple algorithms expressed as an artificial neural system (ANN). The architecture of an ANN is simulated

with the help of the human brain's biological neural network. This makes deep learning superior to conventional machine learning methods.

Deep learning takes into account the neural networks that classify a picture based on its characteristics. This is done in order to construct a comprehensive feature extraction model that is capable of resolving the issues caused by conventional approaches. The extractor of the integrated model must be able to accurately learn to extract distinguishing characteristics from the training batch of images.

Deep learning algorithms are currently producing positive outcomes in fields such as computer vision. Convolutional Neural Network, a technique for machine learning, is used for picture classification.

CNN is a sort of feed-forward artificial neural network that has been applied successfully to image analysis. It is based on biological processes, and the neurons are interconnected like in the visual brain of animals.

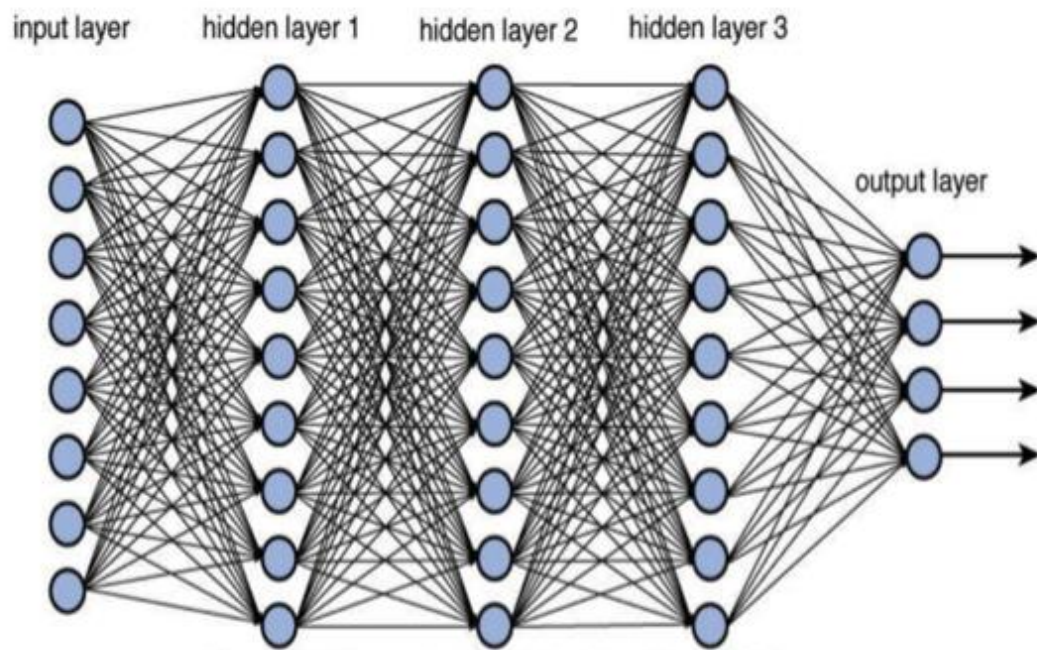


Figure 4. Deep learning layers

CNN is comprised of three distinct levels, including an input layer, hidden layers, and an output layer. Images are created as a matrix of pixels, and these pixel values, together with weights and biases, are fed to the input layer (for non-linearity). Typically, the

output layer will be a completely linked layer that identifies the image's class. The concealed layer may be convolutional, pooling, or completely connected.

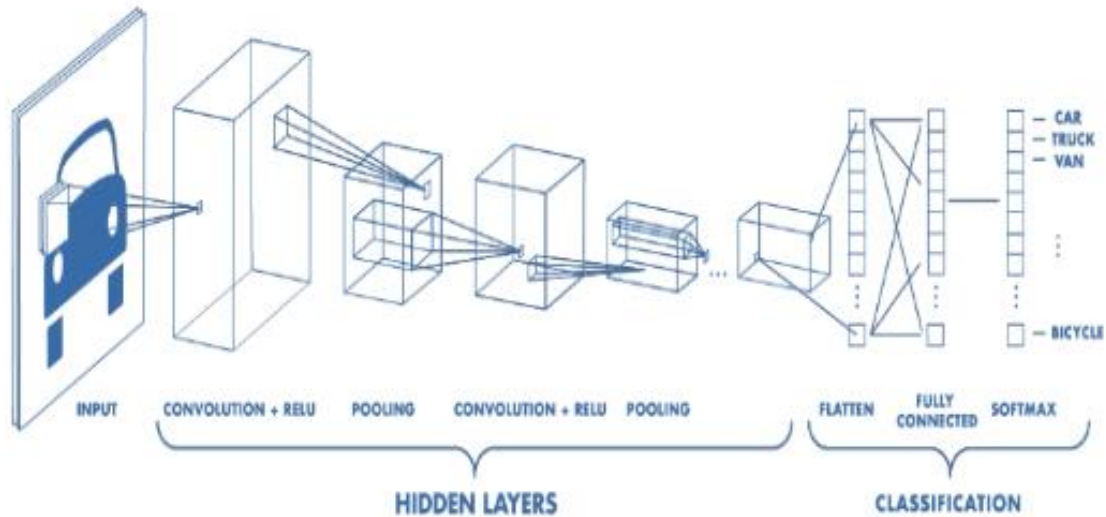


Figure 5. Typical CNN Architecture

The Convolutional layer is a fundamental construction piece and its parameters are learnable filters. Each filter is tiny in size (width and height) but stretches throughout the input volume depth. The 2-dimensional activation map is generated by conducting a dot product between the input and entry of each filter. As a result, the network discovers filters that become active when it recognizes a certain feature at a particular spatial location in the input. The pooling layer is utilized to downsample the image without losing visual information. Max pooling utilizes the maximum value from the prior layer's cluster of neurons. Every neuron in one layer is coupled to every neuron in the other layer via the fully connected layer.

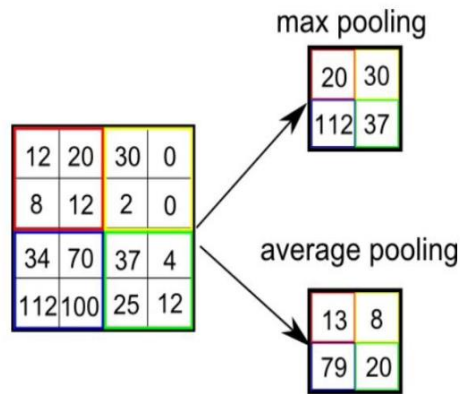


Figure 6. Types of pooling layers

CNNs require less preprocessing than standard classification algorithms, which employ hand-crafted filters. CNN's learning filters are independent of human interaction, which is a significant advantage. CNN is a supervised deep learning method that requires huge amounts of labeled data for network training. After training, the model will learn the weights and the classifier's accuracy will be enhanced.

2.9 Under-fitting and Over-fitting in Machine Learning

Suppose we are constructing a model for machine learning. A model is considered to be a good machine learning model if it generalizes any new problem-domain input data in an appropriate manner. This allows us to forecast future data that the data model has never observed.

Assuming we want to evaluate how well our machine learning model learns and generalizes to the new data, we would conduct a generalization test. The poor performance of machine learning algorithms is largely attributable to overfitting and underfitting.

Under-fitting is the inability of a statistical model or machine learning method to capture the underlying trend of the data. Under-fitting degrades the accuracy of our machine learning model. (It's like attempting to fit tiny jeans!) Its recurrence merely indicates that our model or method does not adequately suit the data. It typically occurs when there are insufficient data to develop an appropriate model and while attempting to build a linear model with non-linear data. In such situations, the machine learning model's rules are too simple and flexible to be applied to such a small amount of data; hence, the model will

likely produce a large number of incorrect predictions. Under-fitting can be avoided by leveraging more data and limiting the number of characteristics through feature selection.

A statistical model is considered to be overfitted when it has been trained with an excessive amount of data (similar to fitting oneself into big jeans!). When a model is trained with such a large amount of data, it begins to learn from the noise and incorrect data entries in our data set. The model fails to appropriately classify the data due to an excess of information and noise. Non-parametric and non-linear approaches are responsible for overfitting since these types of machine learning algorithms have greater leeway in developing the model based on the dataset and can therefore create truly unrealistic models. If we have linear data, we can avoid overfitting by utilizing a linear technique, and if we are utilizing decision trees, we can use parameters such as the maximal depth.

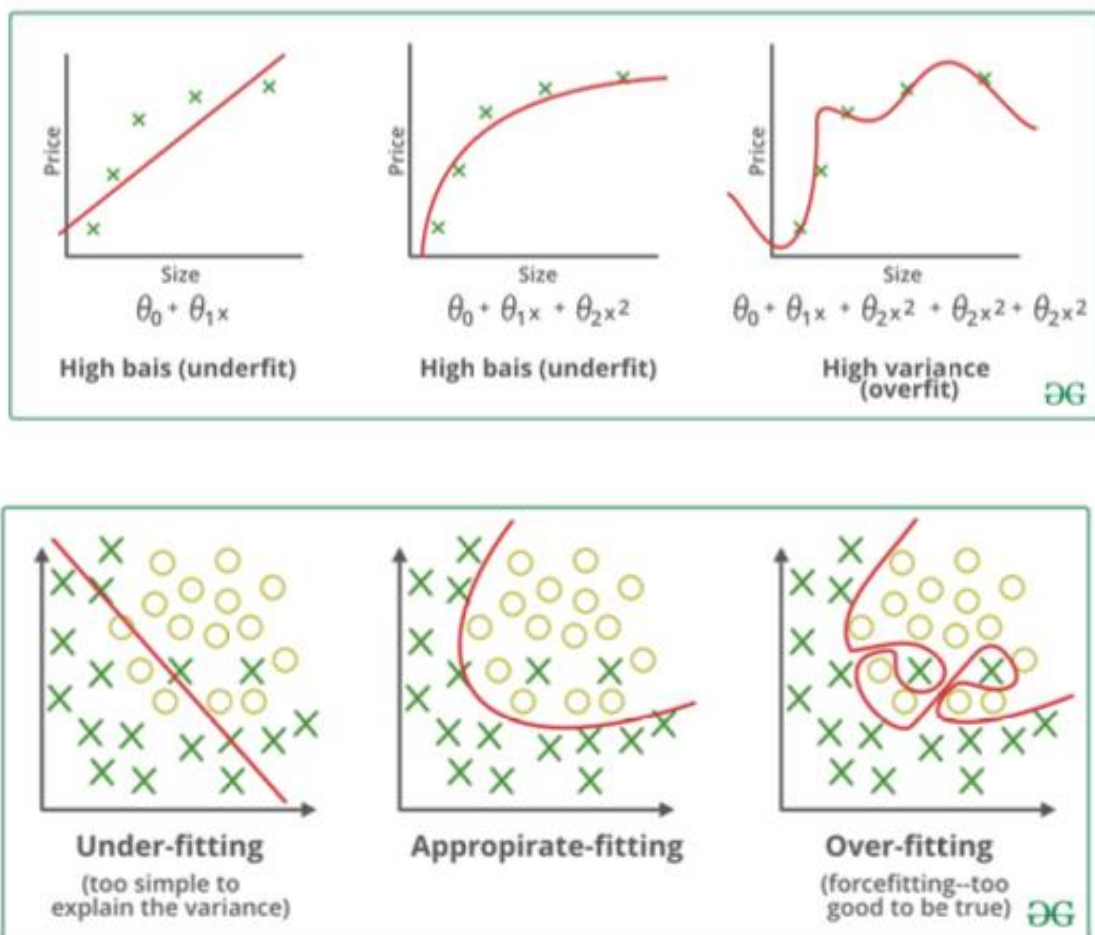


Figure 7. Under-fitting and Overfitting

To avoid these problems here are the commonly used methodologies:

- **Cross- Validation:** Standard procedure for determining out-of-sample prediction error is 5-fold cross validation.
- **Early Stopping:** Its rules tell us how many iterations may be conducted before a learner begins to overfit..
- **Pruning:** Extensive pruning is employed in the construction of related models. It only eliminates nodes that contribute minimal predictive power to the topic at hand.
- **Regularization:** It adds a cost word for incorporating additional features into the objective function. As a result, it attempts to minimize the cost term by pushing the coefficients of numerous variables to zero.

2.10 Dropout Regulation

Dropout regulates Neural Networks to avoid overfitting. Effectively simple. Eliminated neurons don't engage in feedforward or backpropagation during training. Dropout is sampling a neural network throughout the whole network and modifying its parameters depending on incoming input. Dropout layers were always between fully-connected layers. The dropout layer randomly drops units.

2.11 Regularization

Avoiding overfitting is a crucial component of training any machine learning model. If the model is overfit, its accuracy will be low. This occurs when your model tries too hard to catch the noise in your training dataset. By noise, we refer to the data points that do not accurately reflect the actual properties of your data, but rather random chance. Incorporating such data items into your model increases its adaptability at the cost of overfitting. It lowers the calculated coefficients to zero. This method opposes learning a complex or flexible model to prevent overfitting.

Fitting needs an RSS loss function. Loss function is minimized by choosing coefficients. Now, your training data will update the coefficients. If training data are noisy, calculated

coefficients won't generalize effectively. Regularization decreases learnt estimates to zero.

2.12 Good Fit in a Statistical Model

In an ideal situation, when the model provides predictions with zero error, the data fit is said to be good. This circumstance is attainable at a point between over- and under-fitting. In order to comprehend it, we will need to examine the performance of the model over time as it learns from the training dataset.

The model will continue to learn over time, and as a result, its error on training and testing data will continue to decrease. Due to the existence of noise and less valuable features, the model will become more susceptible to overfitting if it continues to learn for too long. Consequently, the model's performance will decline. In order to obtain a satisfactory fit, we will stop right before the error begins to increase. At this stage, the model is deemed to have excellent performance on both the training dataset and our unobserved testing dataset.

2.13 Normalization

Normalization is a technique frequently used in machine learning data preparation. The objective of normalization is to convert the values of numeric columns in a dataset to a similar scale without altering the ranges of the values. For machine learning, normalization is not required for every dataset. It is only necessary when features have distinct ranges.

Consider, for instance, a data set with two attributes, age(x_1) and income (x_2). Where age goes from 0 to 100 and income ranges from 0 to at least \$20,000. The range of income, which is approximately 1,000 times greater than age, is 20,000–500,000. Thus, the ranges of these two characteristics are considerably different. When conducting additional analyses, such as multivariate linear regression, for instance, the attributed income will intrinsically have a greater impact on the outcome due to its greater value. However, this does not necessarily make it a more significant predictor.

2.14 Performance Metrics for Classification problems in Machine Learning

After Feature Engineering, Selection, and Model Construction, the model's efficacy is evaluated using test datasets and a measure. Different Machine Learning Algorithms are evaluated using distinct performance metrics. For the time being, we shall concentrate on those employed in Classification difficulties. Log-loss, accuracy, AUC (area under curve), etc. are classification performance measures. Precision and recall may be used to evaluate machine learning algorithms for search engine sorting.

The criteria selected to evaluate your machine learning model are crucial. Metric selection effects the comparison and measurement of machine learning algorithm performance. Here are some metrics usually required to estimate the performance of ML:

1. Confusion Matrix
2. Classification Accuracy
3. Precision
4. Logarithmic Loss
5. Recall or Sensitivity
6. F1 Score
7. Area under curve
8. Mean Absolute Error
9. Mean Squared Error

2.14.1 Classification Accuracy

When we refer to accuracy, we typically mean Classification Accuracy. It is the proportion of accurate predictions relative to the total number of input samples.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total number of predictions made}}$$

It is essential that there be an equal number of examples coming from each category for it to be successful. Take, for instance, the fact that our training set contains 98 percent samples from class A and only 2 percent samples from class B as an example. Then our model may easily achieve a training accuracy of 98 percent by simply making the correct prediction for each training sample that belongs to class A.

If the same model were to be evaluated using a test set that consisted of sixty percent samples from class A and forty percent samples from class B, the accuracy of the evaluation would fall to sixty percent. Classification Precision is wonderful, but it misleads us into thinking that we have achieved a high level of accuracy.

The true difficulty emerges when the cost of incorrectly classifying the samples from the minor classes is exceptionally large. If we are dealing with an illness that is uncommon but lethal, the cost of missing the diagnosis of the disease in a sick person is going to be far higher than the cost of sending a healthy individual for additional testing.

2.14.2 Logarithmic Loss

Logarithmic loss, often known as log loss, is a method that operates by assigning a penalty on false classifications. It is effective for classification across several categories. When utilizing Log Loss, the classifier is required to assign a probability to each class for each and every sample.

2.14.3 Confusion Matrix

Confusion Matrix, according to its name, generates a matrix as its output and analyzes the model's overall effectiveness in a manner that is summarized by the matrix. Consider the possibility that we are having trouble with binary classification. We have samples that can be classified as either YES or NO. In addition, we have our own algorithm that can determine a category based on the input sample that was provided. When we put our model to the test with 165 different examples, we got the following results:

There are four crucial examples, and they are as follows:

- Instances where our predictions was correct and the actual result was likewise yes are referred to as "true positives."
- Instances that were a "True Negative" consist of those in which we had predicted "No" and the actual output was "No."
- Instances in which we expected YES but the actual output was NO; these are referred to as false positives.
- Instances in which we expected no, but the actual output was yes are examples of false negatives.

The correctness of the matrix can be evaluated by taking the average of the values that run along the "major diagonal," also known as the "long diagonal."

$$Accuracy = \frac{TP + FN}{total\ number\ of\ sampels}$$

Confusion Matrix is the foundation for all other measures.

2.14.4 Area Under Curve

Area Under Curve (AUC) is an evaluation metrics. It is used for challenges involving binary classification. A classifier's AUC corresponds to the likelihood that it would rank a randomly selected positive example higher than a randomly selected negative example.

Before defining AUC, let's define two fundamental terms:

- Rate of True Positives (Sensitivity): True Positive Rate is defined as True Positives / (False Positives plus True Positives). True Positive Rate is the fraction of positive data points that are accurately identified as positive, relative to the total number of positive data points.

$$\text{True Positive Rate} = \frac{TP}{TP + FN}$$

- False Positive Rate is the proportion of negative data values wrongly perceived as positive.

$$\text{False Positive Rate} = \frac{FP}{FP + TN}$$

- True Positive Rate and False Positive Rate are both 0-1 FPR and TPR are calculated at threshold values (0.00, 0.02, 0.04, ..., 1.00) and a graph is created. AUC is the area under the False Positive Rate vs True Positive Rate curve in [0, 1].

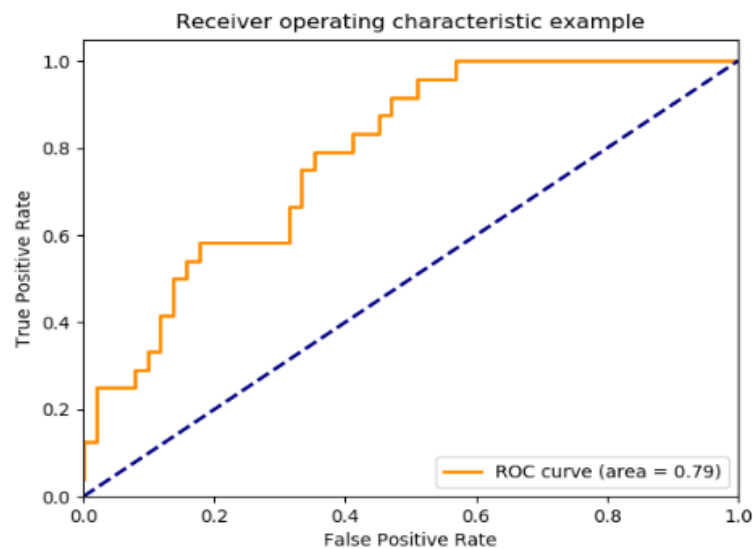


Figure 8. Area Under Curve graph

AUC is [0, 1]. Our model's performance improves with higher values.

2.14.5 F1 Score

The F1 Score is a measurement that determines how accurate a test is; it is the Harmonic Mean between the precision and recall scores. The F1 Score may vary anywhere from 0 to 1. It gives an indication of how durable your classifier is as well as how accurate your classifier is (how many occasions it correctly classifies) (it does not miss a significant number of instances).

A classification system that has high accuracy but poor recall is very exact, but it fails to recognize a significant proportion of instances that are difficult to categorize. When the F1 Score is greater, it indicates that our model is doing better. The F1 Score algorithm makes an effort to find a happy medium between accuracy and recall.

The term "precision" refers to the proportion of correctly positive outcomes that correspond to the total number of positive outcomes that were predicted by the classifier.

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

Recall, it's the total number of relevant samples divided by the number of positive results.

$$Recall = \frac{TP}{TP + FN}$$

2.14.6 Mean Absolute Error

The average disparity between the original values and the predicted values is what's known as the Mean Absolute Error (MAE). It determines how inaccurate the predictions were in comparison to the actual results. On the other hand, they do not reveal if our projections of the data are too low or too high.

2.14.7 Mean Squared Error

MSE is similar to MAE, except it calculates the average square of the gap between estimated and forecasted values. MAE is the average of actual and anticipated values. Mean Squared Error has the benefit of having a gradient that can be computed with less complexity, whereas Mean Absolute Error necessitates the use of complicated linear

programming methods. The influence of larger errors becomes more noticeable than that of smaller errors when the mistake is squared; as a result, the system may now concentrate more on the larger errors.

In this project, the proposed model provides some of the performance metrics discussed in this section like Classification accuracy, Confusion matrix in the experimental result chapter.

Chapter 3

Development Toolkits

3.1 Tensorflow

TensorFlow is a Python-compatible numerical computation library that facilitates machine learning. Complex machine learning. Machine learning frameworks like Google's TensorFlow make acquiring data, training models, delivering predictions, and improving future outcomes easier than in the past.

Google Brain created TensorFlow for numerical computation and large-scale machine learning. TensorFlow uses a common paradigm to connect machine learning and deep learning models and algorithms. Python provides an API for building applications using the framework, while C++ executes the programs.

TensorFlow may be used for handwritten digit classification, picture identification, word embeddings, recurrent neural networks, machine translation, NLP, and PDE-based simulations. TensorFlow forecasts production using the same training models. TensorFlow enables developers build dataflow graph topologies that depict data movement across processing nodes. Each node in the graph reflects a mathematical procedure.

TensorFlow provides all this through Python. Python is easy to learn and use, and it couples high-level abstractions simply. TensorFlow nodes and tensors are Python objects and programs.

Python isn't used for math computations. TensorFlow's transformation libraries are C++ binaries. Python distributes traffic and offers high-level programming abstractions to link components.

TensorFlow apps can run on a local PC, cloud cluster, iOS, Android, CPUs, and GPUs. Using Google's cloud, you can run TensorFlow on TPU chips for acceleration. TensorFlow-generated models may be used to make predictions on almost any device.

TensorFlow's main machine learning advantage is abstraction. The developer may focus on the application's overarching logic instead of creating algorithms or connecting function outputs. TensorFlow handles subtleties.

TensorFlow simplifies debugging and understanding TensorFlow applications. Instead of producing and evaluating the whole graph as a single opaque object, eager execution mode lets you evaluate and change each graph action separately. TensorBoard's web-based interface lets you observe and profile graph execution.

3.2 Keras

TensorFlow is an open-source machine learning framework that covers the whole learning process from beginning to finish. It is a complete and flexible ecosystem consisting of tools, libraries, and other resources that provide workflows with high-level application programming interfaces (APIs). You may design and deploy machine learning models by choosing the appropriate idea level from among the many that are provided by the framework, which gives a number of idea levels. For instance, if you need to execute massive machine learning jobs, you can utilize the Distribution Strategy API to perform distributed hardware configurations, and if you need a complete production machine learning pipeline, you can utilize TensorFlow Extended. Both of these options are available to you (TFX). The following is a summary of few of the most important characteristics:

TensorFlow offers abstraction layers for model creation and training.

Anywhere ML Production: TensorFlow makes training and deploying models easy, independent of language or platform.

Keras Functional API and Model Subclassification API make TensorFlow a great tool for research testing.

Keras operates on TensorFlow, CNTK, and Theano. Keras simplifies and speeds up deep learning development and CPU/GPU operation. Python makes this framework easy to debug and extend. Keras advantages include:

User-Friendly: Keras's interface is straightforward, consistent, and tailored for common use cases, with explicit error feedback.

Keras models are modular and composed of programmable construction pieces.

Keras makes it easy to construct new concepts and research building pieces.

Keras' consistent and easy APIs simplify user activity for typical use cases and provide unambiguous error feedback.

Keras is a wrapper for TensorFlow, therefore comparing it to Tensorflow is inaccurate. All things have advantages and drawbacks, therefore we shouldn't judge only on them. Machine Learning is a complex discipline, and various jobs need different frameworks. Keras is for deep neural networks, whereas TensorFlow is for machine learning. The researcher's job determines the framework to use. This project's Keras model uses a Theano backend.

3.3 OpenCV

The topic of how to give robots the ability to learn from data is at the heart of the field of Artificial Intelligence (AI), specifically the subfield known as Machine Learning. The fundamental concept is to devise a method through which a computer may learn from its own experiences and then use that knowledge to generate accurate judgments and forecasts. The algorithms used in machine learning need expertise from a variety of disciplines, since they frequently overlap with fields such as statistics, mathematics, physics, pattern recognition, and others.

OpenCV, which stands for Open Source Computer Vision which is nothing but a library that may be used for computer vision and also includes a library for machine learning.

3.4 Tensorboard

Machine Learning is a subfield of Artificial Intelligence focused with making machines capable of learning from data. The fundamental concept is to enable a computer to make intelligent decisions and predictions based on historical data. Machine Learning algorithms involve interdisciplinary knowledge and frequently connect with statistics, mathematics, physics, pattern recognition, and other disciplines.

OpenCV (Open Source Computer Vision) is a computer vision library that includes a machine learning library.

3.5 Jupyter Notebook

Creating and sharing documents that include live code, mathematics, graphics, and narrative prose is now feasible thanks to an open-source web tool called Jupyter Notebook. You may get Jupyter Notebook from the internet. Data visualization, statistical modeling, numerical simulation, data cleansing and transformation, machine learning, and a great many more applications are some examples of what may be done with this technology.

3.6 Kaggle Datasets

The owner of the online platform known as Kaggle, which is used by people learning data science and machine learning, is the company known as Google LLC. In a web-based data science environment, users are able to investigate and construct models, discover and publish data sets, engage with other data scientists and machine learning professionals, and compete to solve data science difficulties. Users have the ability to perform all of these things and more on the platform that is Kaggle. Kaggle got its start by hosting competitions in the field of machine learning; now, in addition to such competitions, the firm provides a public data platform, a cloud-based workbench for data science, and short form artificial intelligence courses. On March 8, 2017, Google revealed that it would be purchasing Kaggle, and the revelation was made public.

Chapter 4

Methodology

4.1 Data Collection

Data collection is the process of collecting and analyzing information from a vast array of sources. To develop viable artificial intelligence (AI) and machine learning solutions, we must collect and store data in a manner that makes sense for the business problem at hand.

Collecting data allows us to keep track of prior events so that we can identify recurrent patterns through data analysis. We construct predictive models from these patterns using machine learning algorithms that search for trends and anticipate future changes.

Good data collection procedures are essential to the development of high-performing predictive models because the accuracy of the models is dependent on the data used to construct them. The data must be error-free (garbage in, garbage out) and contain information pertinent to the current activity. In this instance, the relevant Data for the challenge include photos of both melanoma and other (non-melanoma) types of skin cancer. The dataset, which is available on the Kaggle website (www.kaggle.com), is a collection of dermatoscopic images captured and archived by various modalities from various populations.

The final dataset contains 17805 (<https://www.kaggle.com/drscarlat/melanoma>) dermatoscopic images that can be used for academic machine learning applications as a training set. It comprises a representative sample of all significant pigmented lesion diagnostic categories.

The photos are divided into two categories: melanoma-related and non-melanoma-related. In addition, the dataset is subdivided into three subsets, each comprising the equal number of (melanoma/non-melanoma) photos, and utilized as training, validation, and testing datasets.

4.2 Pre-Processing

In order to achieve better outcomes from the used model in Machine Learning projects, the data format must be correct. Certain Machine Learning models require information in a certain format; for instance, the Random Forest method does not tolerate null values; consequently, in order to execute the Random Forest algorithm, null values must be removed from the original raw data set. A data set should also be organized in such a way that multiple Machine Learning and Deep Learning algorithms can be executed in a single data set and the best method selected.

The transformations made to our data prior to feeding it to the algorithm are known as pre-processing. Data Pre-processing is a technique used to transform raw data into a set of clean data. In other words, anytime data is acquired from several sources, it is collected in a way that makes analysis impossible.

4.3 Image scaling

Image Processing is an essential technique for manipulating images. Analyzing and manipulating a digitized image increases its quality. Processing provides several approaches for picture processing, such as Image Resizing and Image Enhancement.

Image scaling refers to the resizing of a digital image in computer graphics and digital imaging. Upscaling or resolution improvement refers to the process of magnifying digital content in video technology.

An picture is nothing more than a two-dimensional array of 0 to 255-valued values (or pixels). It is defined by the function $f(x,y)$, where x and y are the horizontal and vertical coordinates, respectively. The value of $f(x,y)$ at any given point represents the pixel value at that location in an image.

Resizing a picture to a lower resolution entails manipulating and processing images with fewer pixels that represent an image, hence reducing the computation power required to manipulate and process the image. The collection consisted of photos at a resolution of approximately $29700 * 3$ pixels per image ($600 * 450$). The photos were resized to 128×128 for improved and more efficient processing.

4.4 Grayscale Conversion

From black to white, grayscale represents a spectrum of monochromatic tones. Consequently, a grayscale image is comprised of simply shades of gray and no color. While it is possible to store digital photographs as grayscale (or black and white) images, even color images contain grayscale information. This is due to the fact that regardless of color, each pixel has a brightness value. On a scale ranging from black (zero intensity) to white (maximum intensity), luminosity can alternatively be defined as brightness or intensity (full intensity). Most picture file formats include a minimum of 8-bit grayscale, providing 256 levels of luminance per pixel. Some formats offer 16-bit grayscale, which yields 65,536 levels of brightness (216).

This procedure eliminates all color data, leaving only the brightness of each pixel. Because digital images are displayed using a combination of red, green, and blue (RGB) hues, each pixel possesses three distinct brightness values. Therefore, these three values must be merged into a single value in order to remove color from an image. This can be accomplished in numerous ways. Each pixel's brightness values can be averaged as one option. A second technique includes retaining only the brightness values from the red, green, or blue channel.

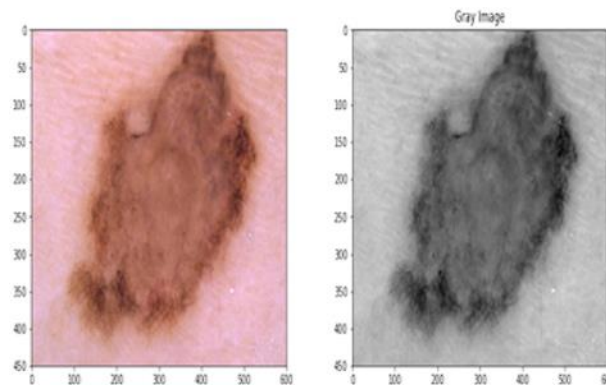


Figure 9. Grayscale Image

I converted all photos to grayscale to facilitate faster computation. It is possible to execute simple pixel-by-pixel processing of a one-million-pixel image in milliseconds with contemporary processors and parallel programming. Facial identification, optical

character recognition, content-aware scaling, mean shift segmentation, and other tasks can take significantly longer. Regardless of how long it takes to modify an image or extract relevant information from it, the majority of customers and users desire a faster processing speed. If we make the hazy assumption that processing a three-channel color image takes three times as long as processing a grayscale image, or possibly four times as long, then the processing time for a three-channel color image is three times that of a grayscale image.

4.5 Median Filer

The Median Filter is a non-linear digital filtering method used to remove image or signal noise. Normal pre-processing to improve future processing (for example, edge detection on an image). Median filtering retains edges while minimizing noise in digital image processing and signal processing.

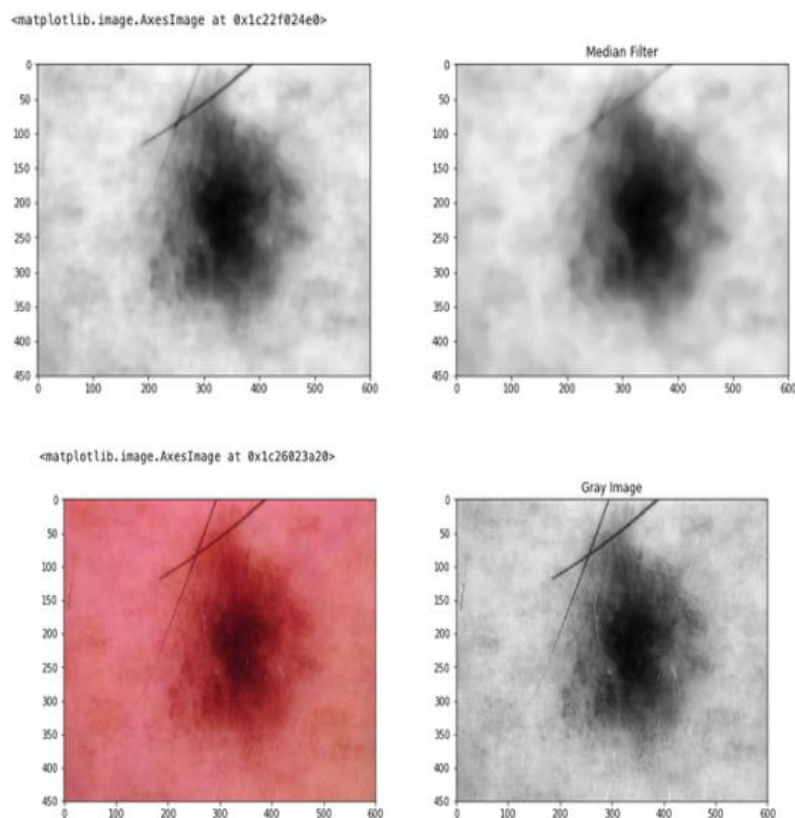


Figure 10. Median filter

4.6 Segmentation

Image segmentation in computer vision involves dividing a digital image into several parts. Segmentation simplifies or transforms an image's representation to make it easier to examine. Photographs are segmented to locate items and boundaries.

Converting an image into a collection of pixel sections that are either represented by a mask or an image that has been labeled is what is meant by the term "image segmentation." You can save time by just processing the parts of a picture that are relevant to the task at hand rather than the complete image if you first divide it up into segments. An image is said to be segmented when it is partitioned into distinct sections that are consistent with regard to a selected characteristic such as brightness, color, or texture, amongst other possible examples.

Segmentation methods can be roughly classified into the following categories:

- Histogram thresholding: The determination of one or more histogram threshold values is included in these approaches. These values serve to differentiate the objects from the background.
- Clustering: Unsupervised clustering algorithms are utilized in these techniques, which result in the segmentation of a color space (or other feature space) into homogenous sections.
- Edge-based: The identification of edges between the regions is accomplished with these methods by utilizing edge operators.
- Region-based: Pixels are grouped together into homogenous regions through the use of region merging, region splitting, or both of these techniques, as part of these methodologies.

- Morphological: The watershed transform is used in these methods for the purpose of determining the contours of objects starting with predefined seeds.
- Model-based: These techniques include modeling images as random fields, with the parameters of the fields being selected by various optimization algorithms.
- contours: The identification of object contours is the focus of these approaches, which make use of curve evolution techniques.
- Soft computing: The classification of pixel data is accomplished by the use of these methods by employing soft computing techniques like as neural networks, fuzzy logic, and evolutionary computation. The picture of Superficial Spreading Melanoma (SSM) that has been segmente may be seen in Figure (2.1).

In this report, the proposed system utilizes the Thresholding segmentation which gave enough results for discovering the features that is required.

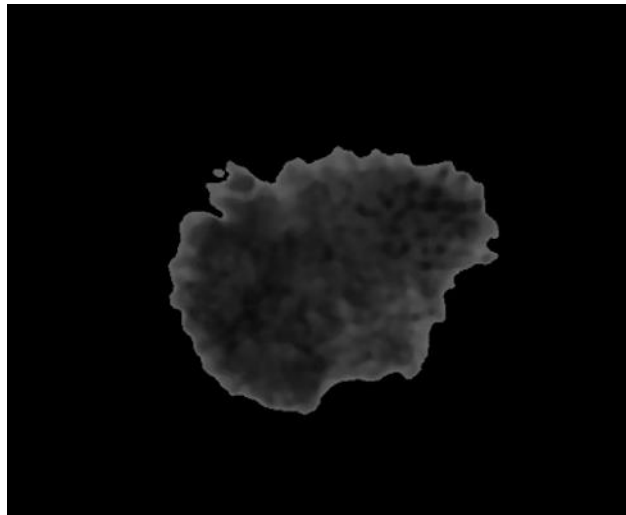


Fig 4. 1 Threshold Segmentation

4.7 Feature Extraction

In feature extraction, the number of resources necessary to describe a huge data set is diminished. When performing analysis on complicated data, the number of variables involved is one of the most significant obstacles. In addition to necessitating a high amount of memory and processing capacity, a large number of variables might cause a classification algorithm to overfit to training examples and generalize poorly to new samples. Feature extraction is an umbrella word encompassing methods of building variable combinations to circumvent these issues while still accurately describing the data. Numerous practitioners of machine learning feel that well optimized feature extraction is the key to efficient model design.

The characteristics we want in this category are the signs and diagnosis of melanoma. Moles, brown patches, and skin growths are mostly innocuous, but not always. The first five letters of the alphabet can be used as a simple guide for recognizing melanoma warning signals.

A represents Asymmetry. A majority of melanomas are asymmetric. If a line is drawn through the centre of the lesion, the two halves do not match, therefore it appears different from a round to oval and symmetrical common mole.



Figure 11. Asymmetry sign

B is for Border. Melanoma borders are typically irregular and may have scalloped or notched edges, but the borders of ordinary moles are typically smoother and more uniform.



Figure 12. Un-even borders

The letter C stands for Color. Variable brown, tan, or black hues may be the first indication of melanoma. As melanomas advance, the hues red, white, or blue may become visible. Typically, benign moles are one to two colors of brown.



Fig 4. 2 Color Variation

D is for Diameter . Melanomas at the time of diagnosis are often larger than a pencil eraser (about 6 mm or 14 inch in diameter). They may also be smaller, however. Additionally, melanomas may seem darker than regular moles. Nonetheless, they can also be lighter or pink in hue. Infrequently, melanotic melanomas lack color. **E** is for Changing through time Any change in the size, shape, color, of place on your skin.

4.8 Classification

We conclude with a discussion of the classification stage, in which machine learning methods are used to the attributes created from the segmented features to predict the presence of melanoma. CNN is a supervised learning approach that requires both input data and desired output data.

A CNN typically consists of three primary components: a Convolutional Layer, a Pooling Layer, and a Fully linked Dense Network. The Convolutional layer applies $m \times n$ filters to the input image to generate a feature map. Next, the feature map is given to the max pool layer, which is primarily used for dimensionality reduction; it selects only the most prominent features from the feature map. All the features are then flattened and provided as input to the fully connected dense neural network, which learns the weights using back - propagation algorithm and produces the classification output.

CNN functions like the visual cortex, which focuses on one item while blurring the others; similarly, CNN identifies one section/window of the input picture at a time.

In the convolutional layer, CNN will generate a feature map for each segment each time. In the Pooling layer, the unnecessary features are eliminated and the most significant characteristics for that area are extracted, hence accomplishing feature extraction. Therefore, using CNNs eliminates the need for an additional feature extraction technique.

CNN assigns photos into one of the two categories (melanoma/non-melanoma) in this instance. The model achieved an accuracy of 85% on the training dataset and 83% on the testing dataset.

Chapter 5

Experimental Result And Discussion

In order to develop this system, which was never an easy task, I tried out and experimented with a large number of different machine learning strategies. An approximate ninetieth of the effort was put into experimenting with various combinations of preprocessing and feature extraction methods, all of which finally gave me the opportunity to develop the most robust model possible. I am going to write up a detailed summary of the tests that I carried out along the way to developing this model.

5.1 Challenges

The following is a list of the challenges I encountered and how I overcame them.

5.1.1 Imbalanced Dataset

The very first step of the data collection process was one of the first places where I tried experimenting with different approaches. The data were unbalanced, and they needed to be balanced, because if a model for machine learning is developed using unbalanced data, the model will never generalize well, and it will typically develop a bias toward one category. The data needed to be balanced because they were unbalanced, and they needed to be balanced. Due to the fact that I was aware of this information, I decided to implement a method that is known as Data Augmentation. Data Augmentation is a method that is used in data analysis to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data derived from existing data. When training a machine learning model, it helps to prevent overfitting by acting as a regularizer and reducing overfitting. It is strongly connected to the practice of oversampling in statistical research.

Then the performance of the models were significantly improved after balancing the dataset, and their performance was even better with the implementation of data augmentation.

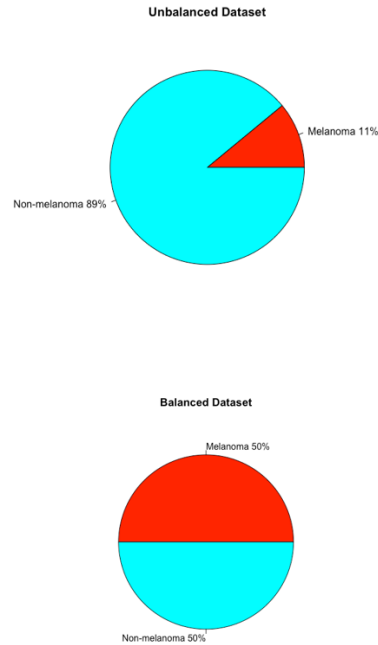


Figure 13 Before & after data augmentation

5.1.2 Computational Difficulties

Because of the nature of the high-resolution photos contained in the dataset. I struggled to achieve the computing requirements required to speed up the process of training a model, so I reduced their size to $128 * 128$ and changed them to Grayscale, which made things faster for me to continue experimenting with the data.

5.1.3 Filtering functions

Filtering functions enables you to filter or remove noise from objects such as hairs. I used a median filter for hair removal, and applying it required me to provide a median parameter. After many failed trials, I discovered the perfect median parameter number, which was 5.

5.1.4 Choosing the right segmentation

Image segmentation is commonly used to locate objects and boundaries in images. In my case, I was just concerned with detecting the skin lesion and neglecting the surrounding

environment. There are many other types of segmentation, but since I was searching for automated segmentation, I went with threshold, specifically a type called Binary Threshold Segmentation. Again, I attempted to find the best threshold setting for my dataset. After much investigation and testing, I discovered that the threshold number 150 worked more consistently and had a positive effect on the model's performance, so I decided to use it.

5.1.5 Performance Tuning

Many moving pieces required to be tweaked when constructing the model to create a robust model. I used a variety of convolutional and pooling layers, which usually resulted in training fragile models. When I employed simple architecture on my CNN, it was unable to fit and suffered from underfitting. However, if I construct a more complicated CNN with many layers, it will not generalize and will begin to memorize by failing on data it has not seen, resulting in overfitting. During the process, I devised a method that included regularization and Dropout. Regularization refers to strategies for calibrating machine learning models in order to minimize the adjusted loss function and avoid overfitting and underfitting. Dropout refers to the removal of units from a neural network, both hidden and visible. By "ignoring," I mean that these units are not taken into account during a certain forward or backward pass.

In the end, the proposed system was the most reliable one, which means it was the most accurate. The CNN had three layers, two layers for pooling information, and an output layer. Below are the experimental results

5.2 Experimental Result

Here are the accuracies that the proposed model generated for each dataset. Below table show the accuracies the model achieved on various datasets and the figures show the visual representation of models while being trained on dataset.

No	Dataset	Accuracy
1	Training Dataset	85%

2	Validation Dataset	84%
3	Testing	82.9%

Table 5. 1 Model accuracy

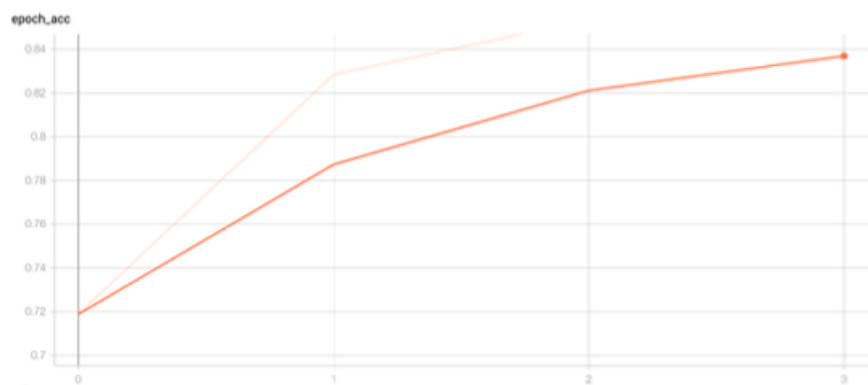


Figure 14. Classification accuracy on training dataset

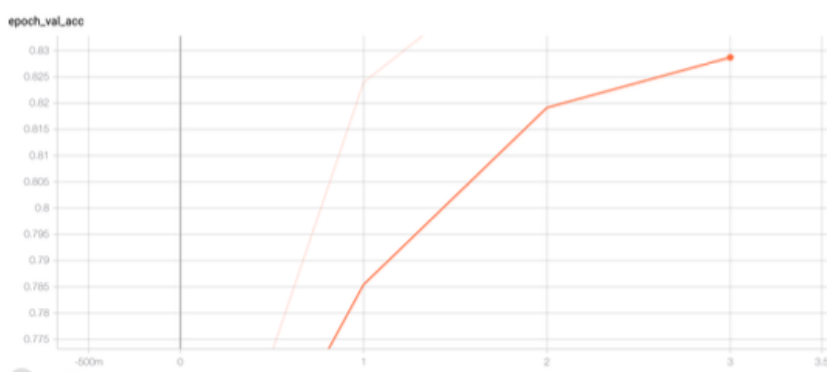


Figure 15. Classification on validation dataset

```

In [20]: # Testing Accuracy
         model.evaluate(x_test,y_test)

3561/3561 [=====] - 9s 2ms/sample - loss: 2.6526 - acc: 0.8295

Out[20]: [2.6525556534336245, 0.8295423]

# -----
Classified: 2954
Miss-Classified: 607
Total: 82.95422634091547

Classified: 1420
Miss-Classified: 361
Melanoma Accuracy 79.73048848961257    loss 20.26951151038742

Classified: 1534
Miss-Classified: 246
non_Melanoma Accuracy 86.17977528089888    loss 13.820224719101123

```

Figure 16. Classification on testing accuracy testing

From figure 16 a testing dataset of 1781 images of melanoma being used, 1420 of them have being classified correctly, but 361 have being miss—classified giving a specific accuracy of 80% of classifying melanoma. Also the same amount of non-melanoma being used, 1534 have being classified correctly while 246 being miss-classified giving us 86% of non-melanoma accuracy.

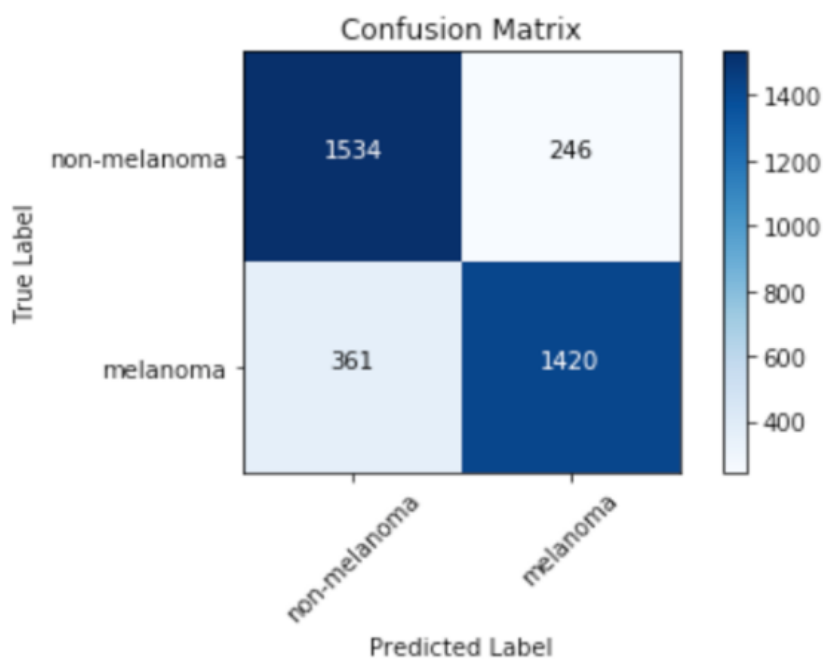


Figure 17. Confusion matrix

Skin Disease	Sensitivity	Specificity	Accuracy	Overall accuracy
Melanoma	79.7%	86%	80%	
Non-melanoma	86%	79.7	86%	82.9%

Table 5. 2 Sensitivity and Specificity

5.3 Discussion

Tensorboard is used to graph the classification accuracy on both training dataset and validation dataset but not on test dataset, the classification accuracy is presented only from the confusion matrix as there is no feature that will allow me to graph the testing accuracy even tensorboard. The proposed work has a special use case and it's limited only to skin cancer images because there is no way the proposed work would know /recognize other images that does not relate to the skin cancer. Therefore the input should always be a skin cancer image, hence an accurate classification is obtained.

5.4 Contribution

My involvement in this project allowed me to acquire the necessary machine learning skills to participate in the implementation of a project that was connected to the medical field. In order to build and train the most effective model, I made use of a large number of different machine learning techniques. I also contributed all of my expertise and knowledge to the process. I had to be patient and put in a lot of effort, but in the end I was able to design an effective architect for a CNN model that achieved more than 80 percent, which is what literature did manage to do.

Chapter 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

I utilized Convolutional Neural Networks (CNN) to classify photos from the Kaggle website in this paper. This data collection is utilized for both training and testing using CNN. It has an accuracy rate of 85%. The images utilized for training purposes are small and grayscale. Compared to other normal JPEG photos, the computer time required to process these images is astronomical. By stacking the model with more layers and training the network with more picture data using GPU clusters, image classification results will be more accurate. Future enhancements will concentrate on classifying large-sized colored images, which is highly important for image segmentation processes.

6.2 Future Work

However, some considerable drawbacks of the proposed system are the insufficient small number of images the system was trained on and the low processing capability that led the system classify image based on one color channel i.e grayscale images. These lead to the future work. The future work will focus on classifying both grayscale and colored images of large size and better accuracy will be generated by covering the obstacles that the present system couldn't cover.

References

- [1] S. Stechschulte, C. Ricotti, C. J. Cockerell, Advances in diagnostic testing for skin cancer, TOUCH BRIEFINGS (2008) 73–76.
- [2] R.Lucas, T.McMichael, W.Smith, B.Armstrong, Solar ultraviolet radiation.global burden of disease from solar ultraviolet radiation, Environmental Burden of Disease. Series 13 (2006) 117.
- [3] D.Rigel, J.Russak, R.Friedman, The evolution of melanoma diagnosis, A Cancer Journal for Clinicians 60 (2010) 301316.
- [4] N. Howlader, A. M. Noone, M. Krapcho et al., —SEER cancer statistics review, National Cancer Institute, April 2015.
- [5] Skin Cancer Foundation. (June 2022). [Online]. Available: <http://www.skincancer.org/skin-cancer-information/melanoma>.
- [6] R. H. Johr, —Dermoscopy: Alternative melanocytic algorithms—the ABCD rule of dermoscopy, menzies scoring method, and 7-point checklist, Clinics in Dermatology, vol. 20, no. 3, pp. 240-247, 2002.
- [7] G. Argenziano, I. Zalaudek, R. Corona, F. Sera, L. Cicale, G. Petrillo, and H. P. Soyer, —Vascular structures in skin tumors: A dermoscopy study, Archives of Dermatology, vol. 140, no. 12, pp. 1485-1489, 2004.
- [8] H. Pehamberger, A. Steiner, and K. Wolff, —In vivo epiluminescence microscopy of pigmented skin lesions. I, Pattern analysis of pigmented skin lesions, Journal of the American Academy of Dermatology, vol. 17, no. 4, 571-583, 1987.
- [9] A. A. Abder-Rahman and T. M. Deserno, —A systematic review of automated melanoma detection in dermoscopic images and its ground truth data, in Proc. SPIE, Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment, 2012, vol. 8318.

- [10] N. K. Mishra and M. E. Celebi. “An overview of melanoma detection in dermoscopy images using image processing and machine learning ” arXiv: 1601.07843 2016.
- [11] F. Nachbar, W. Stolz, T. Merkle, A. B. Cognetta, T. Vogt, M. Landthaler, and G. Plewig, —The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions,|| Journal of the American Academy of Dermatology, vol. 30, no. 4, pp. 551-559, 1994.
- [12] G.D.Leo,G.Fabbrocini,C.Liguori,A.Pietrosanto,andM.Sclavenzi, —ELM image processing for melanocytic skin lesion diagnosis based on 7-point checklist: A preliminary discussion,|| in Proc. Int. Symp. Meas. for Res. and Ind. Appl., Budapest, Hungary: IMEKO, 2004, pp. 474-479.
- [13] Julie Ann A. Salido and Conrado Ruiz Jr. “Using Deep Learning to Detect Melanoma in Dermoscopy Images”, International Journal of Machine Learning and Computing, Vol. 8, No. 1, February 2018.
- [14] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, —PH 2-A dermoscopic image database for research and benchmarking,|| in Proc. 35th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBC), July 2013, pp. 5437-5440.
- [15] M Manoj krishna, M Neelima, M Harshali, M Venu Gopala Rao, “Image classification using Deep learning”, International Journal of Engineering & Technology, 7 (2.7) (2018) 614- 617 .
- [16] Mijwil, M.M. Skin cancer disease images classification using deep learning solutions. *Multimed Tools Appl* 80, 26255–26271 (2021).
- [17] Demir A, Yilmaz F, Köse O (2020) Early detection of skin Cancer using deep learning architectures: Resnet-101 and inception-v3. Med Technol Congress (TIPTEKNO).
- [18] N. C. F. Codella *et al.*, "Deep learning ensembles for melanoma recognition in dermoscopy images," in *IBM Journal of Research and Development*, vol. 61, no. 4/5, pp. 5:1-5:15, 1 July-Sept. 2017.
- [19] Dildar, M.; Akram, S.; Irfan, M.; Khan, H.U.; Ramzan, M.; Mahmood, A.R.; Alsaiani, S.A.; Saeed, A.H.M.; Alraddadi, M.O.; Mahnashi, M.H. Skin Cancer Detection: A Review Using Deep Learning Techniques. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5479.

- [20] A Deep Learning Based Approach to Skin Lesion Border Extraction With a Novel Edge Detector in Dermoscopy Images *2019 International Joint Conference On Neural Networks (IJCNN)* 2019
- [21] S. Gupta, A. Panwar and K. Mishra, "Skin Disease Classification using Dermoscopy Images through Deep Feature Learning Models and Machine Learning Classifiers," *IEEE EUROCON 2021 - 19th International Conference on Smart Technologies*, 2021, pp. 170-174.
- [22] Marwa Kahia a , Amira Ectiouei b , Fathi Kallel c and Ahmed Ben Hamida d. ATMS Lab, "Advanced Technologies for Medicine and Signals", ENIS, Sfax University, Sfax, Tunisia 2022.