University of Mohaghegh Ardabili

Skin_Classification_Report

Mohammad Taha Najafzadeh

9721973144

Software project

Dr.Samad Najjar

2022-09-06

Table of Contents

1. Introduction

In terms of cancer types, skin cancer is the most commonly diagnosed one. There are a number of forms of skin cancer, but melanoma, which is the most common, is responsible for 75% of all deaths caused by skin cancer. According to the American Cancer Society, there will be approximately 100,000 new cases of melanoma diagnosed by 2020, which is an increase of approximately 60 percent from 2011. The national average is expected to be around 7,000 deaths caused by this disease during the course of this year. It can be very helpful if there is an early and accurate diagnosis of this form of cancer, as with other forms of cancer, in order to make treatment more effective-perhaps by using data science assisted by other methods.

There is currently no way for dermatologists to determine a patient's melanoma risk other than to examine every mole on their body hoping to find "ugly ducklings" or unusual lesions that might indicate the disease. Current artificial intelligence approaches have not been able to adequately consider this clinical frame of reference in their design process. Detecting melanoma might become more accurate if dermatologists were to develop an algorithm that could take into consideration "contextual" images within the same patient when determining which image represents melanoma within that patient. When dermatological clinics are able to get hold of classifiers that can be used to improve their accuracy and help them perform their work more efficiently, they will be able to improve their results.

It is the mission of the Society for Imaging Informatics in Medicine to provide education, research, and innovations in medical imaging informatics and to advance this field in a multidisciplinary environment. SIIM and the International Skin Imaging Collaboration (ISIC) are launching a joint initiative that aims at improving the diagnosis of melanoma at an international level through joint collaboration between the two organizations. The ISIC Archive provides the largest collection of quality-controlled dermoscopic images of skin lesions and is the only repository that provides these images at such an extensive level.

In this test you will be asked to identify melanoma using images of skin lesions that have been taken of your skin. Your goal will be to identify those images in a patient that might be indicative of a melanoma based on the type of images within the patient. An advanced image analysis tool that utilizes contextual information provided by patients may be of benefit to clinical dermatologists as they develop tools for image analysis.

It does not matter how deadly Mesothelioma is, in most cases, melanoma can be easily treated with minor surgical procedures if found early enough. It is anticipated that image analysis tools that automate the process of diagnosing melanoma will allow dermatologists to diagnose this disease more accurately in the future. More accurate way to detect melanoma in the future may make a significant difference to millions of people.

2. Objective

There is a need to identify melanoma in images of skin lesions with the aim of identifying melanoma by the current study. The latter involves determining which of the images in the same patient are likely to contain melanoma as we go through the images. Therefore, we must create a model which will make it possible to predict with a high degree of accuracy whether the lesion in the image is malignant or benign, depending on the value of the control points. Value 0 indicates benign, and 1 indicates malignant.

There are a lot of questions that are going to be answered in this report, including the following:

- *How's the data looking?*
- *Do we have complete dataset?*
- *How's the target distribution looking? Is it balanced?*
- *What are the effects of scan site on outcome?*
- *Does age effects skin lesion type?*
- *Is there difference between female and male patients in terms of target?*
- *How many unique patient data we have and how many scans they had? Is it important?*
- *Is image quality, colors, size have meaningful impact on the outcome?*
- *Can we see similar observations when we analyse both train and test dataset, if not why?*

3. Benign and malignant tumors

3.1. What is a benign tumor?

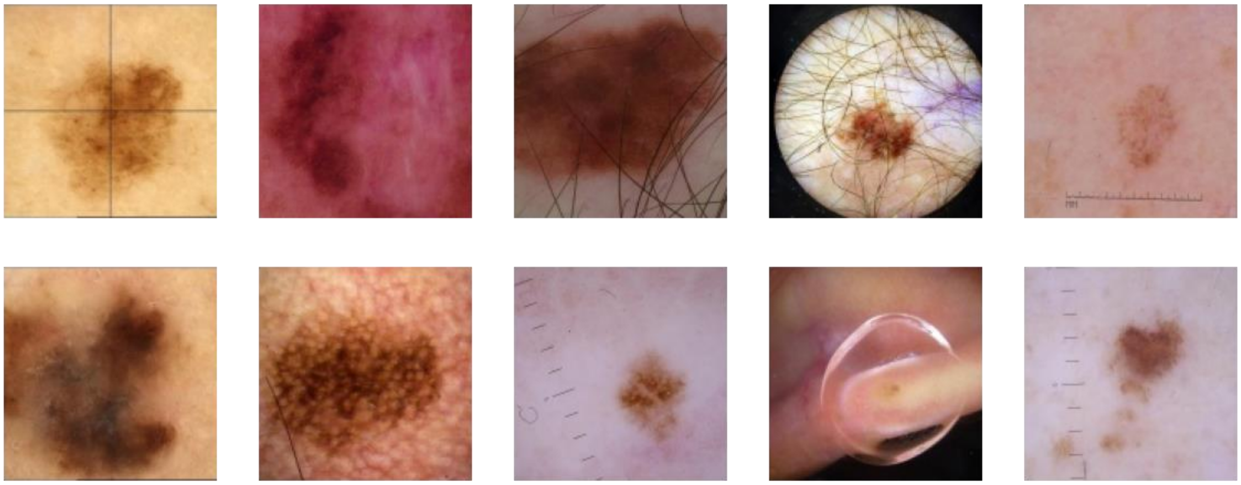
Quite simply, a benign tumor can be defined as one which will not experience any malignant growth in the future. There is no danger to your skin from it, it is just a small blot on the surface of your skin that will not cause any damage to it.

3.2. What is a malignant tumor?

Unlike benign tumors that do not cause cancerous growth, malignant tumors cause cancerous growth within the body.

I think there is a significant difference between benign and malignant tumors in this country. What causes this to be the case? As far as medical science is concerned, it is believed that most cancer cases are diagnosed with malignant tumors in order to reach cancerous conditions. Cancer patients are only able to reach Stage 3 and Stage 4 of the disease system when their condition is at a very critical level. This level includes Stages 3 and 4. There is no good reason for a cell to grow cancerous, and there is always a risk associated with it.

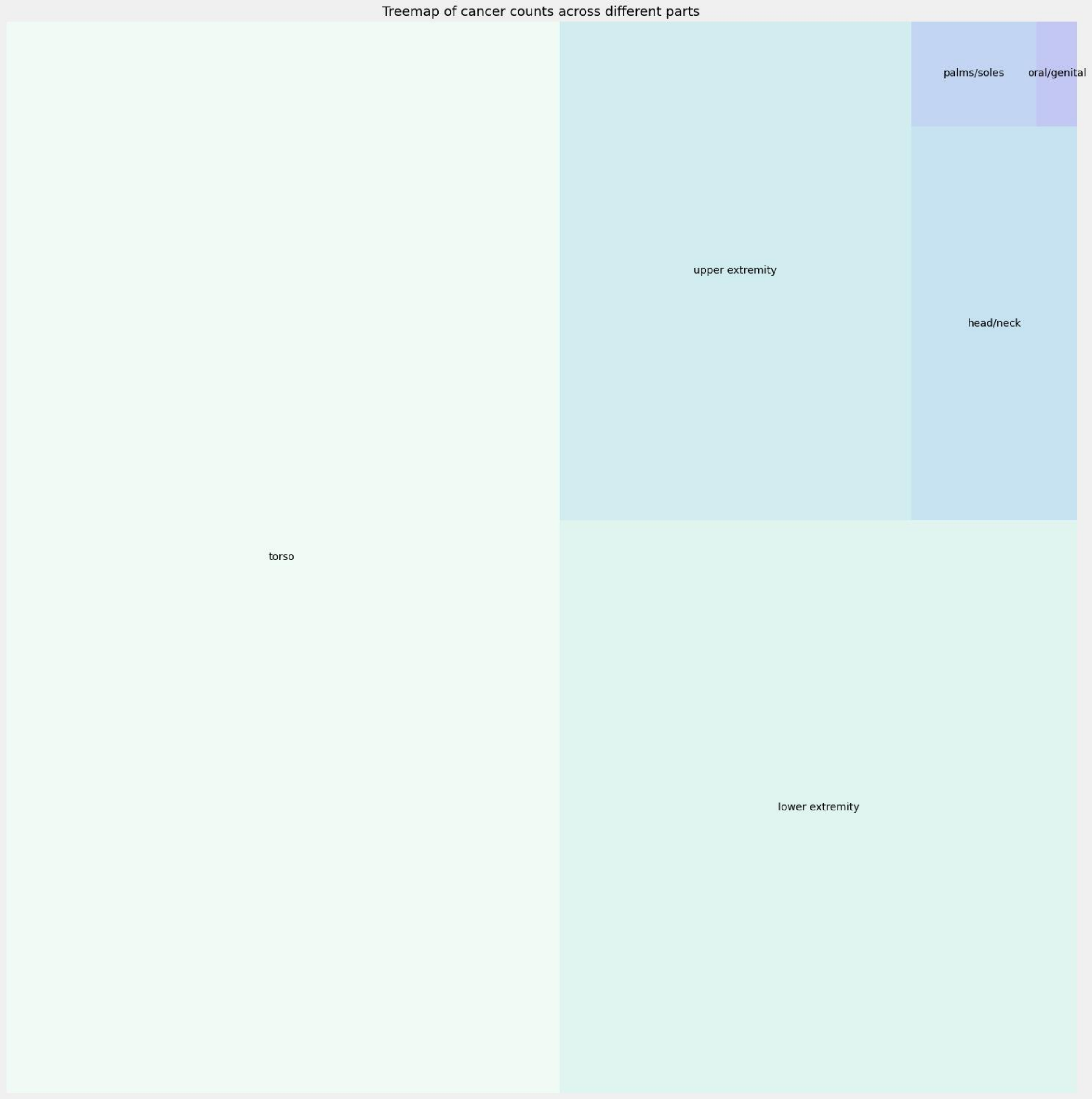
Cancerous growth often begins as a benign tumor and frequently paves the way for malignant tumors by allowing cancerous growth to progress. A malignant tumor can now be viewed in the following way:



3.3. Which part of the body?

There are many different types of skin cancer but you can find melanoma any place where the skin is. As such, it makes it pretty much unavoidable to avoid getting it if you get it because unlike heart and lung cancers, melanoma can grow anywhere on the body; it is not localized to just one spot or region.

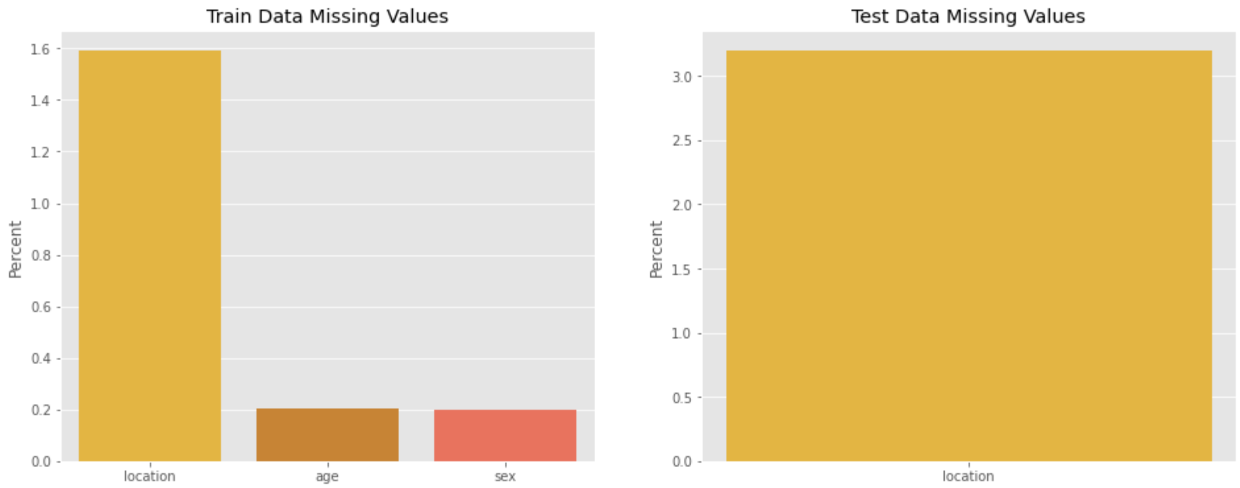
Here is a look at where cancerous growth occurs most frequently in the body:



4. Method

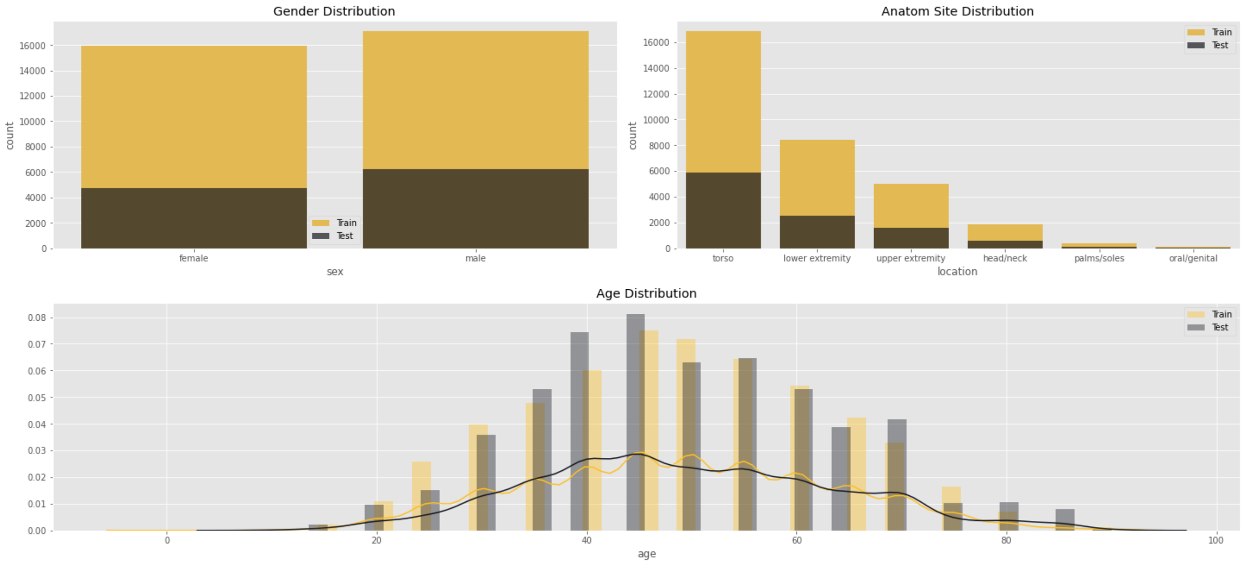
4.1. Missing Values

In the age and sex datasets, there is a small portion of missing values. It is not harmful, if these values are imputed with the most frequent ones, whereas for body parts missing in both datasets, it is better to set 'unknown' for missing values in this dataset. ...



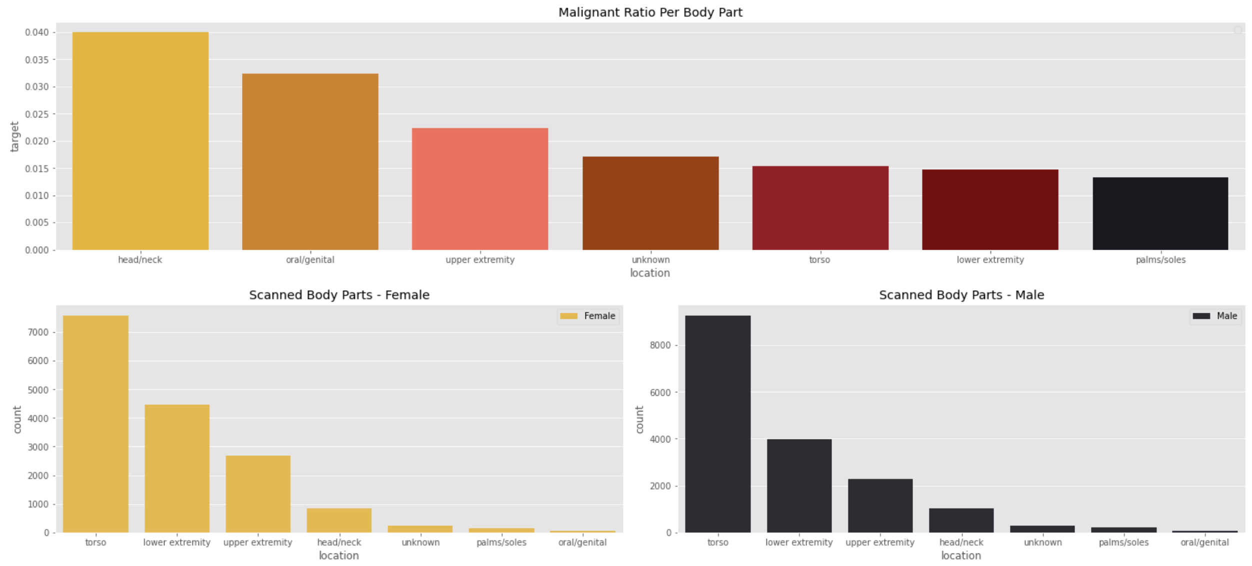
4.2. Checking Variables Before Imputing

As a preliminary step before imputed the missing variables, I just wanted to check their distribution. In the end, it seems that our assumptions were reasonable, so we will be able to proceed with imputing the data...



4.3. Body Part Ratio by Gender and Target

Head and neck cancer has been identified as one of the most common types of cancer, followed closely by oral/genital cancer and upper extremity cancer, in terms of likelihood of malignancy. Males and females differ slightly in terms of the order in which their various body parts are scanned depending on whether they are scanned as a whole or separately.



4.4. A General Look With Sunburst Chart

Sunburst chart is pretty cool looking fella I'd say. It also giving lots of basic information to us. Let's see...

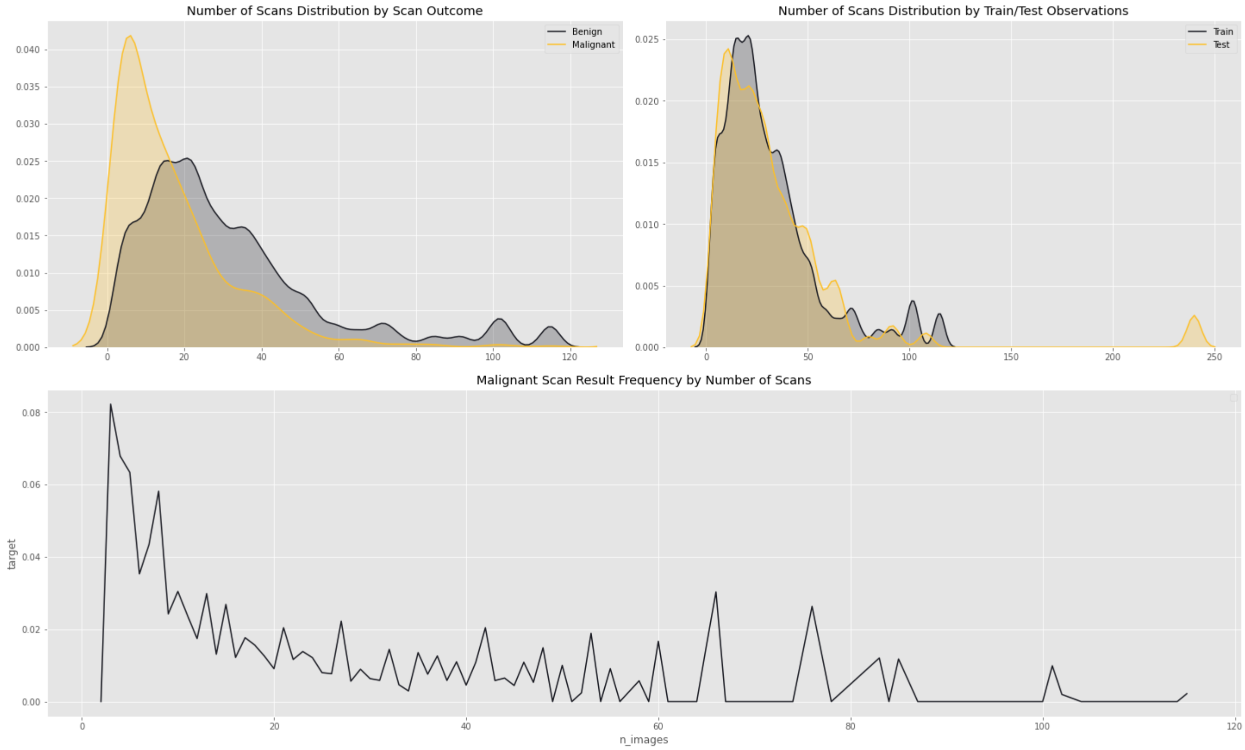
- *It is estimated that only 2% of our targets are malignant*
- *The male gender dominates malignant images with 62% of the images being male*
- *On a gender-specific level, benign images tend to have a higher female-to-male ratio, 52-48%*
- *A malignant image scan may be performed in a different location depending on the gender of the patient:*
    - *It is estimated that 50% of male scans are conducted in the torso, whereas 39% of female scans are conducted in the torso.*
    - *A scan of the male lower extremities reveals around 18% of cases while the scan of the female lower extremities reveals 26% of cases.*
    - *Malignant scans of the upper extremities are more likely to be performed on females than males (23-17%)*

- *The location of benign image scans in males and females is more similar than in either group*

4.5. Unique Patients and Their Scan Images

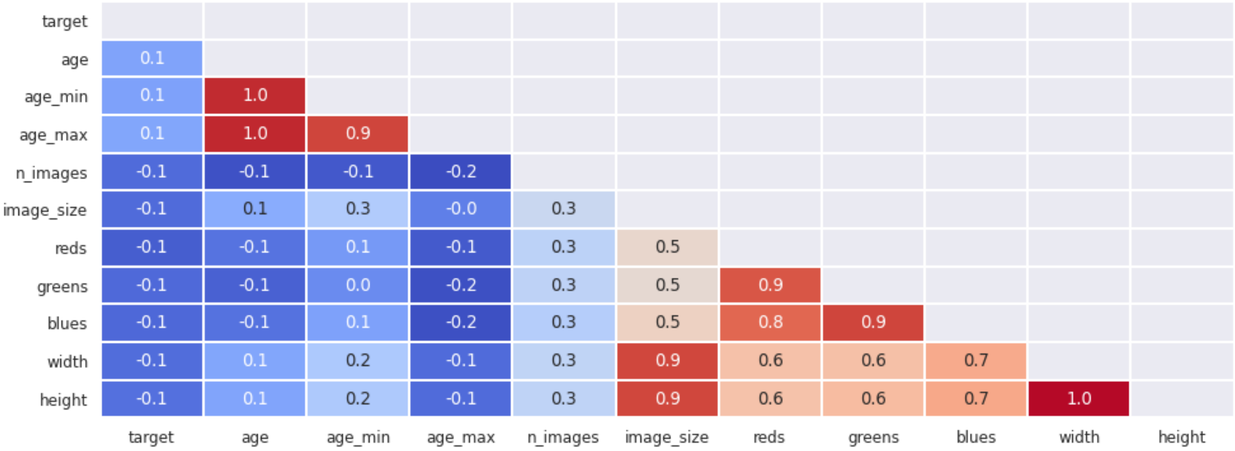
Interestingly, we seem to have multiple scan images for each patient, since the number of unique patients in actual datasets is much lower than what is shown in the images on these datasets. The information that we can obtain about a patient's age can be added to our database, such as when he had his first and last scan. Among the interesting insights we can gain are the following:

- *It is around the first 20 scans that we find the majority of malignant results. As a matter of fact, control scans can be conducted after the diagnosis has been made...*
- *I find it interesting that there are a lot of scan numbers in the first 100 scans, but there are more than 200 scan images for one patient in our dataset, which demonstrates that this is not a case that we have in our training data. There is a risk associated with this, and our model can be affected as a result of it.*
- *A majority of malignant cases consist of fewer than 20 images, but generally speaking, if there are more scan images, we can say that there is a higher chance that the result will be malignant...*



4.6. Correlations Between Features

A machine learning model may not be effective in making the necessary predictions based on all of the features in the dataset, depending on the purpose for which they are built. There are some features, however, that may actually make it worse for the predictions to be made if they are used. Consequently, when it comes to building a machine learning model, feature selection plays a large role in the process. Our objective here is to display the relationship between all the variables in the dataset by calculating correlations between all the metadata. The below analysis demonstrates that the metadata cannot be used in a meaningful way to provide any meaningful insights. In fact, the high correlation values, e.g., the age_min, the age_max, as well as the age variable, can be described as multiple positions of the same variable. This is why we don't use metadata in order to achieve the highest quality results, but rather work direct with the dataset to obtain the highest quality results.



	target	age	age_min	age_max	n_images	image_size	reds	greens	blues	width	height
target											
age	0.1										
age_min	0.1	1.0									
age_max	0.1	1.0	0.9								
n_images	-0.1	-0.1	-0.1	-0.2							
image_size	-0.1	0.1	0.3	-0.0	0.3						
reds	-0.1	-0.1	0.1	-0.1	0.3	0.5					
greens	-0.1	-0.1	0.0	-0.2	0.3	0.5	0.9				
blues	-0.1	-0.1	0.1	-0.2	0.3	0.5	0.8	0.9			
width	-0.1	0.1	0.2	-0.1	0.3	0.9	0.6	0.6	0.7		
height	-0.1	0.1	0.2	-0.1	0.3	0.9	0.6	0.6	0.7	1.0	

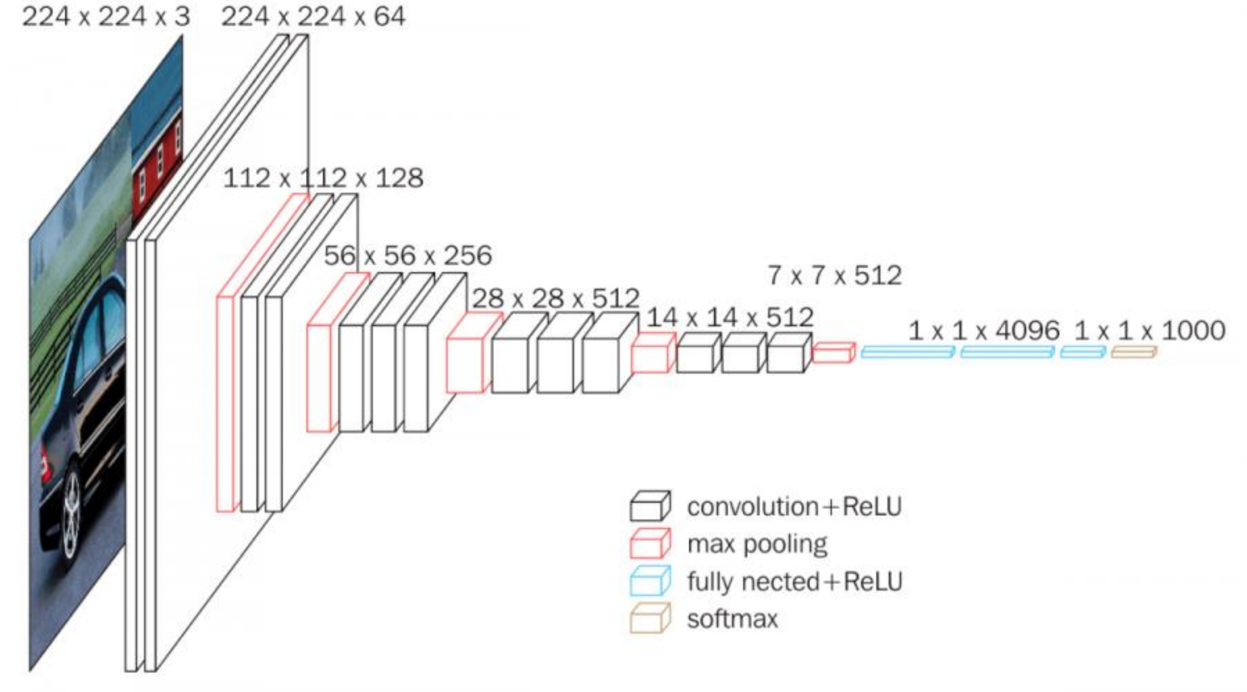
4.7. Preprocessing DIOCOM files

It is the standard for the exchange of information and management of medical images and associated data that is known as the Digital Imaging and Communications in Medicine (DICOM). Using the DICOM technology, medical imaging devices like scanners, servers, workstations, printers, network hardware, and picture archive and communication systems (PACS) from different manufacturers can be integrated for the storage and transmission of medical images.

There is an extension dcm at the end of DICOM images. It is important to understand that DICOM files are divided into two parts: a header and a dataset. An encapsulated dataset is described in a header which contains information about the encapsulated dataset. The preamble of the file is composed of a DICOM prefix, followed by the file meta elements and the file preamble itself. Python provides us with a library called Pydicom which can be used by us to read and manipulate DIOCOM files. Python's Pydicom library is made for interacting with these complex files and provides a convenient way to read and manipulate these files in a natural pythonic way. In order to re-write modified datasets to DICOM format files, they must be modified once again.

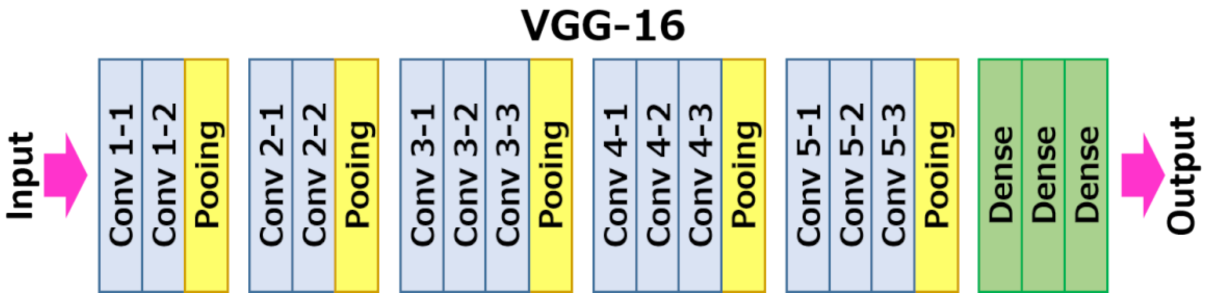
5. Classification Model Architecture

5.1. VGG16



5.1.1. Overview

Several papers have been published on the subject of the development of very deep convolutional neural networks for larger-scale image recognition in which K.Simonyan and A.Zisserman developed a model of convolutional neural networks called the VGG16 model. The accuracy of the test in ImageNet, a dataset containing over 14 million images categorized into over 1,000 categories, was found to be 92.7%, which is among the highest in the test. There is no doubt that it's one of the best models out there, as it was recognized by ILSVRC-2014. Compared to AlexNet, this method improves the accuracy of the convolutional layer by eliminating large kernels-sized filters (8 in the first convolutional layer and 5 in the second convolutional layer, respectively). We trained VGG16 on the NVIDIA Titan Black GPUs over a period of weeks using the GPUs that were provided by NVIDIA.



5.1.2. Pretrained Dataset

There are more than 15 million labeled high-resolution images being used in the ImageNet dataset, which comes with approximately 22,000 categories organized according to the images. The images were collected from the web using a crowdsourcing tool offered by Amazon called Mechanical Turk, and they were labelled manually by humans using a tool called HumanLabeler. This global competition was launched as a part of the Pascal Visual Object Challenge in 2010 and has been running every year since then. The ILSVRC aims to promote large-scale visual recognition. A subset of ImagesNet is used in ILSVRC to select approximately 1000 images from each of 1000 categories in order to meet the criteria for the project. Approximately 1.2 million training images made up of 50,000 validation images and 150,000 testing images made up of 50,000. There will be 12 million total training images used in this project. Images of various resolutions are included in ImageNet, which makes it possible to search for them. The result of this have been a down scaling of the images to a fixed resolution of 256x256 using a mix of different resolutions. It is used to remove the 256x256 pixel patch in the center of a rectangular image given to it after it is rescaled and cropped.

5.1.3. The Architecture

The input picture for the cov1 layer is a RGB image of 224 by 224 pixels. It is used as an input picture for the layer. On the image, we used several convolutional layers (conv.) with a receptive field of 3x3 (which is the smallest size that allows us to capture the concepts of left/right, up/down, and center), in which stacked convolutional layers were used. The 1x1 convolution filters can also be viewed as a linear transformation of the input channels (followed by a non-linear transformation) in one configuration, a process that can be viewed as a linear transformation of the input channels. Convolution stride would be 1 pixel; spatial padding would be 1 pixel, so it is fixed that the stride of this convolution would be 1 pixel. A three-by-three convolution is performed using layer input in order to include the spatial resolution when the convolution is performed, i.e. one pixel is left for the layer padding in a three-by-three convolution. Spatial pooling is carried out using a maximum-pooling algorithm, using five layers within the convolution and a maximum number of convolutions. Max-pooling cannot be applied to all the layers of convolution at the *same time. A maximum-pooling process is performed on a 2x2 pixel window with a stride of two pixels on top of it.*

Three Fully Connected (FC) layers follow a stack of convolutional layers, which vary in depth depending on the architecture, after which there is a stack of convolutional layers. These layers have 4096 channels for each, while the third one uses 1000 channels to perform 1000-way ILSVRC classification, resulting in 1000 channels for each class. I would like to finish off by saying that the final layer is the hard max layer. The fully connected layers of every network have the same configuration as far as their configurations are concerned.

Each of the hidden layers demonstrates the presence of the rectification (ReLU) nonlinearity. In addition, all the networks (with the exception of one network) do not make use of Local Response Normalization (LRN), which does not provide any improvement in performance on the ILSVRC dataset, but instead increases the amount of memory consumed as well as the amount of computation required to generate them.

5.2. VGG19

5.2.1. Overview

VGG19 contains 19 layers of convolution on top of 5 pooled layers, two soft layers, and one convolution layer on top of 16 convolution layers. It is noteworthy that there are other variants of VGG, other than the VGG11, VGG16, and others, that are also using VGG. During the development of VGG19, 19.6 billion FLOPs were produced.

5.2.2. Background

As a result of AlexNet improving the traditional Convolutional Neural Networks in 2012, the Visual Geometry Group at Oxford University developed the Visual Geometry Group. Thus, the name, Visual Geometry Group, was born. The algorithm uses deep convolutional neural layers, as with its predecessors, for improving accuracy and incorporating some ideas that were part of the predecessors' algorithms.

Describe what VGG19 is, compare it to other versions of the VGG architecture, and make use of it in useful and practical ways. These are some of the most useful and practical applications of the VGG architecture.

It is important to start with an understanding of the VGG19 Architecture in order to better understand how it works. We must review both ImageNet and CNNs in order to begin our journey.

- *Convolutional Neural Network(CNN)*

It would be helpful if we could first define the term ImageNet in order to gain a better understanding of it. This database contains approximately 14,197,122 images, all of which are arranged in the same manner as the WordNet hierarchy, making it easy to find the image you are looking for. This initiative is aimed at bringing together professionals and students working in the field of image and vision research, as well as other individuals interested in this field.

In addition to its large-scale visual recognition challenge, ImageNet also conducts competitions related to its large-scale visual recognition challenge, such as the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) specific to ImageNet. Researchers from around the world are challenged with the task of identifying the best solution that is able to achieve the lowest top-1 error rate and top-5 error rate (top-5 error rate refers to the number of images whose labels do not match one of the five most likely labels in the model) for this challenge. We will use 1.2 million images in this competition in order to train 1,000 classes, 50 thousand images in order to validate class classifications, and 150 thousand images in order to test how well our system does.

According to the VGG Architecture, in 2014 during the peak of the state-of-the-art model, it was the fastest growing state-of-the-art model, and it is still being preferred for many difficult problems today.

5.2.3. The Architecture

- *There were 224,224,3 RGB pixels in the RGB image that we provided this network with, resulting in a matrix that was 224,224,3, which means that the matrix shape we gave was 224,224,3.*
- *A preprocessing step has been performed in which the mean RGB value has been subtracted from each pixel over the entire training set, which will be used as a reference for the prediction.*
- *In order to cover both ends of the image it was necessary to use kernels of dimensions (three by three) and stride sizes (one pixel) of 3 pixels.*
- *Spatial padding has been used in the process of preserving the spatial resolution of the image.esolution of the image.*
- *Using Sride 2, the maximum pooling was performed over a window of 2 x 2 pixels in order to achieve maximum pooling.*
- *Thereafter, a modified linear unit (ReLu), which was implemented in order to introduce nonlinearity into the model, as well as to improve its classification performance and its computational time, was used to improve this model since previously, the models had been based on tanh or sigmoid functions, which proved much more effective in comparison with those models.*
- *This was achieved by implementing three layers that were both fully connected, the first two of which had 4096 channels in size and the third layer in which there were 1000 channels for a 1000-way ILSVRC classification. The third layer of the algorithm was the implementation of a softmax function.*

5.3. Inception v3

5.3.1 Overview

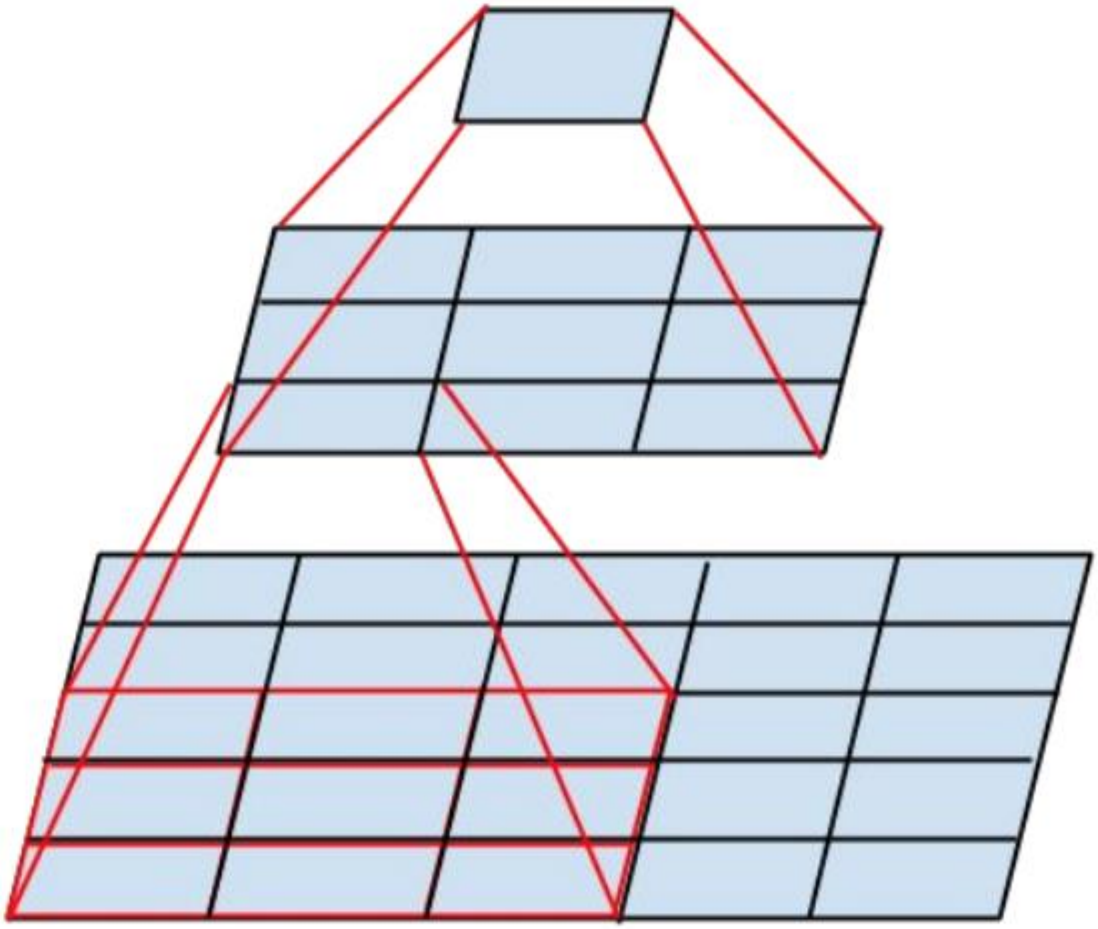
Inception v3 has been updated to include modifications to the underlying architecture of the application, so that it consumes less processing power under the hood. In the paper Rethinking the Inception Architecture for Computer Vision, which was published in 2015, this idea was first introduced as a part of the broader concept of Computer Vision. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and others contributed to the preparation of the manuscript, including a number of co-authors. Having compared the computation efficiency of Inception Networks (GoogleLeNet/Inception v1) with VGGNet, an analysis of both their parameters generation and their cost incurred for memory and other resources indicates that Inception Networks are more efficient, not only in terms of the number of parameters generated, but also in terms of memory and other resources used. Any changes made to an Inception Network should be taken into account in order to preserve the computational advantages of the network and to ensure that those advantages will not be lost in the process. Adapting a network built around Inception for a variety of different applications is difficult because there is no certainty about how efficient the new network will be. The optimization of the network in the Inception v3 model has been achieved using several techniques, in order to facilitate easier model adaptation. Aside from factorized convolutions, regularization, dimension reduction, and parallel computations, it is also possible to utilize several other techniques in addition to factorized convolutions, regularization, and dimension reduction.

5.3.2. Inception v3 Architecture

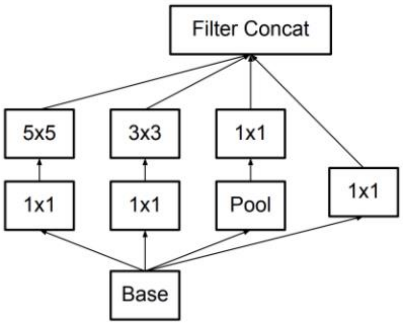
A network will be constructed by gradually undertaking and completing the following steps step by step during the construction of an Inception v3 network:

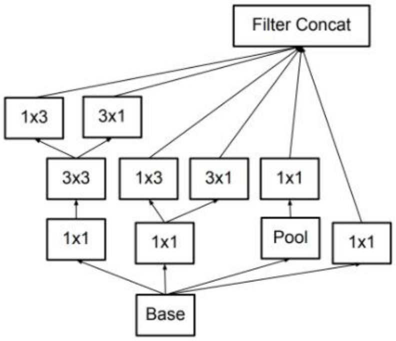
1. In the case of factorized convolutions, the computational efficiency of a network is reduced as a result of the reduction of parameters in the network. As an added benefit, it ensures that the network remains efficient and that it maintains its function as a whole.

2. A study has shown that smaller convolutions are more efficient for training than larger convolutions, as a result of the fact that smaller convolutions have smaller outputs. In this case, two three-dimensional (3-dimensional) filters are replacing the five-dimensional (5-dimensional) convolution, which reduces the number of parameters to 18 instead of 25 in the case of two three-dimensional (3-dimensional) filters.

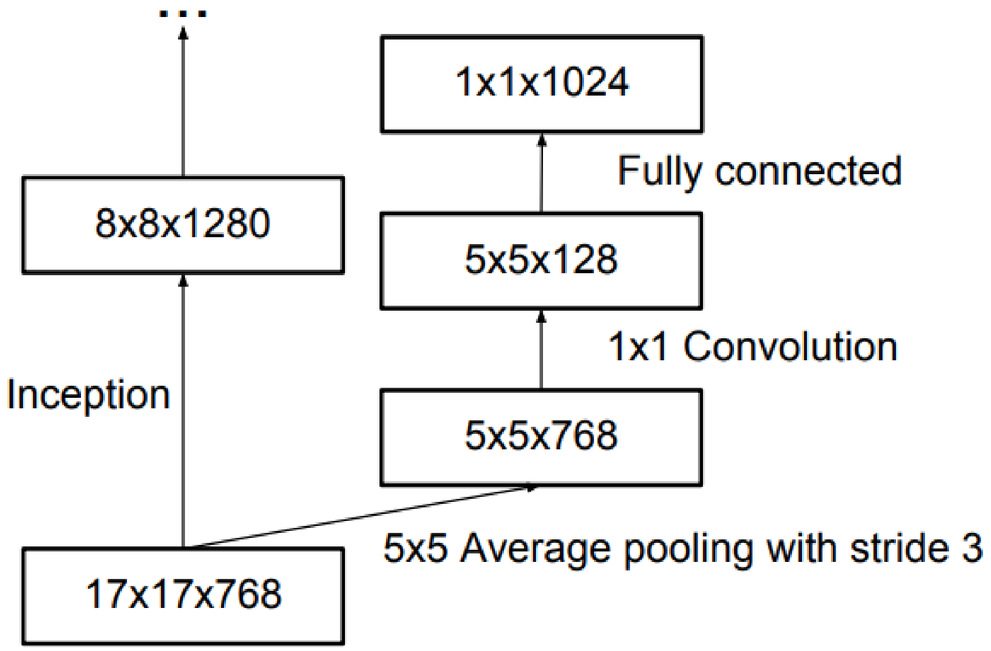


3. It is possible to replace the 3 x 3 convolution with an asymmetric convolution that can work as well. It is possible, however, to replace a 3 x 3 product with a 1 x 3 product followed by a 3 x 1. In the proposed asymmetric convolution, there would be more parameters than in the proposed convolution of 2x2 but the number of parameters would still be slightly higher than the one proposed for convolutions of 3x3.

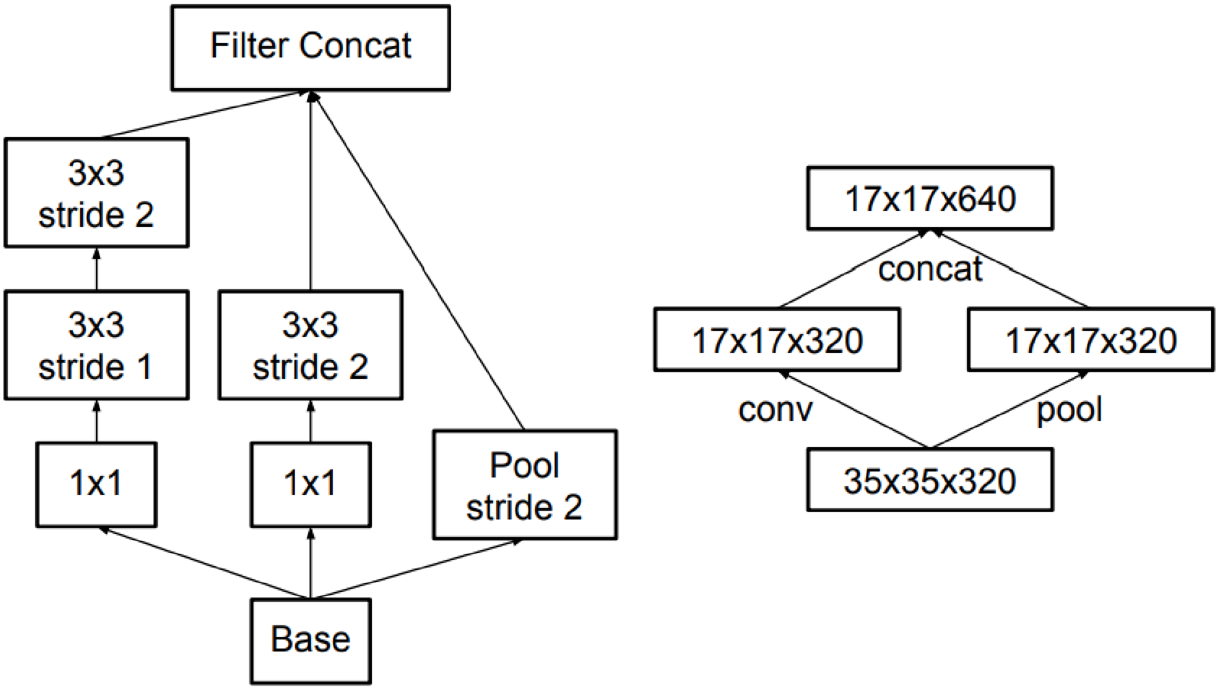




4. An auxiliary classifier is also a concept that can be applied to the training process. It is composed of a smaller CNN that is inserted between layers during the training process and its loss is added to the loss of the main network in order to determine the final result. Auxiliary classifiers are used in GoogleNet to make the network more deep, as opposed to basic classifiers in Inception 3, which are used to make the network more regular.



5. It is common for grid size to be reduced through pooling operations. It is possible, however, to address the bottleneck of computational costs more efficiently by using the following method:



6. Evaluation and Results

6.1. Augmentations and training setup

A significant part of the issue of overfitting can be prevented by enhancing images for small or medium sized datasets. From the Pytorch augmentation library Albumentations [1], which is a powerful and popular library of augmentations, we used the following augmentations in our pipeline: Transpose, Flip, Rotate, RandomBrightness, RandomContrast, MotionBlur, MedianBlur, GaussianBlur, GaussNoise, OpticalDistortion, GridDistortion, ElasticTransform, CLAHE, HueSaturationValue, Shift ScaleRotate, Cutout[2]. A before-and-after picture of these augmentations can be seen in Figure 2. Cosine annealing was used for training schedules with one warm up epoch in order to compute training schedules. As a rule of thumb, most models have a total of 15 epochs.ch is the maximum number of epochs. Depending on the model, the initial learning rate of the cosine cycle is varied, ranging from 1 e 4 to 3 e 4, which is suitable for each situation. Warm-up epochs have always been characterized by a learning rate that is one tenth of the initial learning rate of the cosine cycle during this period. For all models, there is a batch size of 64. Throughout the training process, NVIDIA Tesla V100 GPUs were used in mixed precision mode for all training. It was possible to use up to eight GPUs simultaneously in this study.

6.2. Training results

Initially, you might want to play with the smallest learning rate that you have available to you. The results of our experiments seem to suggest that the size of the first one definitely influences the success and how much greater the amount of learning you are able to achieve with these curves.

In addition, it is also evident that when the score continues to rise, there can be some fluctuations in the loss as well!

	VGG16	*VGG19*	*InceptionV3*
Accuracy	*0.97799511002444*	*0.80032599837000*	*0.96821515892420*
Precision	*0.25714285714285*	*0.01902748414376*	*0.14285714285714*
Recall	*0.24324324324324*	*0.25714285714285*	*0.09433962264150*
F1-Score	*0.25*	*0.3543307086614*	*0.11363636363636*

There is no doubt that F1-scores can represent an overall network's performance best, since they indicate the harmonic mean between precision and recall across the network. In order to give a score for each model in the F1-test, we calculated F1 scores for all three models, of which the VGG19 model had the highest score with 0.35. Since our goal was not to optimise the performance of the network, we did not include any additional steps during our experiments.

Using our curated balanced dataset as the training data, we were able to train all forms of network architectures and get comparable results. We would like to point out, however, that the ISIC 2020 testing set we used to assess the performance of our models is much larger than the training set we used (7848 training examples vs 10,982 test examples). Moreover, performance on ISIC 2020 testing set may be further impacted by the possibility that there may be class imbalances within the set. As well, we would like to point out that we have also identified 78 duplicate image files within the ISIC 2020 test set, based on a FSlint report. Currently, there is no information about the composition of the ISIC 2020 test set, which means that we can only speculate about the effects when obtaining evaluation metrics by using a custom test consisting of 10% of malignant cases and 10% of benign ones.