POLSKO-JAPOŃSKA AKADEMIA TECHNIK KOMPUTEROWYCH

Wydział Informatyki
Katedra Sztucznej Inteligencji
Specjalność: Data Science

Jan Solarz
Nr albumu: 26342

DETEKCJA RAKA PIERSI PRZY UŻYCIU KONWOLUCYJNYCH SIECI NEURONOWYCH

Praca Magisterka
dr hab. inż. Grzegorz Wójcik

słowa kluczowe:
uczenie głębokie, konwolucyjne sieci neuronowe, klasyfikacja obrazu, medycyna

Warszawa, wrzesień, 2023

krótkie streszczenie:

Celem pracy jest próba detekcji nowotworu piersi u kobiet przy użyciu zdjęć mamograficznych. Zostaje zobrazowany problem procedur i złożoności procesu radiologicznej oceny zdjęć i konsekwencji jakie w związku z nim powstają. Pojawia się aspekt odpowiedniej automatyzacji prac detekcji chorób przy wsparciu rozwiązań uczenia głębokiego w medycynie. W niniejszej pracy występuje przegląd wybranych konwolucyjnych sieci neuronowych oraz ich omówienie pod kątem budowy. Następnie wykorzystując przedstawione podejścia zostają one zaimplementowanie w celu uzyskania jak najlepszych rezultatów na prawdziwych danych mammograficznych.

POLISH-JAPANESE ACADEMY
OF INFORMATION TECHNOLOGY

Faculty of Computer Science
Artificial Intelligence Cathedral
Specialty: Data Science

Jan Solarz
Album nr: 26342

SCREENING MAMMOGRAPHY BREAST CANCER DETECTION USING CONVOLUTIONAL NEURAL NETWORKS

Master thesis
Dr. habil. Grzegorz Wójcik

keywords:
deep learning, convulutional neural netowrks, image classification, medicine.

Warszawa, September, 2023

short summary:

The aim of the work is an attempt to detect breast cancer in women using mammographic images. The problem of procedures and complexity of the process of radiological assessment of images and the consequences that arise in connection with it are illustrated. There is shown the aspect of proper automation of disease detection work with the support of deep learning solutions in medicine. In this work, there is a review of selected convolutional neural networks and their discussion in terms of construction. Then, using the presented approaches, they are implemented in order to obtain the best results on real mammography data.

Table of Contents

Introduction

Cancer is a leading cause of death worldwide, and early detection is crucial for improving patient survival rates. Radiologists play a critical role in cancer detection, and the accuracy of their diagnosis can significantly impact patient outcomes.

Deep learning has emerged as a powerful tool to assist radiologists in identifying subtle patterns in medical images that may not be visible to the naked eye. By training deep learning algorithms on large datasets of medical images, radiologists can improve the accuracy and speed of cancer detection.
Radiology has been one of the primary fields in medical imaging that has advanced significantly in recent years with the integration of deep learning algorithms. That field of data science has shown remarkable progress in various applications, including medical imaging analysis, and has demonstrated great potential in cancer detection. The ability of deep learning algorithms to extract and analyze complex patterns in medical images has made it an invaluable tool for radiologists to accurately diagnose cancer at an early stage, thus improving patient outcomes.

The number of people being diagnosed with cancer by 2040 in EU and EFTA countries is estimated to increase by 21% compared to 2020. This is the finding of JRC experts who studied the impact of population aging on the future cancer burden. [4]
Based on that fact which is related with climate change and development in ways of feeding, the population should be prepared to treat and detect cancer more efficiently, due to the higher probability of the diseases that are already known.

In this thesis, it will be explored the current state of deep learning in radiology, specifically in breast cancer detection as binary classification problem. It will be examined the various deep learning techniques used in medical imaging analysis, including convolutional neural networks (CNNs) and autoencoders, and their applications in cancer detection. Additionally, it will be discussed the challenges and limitations of deep learning in radiology, such as the need for large amounts of high-quality data, the lack of interpretability of deep learning models, and the potential for bias in algorithm development. Finally, we will examine the future directions of deep learning in radiology and discuss the potential for the integration of other technologies, such as augmented reality, to improve cancer detection and diagnosis.

Overall, the work aims to provide an in-depth analysis of the current state of deep learning in radiology, specifically in cancer detection, and to identify the potential future directions of this exciting field.
Deep learning has been making significant progress in the field of radiology over the past few years. Radiologists have been utilizing deep learning algorithms to help them analyze

medical images and detect anomalies more accurately and efficiently, but still more in the scientific way. Deep learning models are capable of identifying patterns in medical images that might be difficult for the human eye to detect, making it an invaluable tool for radiologists.

> "What the AI tools are doing is they're extracting information that my eye and my brain can't" *Constance Lehman, radiologist, Massachusetts General Hospital.* [20]

The combination of deep learning and radiology has opened up new possibilities for the detection of cancer. These technologies have the potential to improve the accuracy and efficiency of cancer diagnosis, ultimately leading to better patient outcomes.

Overall, radiologists seem to be optimistic about the potential of deep learning to enhance their work. However, there are also concerns about the accuracy and reliability of deep learning algorithms. Some radiologists worry that these models may not always provide consistent results, and there is a risk of false positives or false negatives.

Mammograms can detect breast cancer in the early stages, even before someone notices a lump.

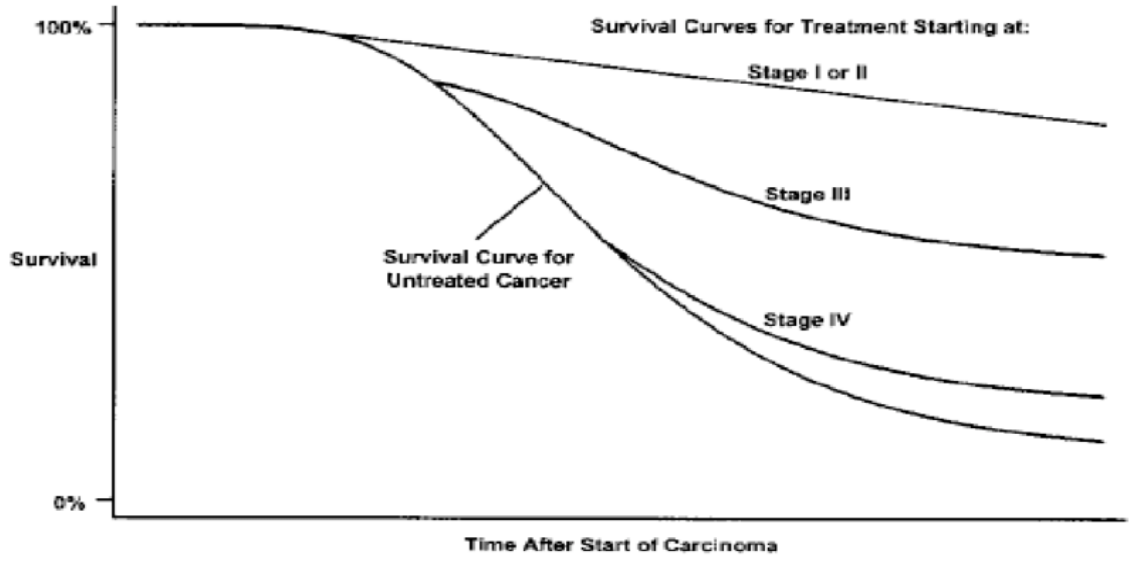

Figure 1: Chart of breast cancer survival. A conceptual graph showing the percentage of breast cancer survivors as a function of the stage in which they received treatment. [6]

In the Figure 1 above is shown the difference between stages of cancer which are related to spreading to other tissues and the following consequences. Stage I means that the tumor is localized to a small area and hasn't spread to lymph nodes or other tissues and stage IV refers that the cancer has spread to other organs or areas of the body metastatic or advanced cancer (metastatic or advanced cancer).

Early detection allows doctors to treat breast cancer more successfully. The American Cancer Society (ACS) states that regular mammograms are an essential part of routine annual healthcare for females specially in the group over the age of 45 years with an already average risk of breast cancer.

Chapter 1

Radiology as a challenge in Deep Learning

Breast cancer is the second most fatal disease in women and is a leading cause of death for millions of women around the world [27]. According to the American Cancer Society, approximately 20% of women who have been diagnosed with breast cancer die [2].

Mammography is a type of medical imaging that uses low-dose X-rays to create images of the breast tissue. The procedure is typically performed to screen for breast cancer or to investigate abnormalities found during a clinical breast exam. The following are the stages involved in the mammography process.

The accuracy of cancer classification in mammography for radiologists can vary depending on a number of factors, including the quality of the images, the experience and skill of the radiologist, and the specific characteristics of the cancer being detected. In general, however, mammography is considered a highly accurate screening tool for detecting breast cancer.

In Europe, breast cancer screening programs vary by country and are typically offered to women between the ages of 50 and 69, although the age range may vary depending on the country. According to a report by the European Commission, the percentage of eligible women participating in breast cancer screening programs in the European Union ranged from 15% to 89% in 2016, with an overall participation rate of about 50%.
In the United States, the American Cancer Society recommends that women at average risk of breast cancer begin getting yearly mammograms at age 45, with the option to begin screening as early as age 40. According to the National Cancer Institute, an estimated 39 million mammograms were performed in the United States in 2018.

Globally, the utilization of mammography screening programs varies widely depending on factors such as access to healthcare and cultural attitudes towards cancer screening. According to a report by the International Atomic Energy Agency, breast cancer screening programs are most commonly available in high-income countries, with low- and middle-income countries having more limited access to screening services. However, efforts are being made to increase access to mammography screening in low- and middle-income countries.
In 2018, almost 67% of females over the age of 40 years in the United States had a mammogram to check their breast health. Sometimes, knowing what to expect before and

after can help people feel more at ease during the process.

The mammogram results indicate whether a person has signs of abnormalities in the breast. Abnormal findings do not necessarily mean that the person has breast cancer. The ACS states that less than 10% of people with an abnormal mammogram have cancer.
However, doctors typically call back anyone with abnormal mammogram results to rule out any problems or begin any necessary treatment as soon as possible. Typically, a person will learn the results during this follow-up appointment.
When the person returns, they will likely have a diagnostic mammogram rather than a screening mammogram, as a diagnostic mammogram takes more pictures. They may also have an ultrasound test, which creates an image of the inside of the breasts using sound waves.
A radiologist may decide that an MRI scan of the breast will provide more information about the abnormality. In some cases, they may ask the individual to return to have a biopsy of the breast tissue.
A surgical or needle biopsy takes a piece of the abnormal breast tissue so that a technician can examine the cells under a microscope to determine whether they are cancerous.

1.1 Factors of breast images

In this section it will be approximate issues regarding how exactly the cancer look like on the imaging and at which factors should take attention a person not from the field of oncology.
Like other X-ray images, mammograms appear in shades of black, gray and white, depending on the density of the tissue.

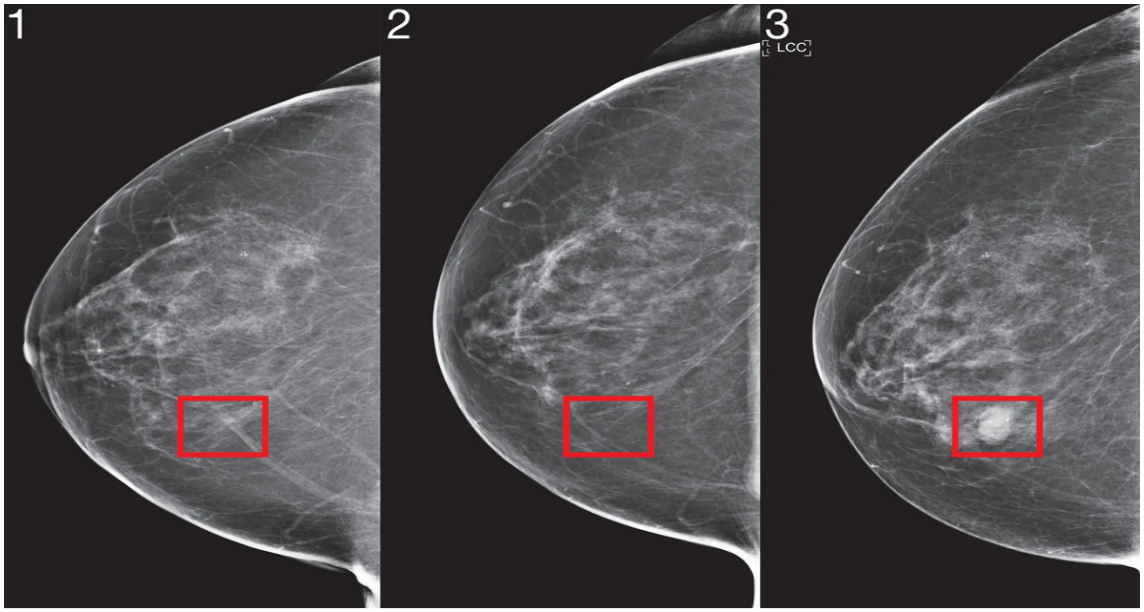

Figure 1.1: Images from a mammogram of a patient whom the algorithm identified as high risk four years before cancer was diagnosed. Courtesy of MIT. [21]

Very dense tissue, like bone, shows up as white on an X-ray. Fat looks dark gray on an X-ray.

Breast cancer and some benign breast conditions are denser than fat and appear a **lighter shade of gray or white** on a mammogram. Dense breast tissue can look **light gray or white** on a mammogram. This can make abnormal findings on a mammogram hard to see.
Younger women tend to have dense breast tissue, so their mammograms can be harder to read than the mammograms of older women.
Women with dense breasts (as seen on a mammogram) have a somewhat higher risk of breast cancer than women with fatty breasts [7]. However, breast density does not appear to be related to breast cancer survival [9].

1.2 All aspects of radiology procedures. The following radiology procedures

There are several important stages during the mammography test, which is important to be aware on their course from the radiological side.
At first there is the image acquisition, based on the taking images from different angles during compressing the breast tissue between two plates. Then follows the interpretation procedure which means the mammogram images are sent to a radiologist, who will examine them for any signs of breast cancer or other abnormalities. The radiologist will look for any lumps, classifications, or other irregularities in the breast tissue. On the reporting stage it will be prepared a report of their findings and sharing the results with the patient's healthcare provider. If any abnormalities are found, the healthcare provider may order further testing or refer the patient to a specialist.

The time and cost associated with mammography results can vary depending on factors such as the type of mammogram performed, the location of the imaging facility, and the interpretation of the results by the radiologist. In general, mammography is considered a relatively quick and inexpensive procedure compared to other medical imaging tests, and it is often covered by insurance.

The time it takes for a radiologist to check mammography images and interpret them for cancer detection can vary depending on a number of factors, including the complexity of the case and the workload of the radiologist. In general, however, it typically takes around 10-15 minutes for a radiologist to review and interpret a mammogram. This can vary depending on the number of images taken and the quality of those images, as well as the experience and skill of the radiologist performing the interpretation.
It's important to note that mammography is not always 100% accurate, and some breast cancers may not be visible on mammogram images. In cases where the mammogram results are inconclusive or there is a suspicious finding, additional imaging tests or a biopsy may be necessary to confirm or rule out a cancer diagnosis.

1.3 The classification problem as cancer detection

Unlike traditional machine learning methods that require hand-engineered feature extraction from input images, deep learning methods learn the image features by which to classify data.

Image classification is the task of predicting the class or label of an entire image and can be binary (two classes) or multiclass (more than two).
There are several paths of goals to achieve in radiology in case of working with images: image classification, object detection, semantic segmentation or instance segmentation. In classification, the data are images with category labels. In detection, the data are images and rectangular bounding box coordinates delimiting features of interest. In segmentation, the data are images and image masks that provide labels for each pixel or voxel.
Medical images need labels to be used for supervised learning, the most common form of machine learning, in which the goal is to predict labels for new inputs. Depending on the task, labels for classification may arise from radiology reports, expert reviews, or clinical or pathologic data. Labels for detection and segmentation tasks are more complicated and time-consuming to create compared with classification datasets. Distributing the labeling task among more human labelers reduces the labeling burden on individuals but increases overall labeling work and raises consistency issues that may require averaged or consensus labels among several labelers.

The examples of results which are shown in this chapter were evaluated by radiologists, not by any AI algorithms. It is also important to mention that however the tests come from different sources and were tested on different group of women (which depend on age, nationality, habits etc.)

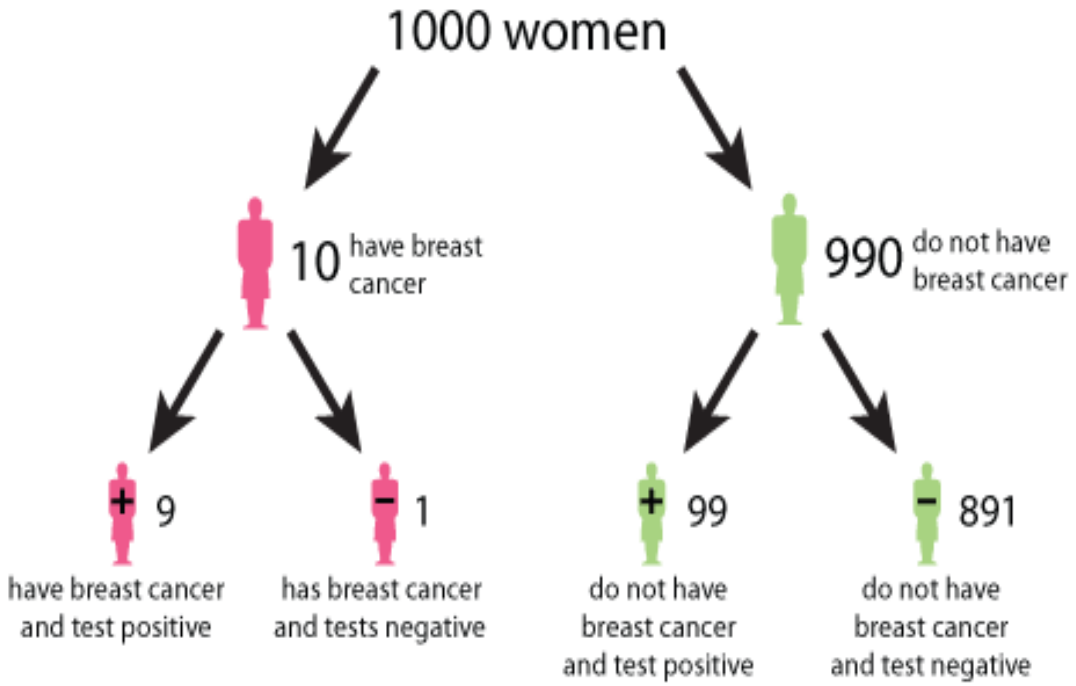

Figure 1.2: A tree diagram describing the outcomes of a mammography test. [15]

According to the American Cancer Society, mammography has a sensitivity of about 87% for detecting breast cancer, meaning that it correctly identifies about 87 out of 100 cases of breast cancer. The specificity of mammography, or the ability to correctly identify cases that are not cancer, is also high, at about 95%.
It's important to note that mammography is not infallible, and some breast cancers may

not be visible on mammogram images. In some cases, additional imaging tests or a biopsy may be necessary to confirm or rule out a cancer diagnosis. Additionally, false positives (when a mammogram indicates a possible cancer that is later found not to be cancer) and false negatives (when a mammogram misses a cancer that is later diagnosed) can occur, although the rate of false positives is generally higher than the rate of false negatives.

It is obvious that classification tasks, specially binary should be the simplest. On the other hand if the goal is to predict if someone does have the cancer or no the outcome of the prediction should be as much accurate as it is possible, taking all of the factors like patient's mental safety into consideration.

1.4 False positive cases

False-positive cancer screening test results are common. Over 10 years, approximately 50–61% of women undergoing annual mammography and 10–12% of men undergoing regular PSA testing will experience a false-positive result. [17]

False-positive cancer screening test results may affect individuals' willingness to continue screening for cancer in the future; about 40% of women experiencing a false-positive mammogram labeled the experience as "very scary" or the "scariest time of my life"[19]. On the figure below is shown interpretation of the positive cases from another angle of view.

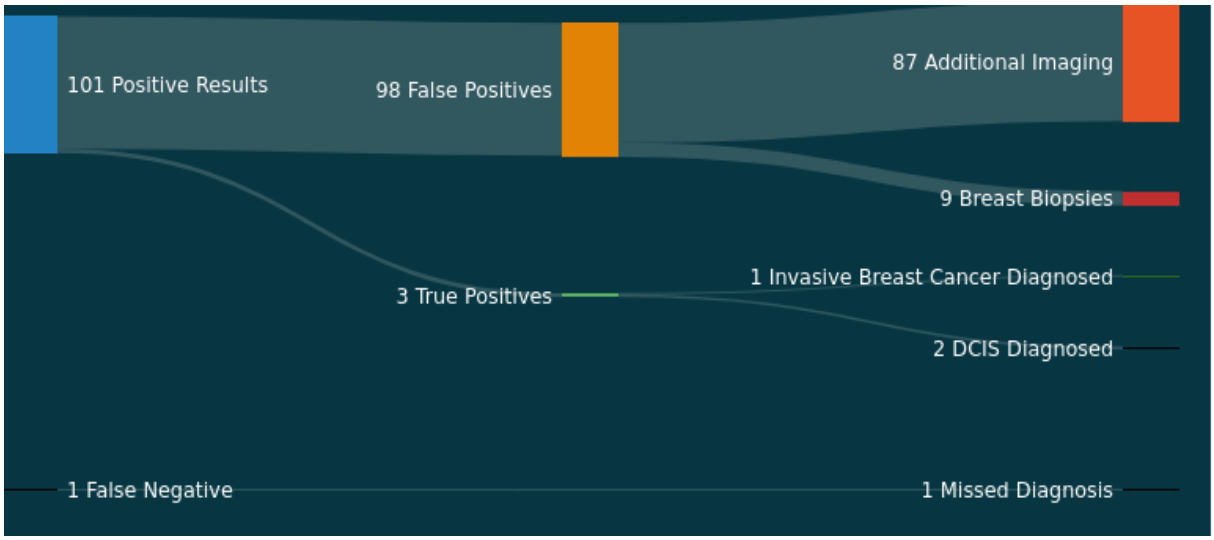

Figure 1.3: The results of finding from the U.S. Preventative Services Task Force on Breast Cancer Screening for women in their 40s.

One of the most significant consequences of false positives is the unnecessary anxiety and stress that they can cause for patients. A false positive result can lead patients to undergo further testing and treatment, which can be invasive, uncomfortable, and time-consuming. This can cause unnecessary emotional distress and impact a patient's quality of life.

False positives can also result in unnecessary medical costs. Patients may incur expenses related to additional testing, follow-up visits, and even treatment for cancer that they do not have. The healthcare system may also bear the cost of unnecessary medical procedures and testing, which can strain already limited resources.

Another consequence of false positives is the potential for over diagnosis and over treatment. Over diagnosis occurs when a cancer is detected and treated, but it would not have caused any harm if left untreated. Over treatment can result in unnecessary side effects and complications, such as pain, infection, and reduced quality of life. These issues can also result in additional healthcare costs.

False positives can also impact healthcare providers, as they can lead to increased workload, stress, and the potential for medical malpractice claims. False positives can also undermine the trust between patients and healthcare providers, which can impact patient outcomes.

On the figure above is presented the distribution of cases in mammography's in U.S. which were classified as positive. Theses are the simplest findings so far:

- around 10% of breast mammographies imaging results are returned as positive.
- 97-98 % of the positive cases are not positive in real, and the 87% of patience will be made addtional imaging (perhaps 9% of them will have unnecessary biopsy).
- 2-3% of the positive cases are positive in real (on the figure that perecentage of cases was around 8%, but it was said that the tests were made in different groups of patiences- the goal here is to show the proportions between group cases).

1.5 False negative cases

In comparison with false positive cases the false negative group is less complex and just much smaller. It is around 10 % (FP) vs 0.1 % (FN) of the whole sample. Both presented studies are showing that in general there is only 1 patient in a group of 1000 which was became after the diagnoses by the doctors as false positive what corresponds the 0.1 % of the population. That is a really optimistic information about the cancer detection- the most important goal should be to minimise the risk of FN.

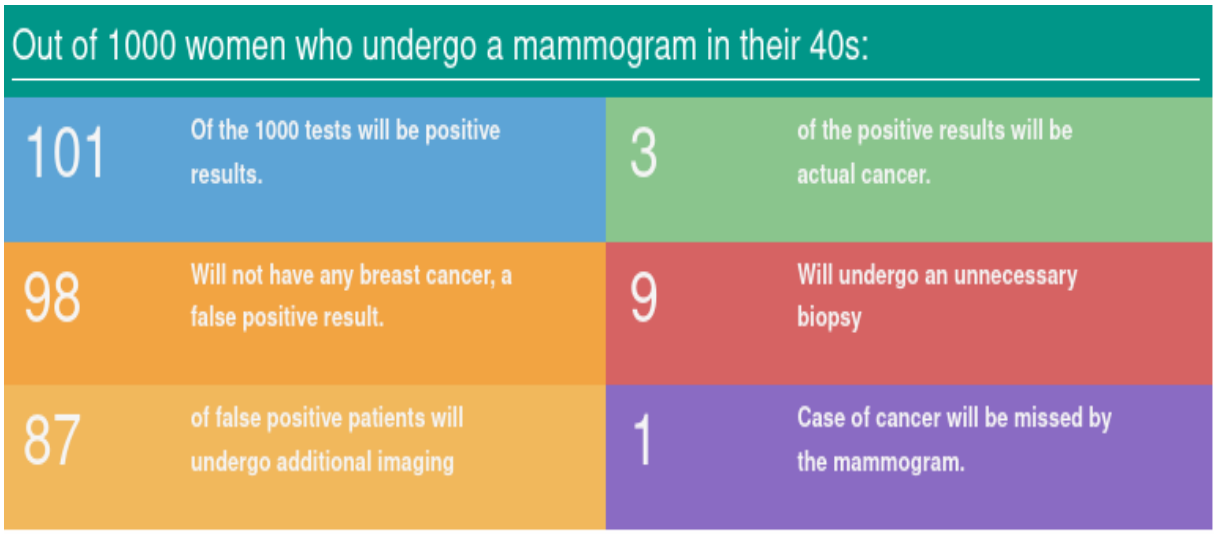

Figure 1.4: The explanation of finding from the U.S. Preventative Services Task Force on Breast Cancer Screening for women in their 40s.

Then the chances of being correctly tested for that groups of patients are related to quick retesting after the FP result.

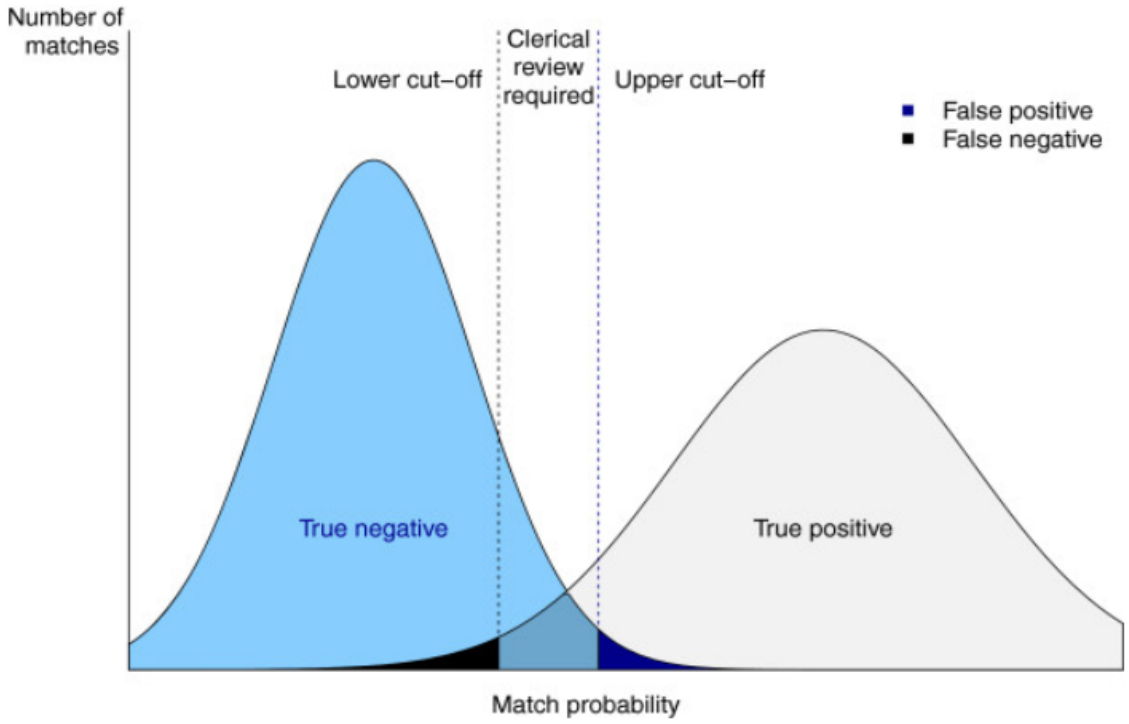

Figure 1.5: Thresholds to minimise false positive and negative matched records. The clerical review required range refers to the area of cases beeing correctly classified. [5]

As is it shown on the plot above during the classification phase the goal is to find the best fitted lowest and upper cut-off as the proper tresholds. In case of binary classification the range of value is the same as in the propability scale. As it is needed to classify some image which has representation in the final propability range of values the key is to find the best number as the treshold between two of the classes. In practise the main idea is to find a kind of settlement between minimising FP and FN and decide which are for the case study errors may be more important and can have more significant impact.

Chapter 2

General approach to data in case of cancer detection

In this section the attention will be given specificaly into data regarding to the Deep Learning solutions. It will be discussed than the most important aspects in the field of handling with data.

2.1 Unbalanced data

Let's mention some of the problems that one might encounter without talking much about data preprocessing.

The general dataset regarding to the cancer detection might be skewed, as it was already shown there might be 1% of the positive class, and 99% of the negative class. As a performance criterion, accuracy will not work and it is more reasonable to maximize the true positive rate (recall) in a grid search (as a function of one or more hyperparameters). In addition, there might be some outliers present which have to be removed first, for example by calculating standard scores and removing extreme values.

Unbalanced image data refers to a dataset where the distribution of samples across different classes is not equal. In the context of breast cancer detection, unbalanced image data means that there are fewer examples of images with malignant (cancerous) cells than images with benign (non-cancerous) cells. This can pose a significant problem for machine learning models, as they may be biased towards the majority class (benign) and perform poorly on the minority class (malignant).

Imbalanced data frequently occurs in real-world problems (specially when the goal is to predict if something is seek or not), so it's a scenario that data scientists often have to deal with.

Dealing with unbalanced image data in breast cancer detection requires special attention and techniques such as data augmentation, oversampling, undersampling, or a combination of these. These techniques aim to increase the representation of the minority class in the dataset, which can improve the performance of the machine learning models. It is important to note that choosing the right technique for dealing with unbalanced image data is problem-specific, and a careful evaluation is needed to ensure that the model's

performance is not negatively impacted.

Imbalanced data can cause issues in understanding the performance of a model. When evaluating performance on imbalanced data, models that only predict well for the majority class will seem to be highly performed when looking at simple metrics such as accuracy, whilst in actuality the model is performing poorly.

This means that metric choice becomes even more important in these kind of situations.

2.2 The proper metrics

On the example of 100 mammography images, where only one of the patients has a real cancer shown on the image it can be easily illustrated the problem of unbalanced data and proper choice of used metrics.
If this single observation of cancer would be incorrectly classified as the model only predicts for the majority negative class. This is clearly not a well-performing model, so we should expect our metrics to reflect this bad performance.

Here it can be shown that by only predicting for the majority class, it seems like the model is performing incredibly well when we look at accuracy (99%).

Tasked to detect the cancer it is more important to reduce the false negative cases rather than the false positive. In that case from the medical point of view it would be better if the prediction of classification outcome with even 10 false positive and 0 false negative than 0 false positive and 1 false negative. The effect of such a situation would be just doing further tests on the FP patients and no one would be than left out with a real cancer.
This demonstrates the importance of choosing the right metrics, to truly understand performance.

It is a good practice to track multiple metrics when developing a machine learning model as each highlights different aspects of model performance.

Main indicators of model fit:

- **Balanced Accuracy**:

$$B_{Acc} = \frac{\text{sensitivity} + \text{specificity}}{2} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{2.1}$$

  Balanced accuracy works well with imbalanced datasets, while *Accuracy* performs poorly in these situations, often leading to misleading results. It takes into account the model's recall ability across all classes, while accuracy does not and is much more simplistic.

- **F1 score**:

$$F1 = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = 2\frac{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \tag{2.2}$$

  $F1$ score is still able to relay true model performance when the dataset is imbalanced, which is one of the reasons it is such a common metric to use. F1 is able to do this because it is calculated as the harmonic mean of both precision and recall for the minority positive class.

- **Micro F1 score**:

$$MicF1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{2.3}$$

  Micro F1 score is the normal F1 formula but calculated using the total number of True Positives (TP), False Positives (FP), and False Negatives (FN), instead of individually for each class.

- **Macro F1 score**:

$$MacF1 = \frac{\text{sum(F1)}}{2} \tag{2.4}$$

  Macro F1 score is the unweighted mean of the F1 scores calculated per class. It is the simplest aggregation for F1 score.

- **ROC-AUC**

  Area under Curve (AUC) or Receiver operating characteristic (ROC) curve is used to evaluate and compare the performance of binary classification model. It measures discrimination power of the predictive classification model. In simple words, it checks how well the model is able to distinguish (separate) events and non-events. A value of 0.5 indicates that the model performs no better than random guessing, while a value of 1 indicates perfect discrimination between positive and negative classes.

- **AUPRC**

  The AUPRC metric evaluates the precision-recall trade-off of a binary classification model. The precision-recall curve is a graph of the precision (TP / TP + FP) against the recall (TP / TP + FN) at various classification thresholds. The area under the precision-recall curve (AUPRC) is a measure of the overall performance of the model, with a value ranging from 0 to 1. A value of 0 indicates that the model has no predictive power, while a value of 1 indicates perfect precision and recall. AUPRC is particularly useful when dealing with imbalanced datasets, where the number of positive and negative samples is heavily skewed.

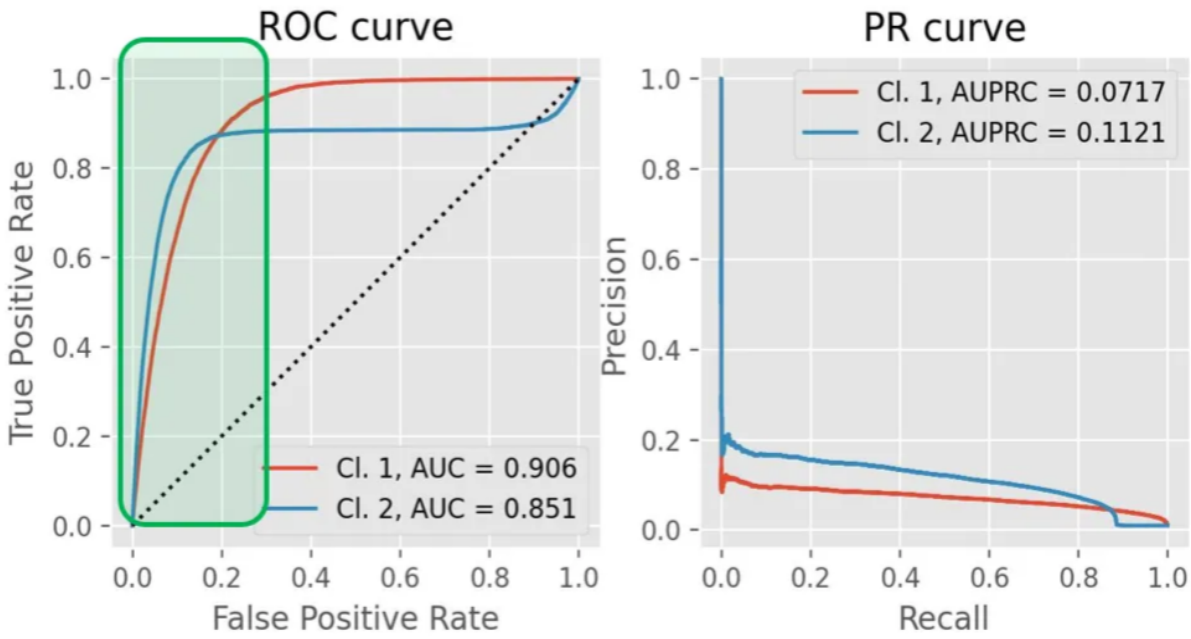

Figure 2.1: Demystifying ROC and precision-recall curves.

In the ROC plot, both curves attain relatively quickly a high true positive rate while having a low false positive rate. Likely, the interest is in the area with a small false positive rate, here below 0.2, which is highlighted with a green rectangle. The reason why only this area is because when the further increasing of the decision threshold takes place, the true positive rate will increase only marginally whereas the false positive rate will increase considerably. This is a consequence of the fact there is class imbalance and that the data is easily classifiable.

For imbalanced data, the PR curve and the AUPRC automatically tend to focus more on areas with small false positive rates. That is why according to the PR curve and the AUPRC in the example above the classifier 2 is a better choice.

Chapter 3

Set of mammography images

3.1 Data acquisition- the RSNA competition

Radiological Society of North America (RSNA) is a non-profit organization that represents 31 radiologic subspecialties from 145 countries around the world. RSNA promotes excellence in patient care and health care delivery through education, research, and technological innovation.

On the November of 2022 they launched a competition to identify breast cancer [22]. The task was to train a model based on deep learning with screening mammograms obtained from regular screening. The potential efforts in this competition could help extend the benefits of early detection to a broader population. Greater access could further reduce breast cancer mortality worldwide.

The image data used in the thesis where collected from *kaggle* platform where RSNA has published them for competition and research purposes.
Data contain 314.72 Gb of files. The main core are the mammography images. There is also numeric data frame added.

The mammograms immages are in dicom format. There are usually but not always 4 images per patient.
Metadata (the numeric data frame) consits information for each patient and image.

The metadata's variables:

- site_id - Id code for the source hospital.
- patient_id - Id code for the patient.
- image_id - Id code for the image.
- laterality - Whether the image is of the left or right breast.
- view - The orientation of the image. The default for a screening exam is to capture two views per breast.
- age - The patient's age in years.
- implant - Whether or not the patient had breast implants. Site 1 only provides breast implant information at the patient level, not at the breast level.
- density - A rating for how dense the breast tissue is, with A being the least dense and D being the most dense. Extremely dense tissue can make diagnosis more difficult.
- machine_id - An id code for the imaging device.
- cancer - Whether or not the breast was positive for malignant cancer. The target value.
- biopsy - Whether or not a follow-up biopsy was performed on the breast.
- invasive - If the breast is positive for cancer, whether or not the cancer proved to be invasive.
- BIRADS - 0 if the breast required follow-up, 1 if the breast was rated as negative for cancer, and 2 if the breast was rated as normal. Only provided for train.
- prediction_id - The id for the matching submission row. Multiple images will share the same prediction ID.
- difficult_negative_case - True if the case was unusually difficult.

3.2 Preview and initial analysis

In this section will be presented an overview of the data, observations after the firsts data analyses.
Generally, mammography is conducted with 2 views (from 2 angles), because the position of breast cancer is various, although the upper outer quadrant of the breast is the most common site of breast cancer occurrence. But the shape of the breast is almost the same regardless of different views. Moreover, the left and right breast generally have the same view.
So if there is already known with what kind of data were established it is time to set up some assumptions and approaches about the ways to use data for the modeling process.

General informations and observation about the given dataset about patients:

- The number of total patients - 54706.
- There are 2 total hospitals from where the records were gathered, split roughly 50-50.
- Each patient has an average of 4.5 breast scans (with 4 being the least number of scans and 14 being the maximum number of scans per patient).
- There are slightly more images for the right breast (27,439) than for left (27,267).
- The number of patients having implant - 1477.
- The orientation of images is occur in 6 views. MLO (mediolateral oblique) - 27903 images. CC (craniocaudal) - 26765 images. AT (axillary tail) - 19 images. LM (latero-medial) - 10 images. ML (medio-lateral) - 8 images. LMO (latero-medial oblique) - 1 images.
- The number of patients having malignant cancer is 1158 including these whose malignant cancer is invasive - 818.
- The number of patient who took biopsy - 2969.
- The average age is 58 years old, the range is from 26 to 89.
- The class imbalance is very big, with only 2% of images being cancer positive.
- For cases when cancer was present, another flag that states if the cancer was invasive or not. Around 70% of cases that had cancer were invasive.
- There have been around 7705 cases (around 16%) that were exceptionally difficult in the train dataset.

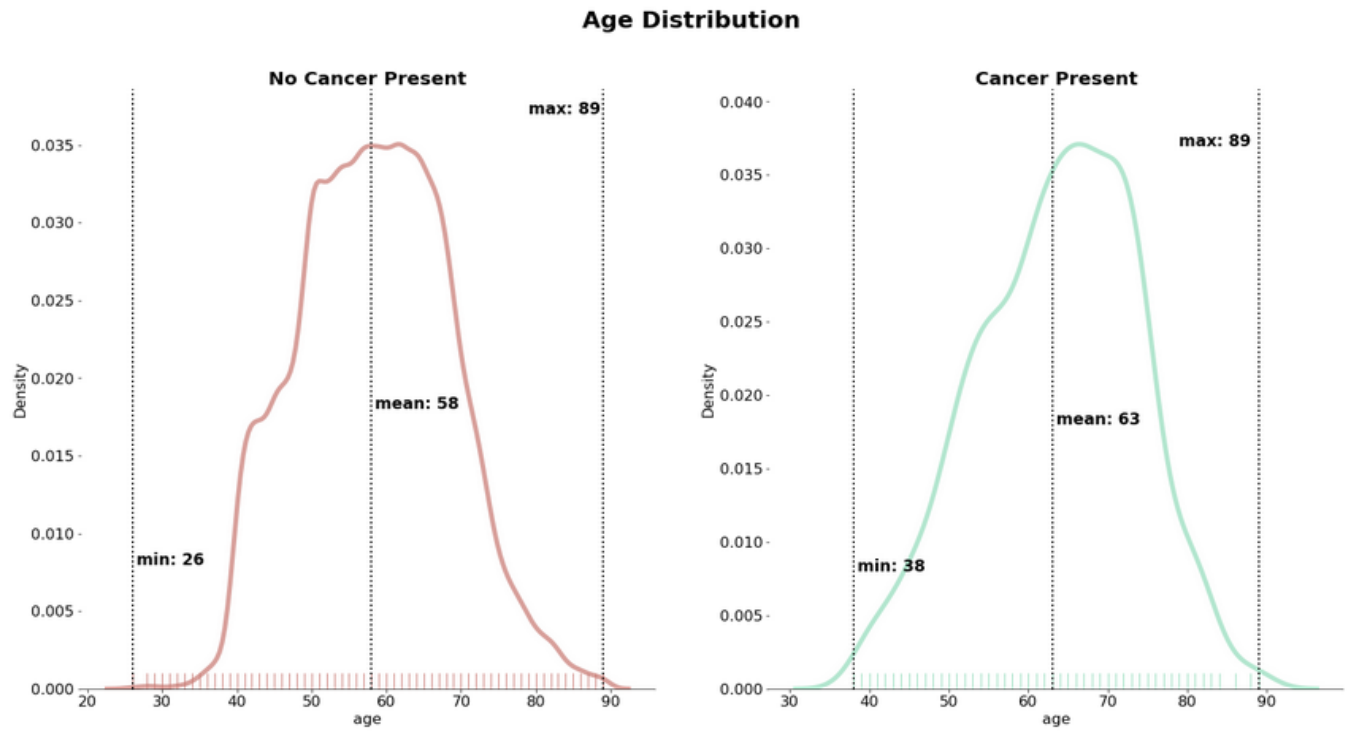

Figure 3.1: Distributions of age variables in groups of patients with or without cancer flag.

Both of the distributions resemble a normal distribution. The group of patients with cancer is more left-skewed, which confirm the intuitive that there exists a correlation between the age and cancer appearing. The average age is 58 years old, with the vast majority of the patients having between 50 and 65 years old. There are a few outliers with very young patients (26-30 years old), as well as a few more senior patients (89 years old).

It looks like the minimum and average age have shifted. The youngest patient to have cancer is 38 years old, while the mean of those patients is 63 years old.

3.3 Image data and the numeric data in Deep Learning

Image preprocessing is a crucial step in deep learning for cancer breast detection in mammography images. The goal of preprocessing is to enhance the quality of the images and to remove any artifacts or noise that may interfere with the analysis.

Overall it can help to improve the quality of the images and can make the deep learning model more robust and accurate.

After that section both of the types of data should be well processed and be ready to use in a proper unchangeable form as an input to combine them together than in some of the steps during neuron networks modelling.

3.3.1 Mammography images processing

In mammography imaging, different colors can be used to represent different levels of tissue density or different types of tissue.

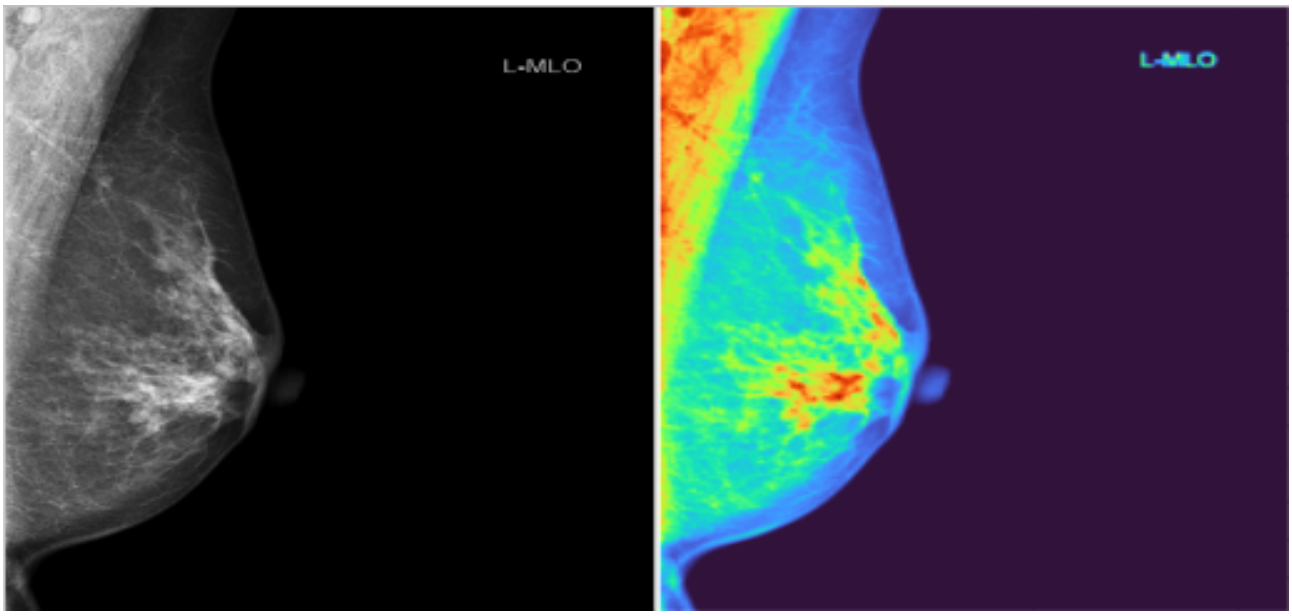

Figure 3.2: The examples of imaging of a case marked as difficult and wihout cancer detected in left MLO view of mammogram in 2 versions with 1 channel and 3 channels of colours.

For comparison it is presented the method of using colours in RGB for better density differences in the breasts.

In digital mammography, after the image is acquired, it is processed to enhance its contrast and sharpness, and to remove noise. During this process, the image is often color-coded to make it easier for radiologists to interpret.

Typically, black represents areas of low density, such as air or fat, while white represents areas of high density, such as bone or metal. Shades of gray are used to represent the different levels of tissue density between black and white.

In addition to grayscale, some mammography systems use color coding to highlight specific features. For example, some systems use green to represent fatty tissue, blue to represent blood vessels, and red to represent areas of increased blood flow.

It's important to note that different mammography systems and software may use different color coding schemes, so radiologists need to be trained to interpret the images based on the specific system being used. Ultimately, the goal of color coding is to help radiologists identify and diagnose abnormalities in the breast tissue.

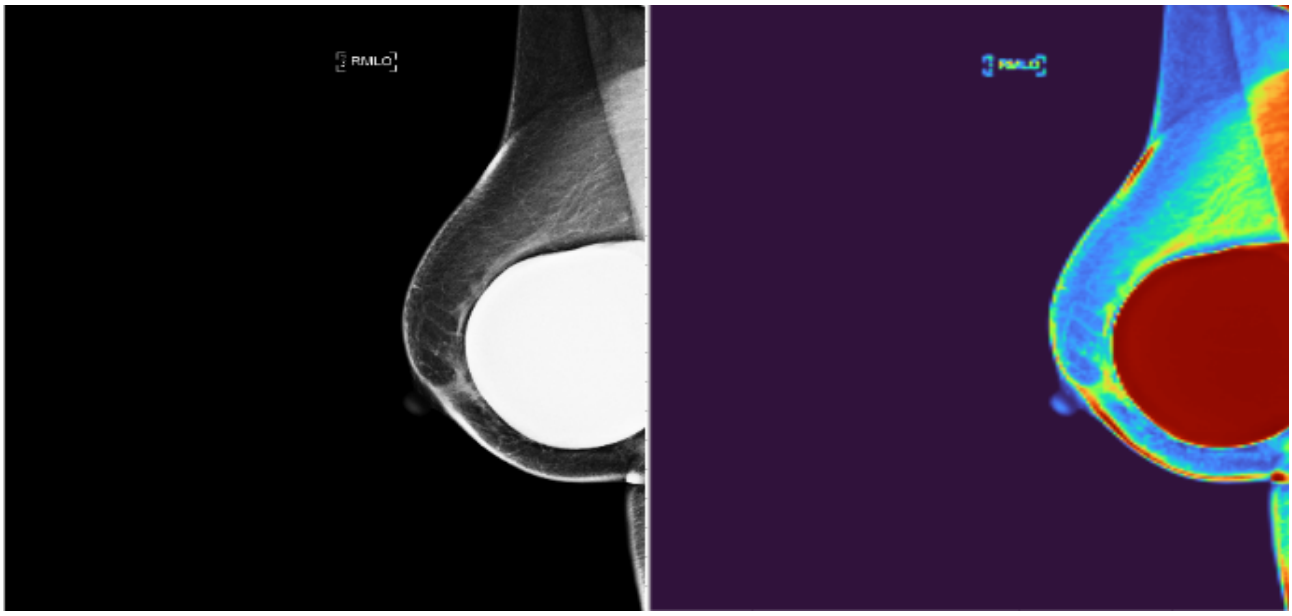

Figure 3.3: The examples of imaging of a case with existing implant and detected cancer in right MLO view of mammogram in 2 versions with 1 channel and 3 channels of colours.

Data augmentation techniques, such as rotation, flipping, or zooming, can be used to create additional training samples from the original mammography images [8]. This can help to increase the size of the training dataset and can improve the generalization performance of the deep learning model.
It helps deep learning models generalise better by increasing the size of the dataset and exposing the model to additional variations of the same images. This can improve the model's accuracy and reduce overfitting.

Following procedures has taken place on the original images of the dataset.

- The images (.dcm format) were processed as input of shape $[3, 3, 224, 224]$, what means that the data were used with 3 channels in RGB colours. The batch size required more iterations to converge but resulted in better generalization and allow for more precise gradient updates. That is why it was set as 3.

- Cropping a random portion of image and resizing it to the given size -($height = 224, width = 224$). A crop of the original image is made: the crop has a random area (H * W) and a random aspect ratio. This crop is finally resized to the given size.

- Increase the variety of points of view on an object in the training set. This approach creates the needed diversity without the need to find and label more data.
  Shift factor - specifies a specific range from which a random shift factor is picked and used to shift an image both horizontally and vertically ($shift\_limit = 0.0625$).
  Scale factor - specifies a specific range from which a random scale factor is picked and used to rescale an image ($scale\_limit = [0.8, 1.2]$).
  Rotation range - specifies a specific range from which a random angle (in degrees) is picked and used to rotate an image ($rotate\_limit = 90$).

- Horizontal and vertical flips were used to increase the size of a dataset by flipping images horizontally pr vertically. This can help improve the accuracy of a model by exposing it to additional variations of the same images.

The following data augmentations procedures, one of the most important goals in the breast cancer detection was to standardise all of the images which were collected from different machine devises. Another reason of using this kind of data treatments was to decrease the level of deviation and differences between the ranges of images. That factor was related to the way how the breast is placed in the right place on the imaging machine.

3.3.2 Metadata processing

Because the core of the data are images in case of metadata processing the goal was to gather the most relevant and useful information which would be good for modelling, combining with images.

The following decision have been made:

- Standard approach regarding to the direct cancer variables, they will be removed also in the training process.
- Removing the records which are type of outliers in *view* variable (all instead of MLO and CC have appear less than 30 times in the whole population).
- Binary encoding the *view*, *letarity* and *implant* variables.
- Detecting records with not available values in *dense* and *age* variables and filling them using the mean values and then normalizing, so in that case:
- Encoding *dense* variable for 4 categories.

The final metadata variables used to modelling are presented below.

Table 3.1: Table of unique values of final variables in the input metadata

View	Leterility	Age	Density	Implant
MLO (mediolateral oblique)	0 (left)	26	A (the lest dense)	0 (does not contain implant)
CC (craniocaudal)	1 (right)	...	B	1 (contains implant)
-	-	...	C	-
-	-	89	D	-

Chapter 4

Look into Deep Learning. The overview of the neural networks used in analyzes

This chapter aims to enlighten readers on the complex but intriguing theory of neural networks, the backbone of modern AI systems. Neural networks, inspired by the workings of the human brain, have been instrumental in reshaping numerous sectors of society, and one such notable field is healthcare.

It will be delved into the intricate architecture of neural networks focusing on convolutional neural networks (CNNs). Exploration the anatomy of these networks will take place, explaining nodes, layers, and connections, and how they work in tandem to make sense of the complex data they are fed. Next,it is gonna be shown how these neural networks 'learn' from data, covering essential aspects.

4.1 An introduction to the issues covered in the section

In this section will be defined and explained the definition of the issues on which based are the models presented in the chapter.

ImageNet is a large dataset that was designed for use in visual object recognition software research. It contains over 14 million labeled images that cover more than 20,000 categories of objects. This immense and varied collection of data has been a key driver in the advancement of deep learning, specifically in the realm of Convolutional Neural Networks (CNNs) [16].

CNNs are a specialized kind of neural network designed to process data with a grid-like topology, such as an image. They excel in identifying patterns within images because they take into account the spatial nature of the data, maintaining the relative positional information. Each layer of a CNN applies different filters, identifying various features at increasing levels of complexity.

Now, how does ImageNet fit into this, and more specifically, how does it relate to breast cancer classification? While ImageNet itself does not contain medical images like mammograms, it plays a vital role in the process. The technique used here is known as transfer learning.

Transfer learning allows us to use pre-trained models (such as those trained on ImageNet) and fine-tune them for a specific task, like breast cancer detection in this case. Even though ImageNet images are not related to mammograms, the models trained on them learn a rich set of features that can be useful for many visual recognition tasks. For instance, a model trained on ImageNet can recognize textures, shapes, and patterns - key elements that are also relevant in analyzing mammographic images.

So, when working on breast cancer classification, we can take a CNN model trained on ImageNet, replace the last few layers, and then train this modified network on a dataset of mammograms. The idea is that the early layers of the model, which have been trained on ImageNet, can effectively act as feature detectors for the mammogram images. The final layers that we've added can then learn to classify these features in the context of breast cancer.

Using pre-trained models from ImageNet has demonstrated impressive results in medical image analysis, including breast cancer classification. It allows researchers and clinicians to leverage the power of deep learning without requiring a prohibitively large labeled dataset, which is often a challenge in medical settings. In this way, ImageNet and CNNs together are helping push the boundaries of automated breast cancer detection.

Ensemble learning involves combining multiple models to improve the accuracy and stability of the classification. This approach has been shown to be effective in breast cancer classification by combining different neural network architectures, connectin the image data with metadata.

4.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are widely used in medical image analysis, including breast cancer classification. They can learn spatial features from mammography images, which can be useful in distinguishing between benign and malignant tumors. CNNs are a class of Deep Neural Networks that can recognize and classify particular features from images and are widely used for analyzing visual images. Various modifications of CNNs, such as ResNet, InceptionNet, and DenseNet, have shown promising results in breast cancer classification [11].

The term **'Convolution" in CNN denotes the mathematical function of convolution which is a special kind of linear operation wherein two functions are multiplied to produce a third function which expresses how the shape of one function is modified by the other.** [13] In simple terms, two images which can be represented as matrices are multiplied to give an output that is used to extract features from the image.

The choice of neural network architecture will depend on the specific requirements

of the breast cancer classification task. It may be necessary to experiment with different models and architectures to find the most efficient and accurate solution, which will be presented in the next chapter.

Convolutional Neural Networks (CNNs) are considered the best choice for image classification in the case of breast cancer detection for several reasons:

- CNNs are designed to handle the high-dimensional input data (i.e., images) and can automatically learn relevant features from them. They are particularly well-suited for image classification tasks because they can effectively capture local patterns and structures in the image, which is important in detecting subtle details that may indicate the presence of cancer.
- CNNs can handle images of different sizes and resolutions, making them useful for analyzing medical images, which can vary in quality and size depending on the equipment used to generate them.
- CNNs can be trained on large datasets and can effectively generalize to new, unseen data, which is important for detecting cancer accurately and reliably.
- CNNs can be fine-tuned for specific tasks such as breast cancer detection, by modifying the architecture and the training process to optimize performance on this specific problem.

Overall, CNNs are the best choice for image classification tasks such as breast cancer detection because they can learn relevant features from high-dimensional input data, can handle images of different sizes and resolutions, and can be trained and optimized for specific tasks.

Breast cancer classification is a challenging task, and the choice of neural network architecture will depend on several factors, such as the size and complexity of the dataset, the desired accuracy and efficiency of the model, and the availability of computing resources.

4.2.1 General CNN architecture

The main idea behind a CNN architecture is to apply a series of convolutional, activation, and pooling layers to the input image, followed by one or more fully connected layers to produce the final classification output. Here is an explanation of the key elements in this architecture.

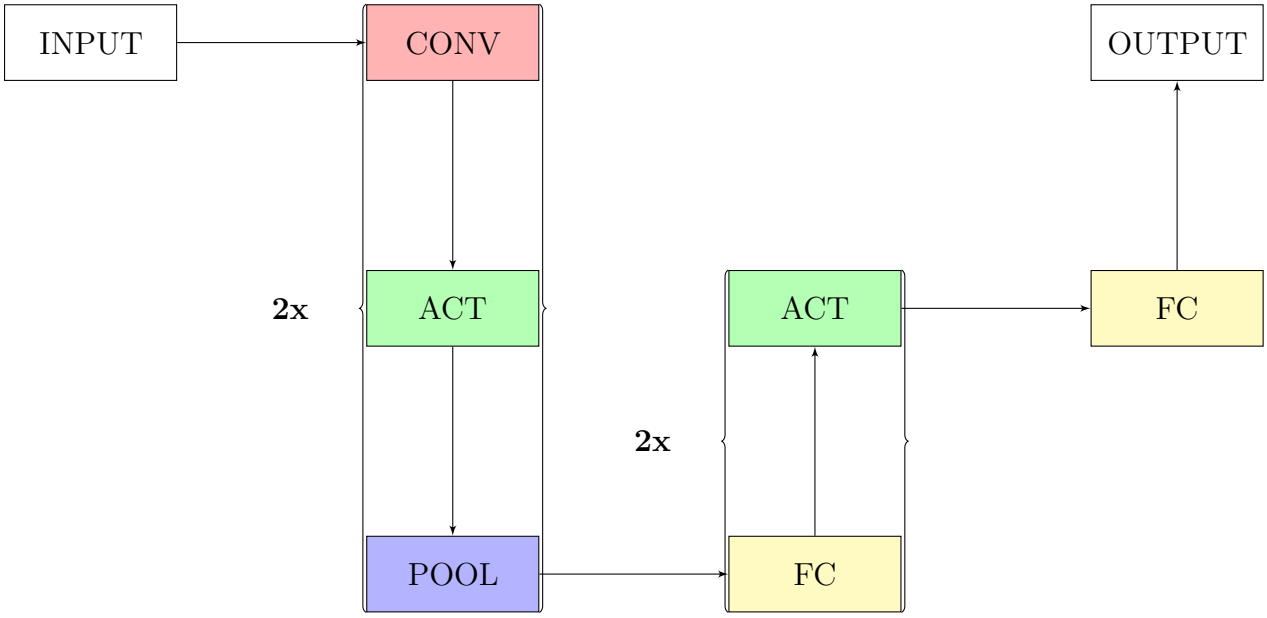

Figure 4.1: The schema of structure of the CNN

- INPUT: The initial input image, which is typically preprocessed to a fixed size and format, such as a specific number of pixels and color channels.
- CONV: This layer applies a set of learnable filters to the input image, each of which is convolved over the entire image to produce a set of output feature maps. These filters help identify local patterns and structures in the image. Its purpose is to detect features such as edges, corners, and other intricate details from the input data. In the context of a CNN used for breast cancer detection, the convolutional layer is of crucial importance. It could be trained to identify various visual features in mammographic images that may be indicative of cancer. Lower-level layers might learn to detect simple patterns in cell structures, while more complex patterns that might be indicative of a tumor could be detected by higher layers. By stacking multiple convolutional layers, the network can learn more abstract and complex visual features to perform accurate classification.
- ACT: It applies an element-wise activation function called Rectified Linear Unit (ReLU) to the output of the convolutional layer, which introduces nonlinearity and helps to further increase the model's ability to identify important features in the image [25].
- POOL: It downsamples the output feature maps of the convolutional layer by taking the maximum or average value in each local pooling region. This helps to reduce the spatial dimensions of the output feature maps and introduces some degree of invariance to small variations in the input image. A pooling layer is another essential component in a Convolutional Neural Network. After each convolutional layer, a pooling layer often follows. Its primary function is to reduce the spatial dimensions (width and height) of the input volume while keeping the depth intact. This operation reduces computational complexity, controls overfitting, and abstracts the features. Regarding breast cancer detection, pooling layers can help to create invariant representations of the images. For instance, max pooling might enable the

network to recognize important patterns, like the shape of a cell nucleus, regardless of their size or position in the image. This invariance to local translation can make the CNN more robust and accurate in its predictions.

- The repeated sequence of convolutional, activation, and pooling layers that is commonly used in many CNN architectures.

$$[CONV- > ACT- > POOL] * 2$$

  By stacking multiple such layers, the model can learn increasingly abstract and high-level features from the input image.

- FC: This layer is a fully connected layer, which takes the flattened output of the previous layers and applies a set of learnable weights to produce a set of output activations. These weights help combine the learned features from the previous layers to make a final classification decision.

- OUTPUT: The final output layer, which produces a set of class probabilities or scores based on the activations of the previous layer. The predicted class is typically the one with the highest probability score.

Additionaly after the fundamental elemets of the network there occur also other types of layers.

The **Dropout** is a regularization technique used in neural networks, including convolutional neural networks, to prevent overfitting [12]. Overfitting occurs when the model learns the training data too well and captures noise along with the underlying pattern. As a result, it performs poorly when presented with new, unseen data because it fails to generalize from the training data.

The dropout layer plays a critical role in addressing this issue. During the training process, a dropout layer randomly sets a fraction of input units to 0 at each update, effectively "dropping out"those units from the network for that specific update. The fraction of input units to drop is a parameter that can be set and is often set to a value like 0.5. This means that approximately half of the input units are randomly dropped during the training phase.

The key effect of this is that it makes the network less confident about the exact configuration of features it has learned from the input data, forcing it to become more robust. The model is compelled to learn more generalized features in the data rather than relying on specific, potentially noisy, details of the training data.

In the case of CNNs, a dropout layer is usually applied before the fully connected layers, after the convolutional and pooling layers. The CNN learns to extract useful features from the images through its convolutional layers, and the dropout layer helps ensure these features are robust and generalizable.

In the specific case of breast cancer detection, dropout layers can help ensure the model is not overly reliant on specific features or configurations in the mammogram data that might not generalize well to all cases of breast cancer. By adding dropout layers, we improve the model's ability to identify the fundamental, generalizable patterns that truly signal the presence of cancerous tissues.

A **flatten layer** is used in a CNN as a bridge between convolutional and pooling layers, and the fully connected layers. The primary function of a flatten layer is to convert the multi-dimensional input tensors into a one-dimensional vector, which can then be fed into the subsequent fully connected layers. In a typical CNN architecture, a flatten layer is usually positioned after all the convolutional and pooling layers and before the fully connected layers. In the context of breast cancer detection using image data, the flatten layer would be responsible for reshaping the output tensors from the previous layers, preparing them for the classification process in the dense layers that follow. The dense layers then make the final prediction, such as whether a tumor is benign or malignant, based on the input they receive from the flatten layer.

4.2.2 The hyperparameters for the convolutional layers

Kernel size

The kernel, also referred to as a filter, is a small matrix that traverses through the input data, performing a convolution operation at each step.

The kernel size is the dimensionality (width and height) of this kernel. Common kernel sizes include 1x1, 3x3, 5x5, and 7x7. A larger kernel will cover a larger area of the input when performing the convolution, which might help to capture larger scale features. However, larger kernels will also result in a larger number of parameters and thus increase the computational complexity.

Smaller kernels, on the other hand, cover less input area but can capture finer-grained details. It's worth noting that the recent trend in deep learning is to use 3x3 kernels, which seem to offer a good trade-off between computational efficiency and the ability to capture complex features. Some networks also employ multiple parallel branches of different kernel sizes (as in the Inception architecture) to simultaneously capture features at various scales.

In the context of breast cancer detection, the choice of kernel size might depend on the scale of features that are relevant for the task. For example, if finer details in the cellular structure are crucial to determine malignancy, smaller kernel sizes may be preferable. Conversely, if larger patterns or abnormalities need to be detected, larger kernel sizes may be more appropriate.

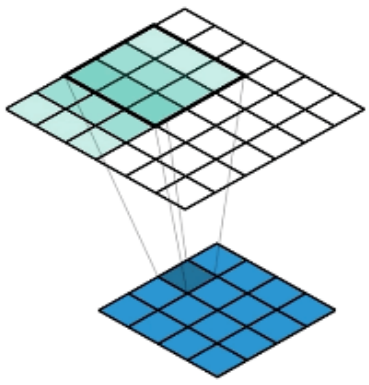

Figure 4.2: The size of the filter (kernel) refers to the dimensions of the sliding window above the input image. In this case it has size 3x3. Graphic source: [14]

Strides

The stride is a hyperparameter which determines how much the kernel or filter moves across the input image or feature map.

In other words, stride refers to the number of pixels shifts over the input matrix when the convolution operation is performed. For example, when the stride is 1, we move the filters one pixel at a time. When the stride is 2, we move the filters two pixels at a time, and so on. This implies that the filters overlap with each other when the stride is less than the kernel size.

The primary effect of using a larger stride is that the dimensionality of the output feature maps decreases. For instance, when a stride of 1 is used, the output feature map will have the same width and height as the input (assuming padding is used). However, with a stride of 2, the width and height of the output feature map will be roughly half that of the input.

Using larger strides can help reduce computational requirements and control the amount of information flowing through the network, but at the expense of losing some detailed information. Smaller strides, on the other hand, tend to preserve more detail and produce denser representations, but at the cost of increased computational complexity.

In the context of CNNs used for breast cancer detection, stride could play a significant role. A small stride may allow the model to capture fine-grained details in the tissue scans, which might be crucial for detecting early-stage tumors. On the other hand, a larger stride might be sufficient for applications where only larger-scale, more obvious features are relevant for the diagnosis.

Padding

Padding refers to the process of adding extra artificial pixels around the border of the input image or feature map.

There are two primary types of padding:

- Valid Padding (No Padding): In this case, no padding is added to the input. The filter is only applied to the valid locations on the input volume, leading to reduced dimensions in the output volume.
- Same Padding: Here, padding is added in such a way that the output dimensions are the same as the input dimensions. This is done by adding $(kernelsize - 1)/2$ number of pixels on each side of the input volume. The pixels added are typically zeroes.

The reason padding is important can be broken down into two main factors:

- Control of Output Dimensions: Padding allows us to control the spatial size of the output volumes, often preserving the spatial dimensions of the input volume. This can be particularly useful when we want to design deeper networks, as without padding, the size of the volumes would reduce at a rate that we might lose too much information too quickly.
- Preservation of Information at the Borders: Without padding, the pixels on the corners and the edges of the input volume are taken into account by a much smaller number of filters than the central pixels. Padding ensures that these pixels are processed by an equivalent number of filters, preserving information at the borders.

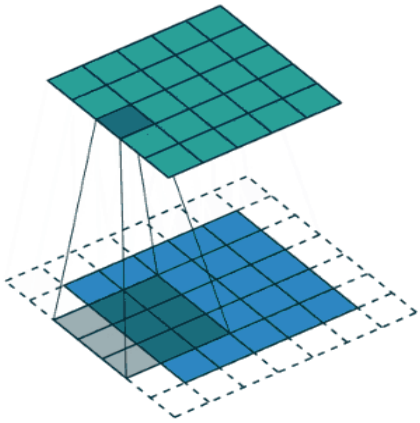

Figure 4.3: As it is shown, the output size (green square) is 5x5 with the added "frame". Same as input image (blue square). If there was no frame added, then the output size would be 3×3. Graphic source: [14]

4.3 VGG19

VGG19 is a variant of the VGG (Visual Geometry Group) model. It's a convolutional neural network (CNN) model, widely used in the field of computer vision. The "19"in VGG19 refers to the number of layers with trainable weights in the network, which includes 16 convolutional layers and 3 fully connected layers. VGG19 has 19.6 billion FLOPs.

The VGG19 model is a good choice for breast cancer classification for several reasons:

- **Feature Extraction**: The convolutional layers in VGG19 are excellent at extracting features from images. This is particularly useful in medical imaging, where subtle features in the images can indicate the presence or absence of cancer. The multiple layers of the network allow it to learn complex patterns and features in the data.
- **Transfer Learning**: VGG19, pre-trained on the ImageNet dataset, can be fine-tuned on a smaller dataset for breast cancer classification. This is beneficial because medical imaging datasets are often smaller due to privacy concerns and the difficulty of obtaining labeled data. Transfer learning allows the model to leverage the knowledge gained from the large ImageNet dataset, which can lead to better performance on the smaller breast cancer dataset.
- **Depth of the Network**: The depth of the VGG19 model allows it to learn more complex features. This can be particularly useful in medical imaging, where the difference between benign and malignant tumors can be subtle and complex.
- **Localization**: The convolutional nature of VGG19 allows it to maintain spatial information about the images. This can be useful in identifying the location of tumors in the breast tissue.
- **Simplicity**: Despite its depth, VGG19 has a very uniform architecture. It uses 3x3 convolutional layers stacked on top of each other in increasing depth. This simplicity makes the model easier to understand and modify if necessary.

For instance, a study by Antropova, Huynh, and Giger (2017) used transfer learning with a VGG19 model for breast cancer risk assessment. They found that the model, pre-trained on ImageNet and fine-tuned on a mammography dataset, outperformed traditional machine learning methods [3].

Another study by Wang et al. (2019) used a VGG19 model for breast cancer classification based on histopathological images. They found that the model achieved high accuracy, demonstrating the effectiveness of deep learning methods for this task [23].

4.3.1 Network architecture

In terms of CNN aspects, VGG19 uses several key components.
The **input Layer** takes in this case an input image of a fixed size (224 x 224 RGB image).

There are 16 **convolutional layers** in VGG19. These layers use small 3x3 receptive fields and are stacked on top of each other. The number of filters in these layers starts from 64 in the first layer and increases by a factor of 2 after every max-pooling layer, going up to 512.

The network contains 5 **max-pooling layers**. Each of them is placed after the convolutional layers blocks. They perform a 2x2 max pooling operation with stride of 2 value. This was followed by ReLu to introduce non-linearity to make the model classify better and to improve computational time.

Next in the architecture are 3 **fully connected layers**. The first two have 4096 channels each, and the third one performs with 1000 channels. The reason of that is that it was developed and tested as part of ImageNet Large Scale Visual Recognition Challenge competition, and its performance ILSVRC classification task involves classifying the objects in the images into one of 1,000 categories.

The **softmax Layer** is the final layer in the network. It uses the softmax function to output a probability distribution over the classes.

The **ReLU activation function** is applied to the output of every convolutional and fully connected layer.

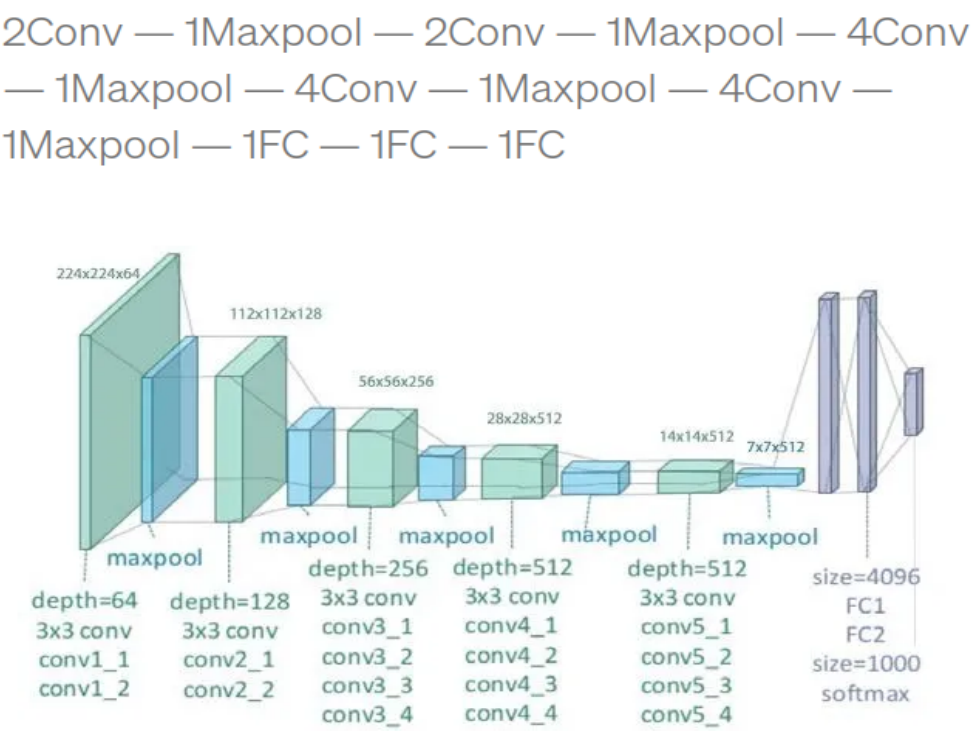

Figure 4.4: VGG-19 Architecture (Image source: researchgate.net) [26]

Here is a additional breakdown of its architecture:

- The only kind of preprocessing phase that is done in the network is that it subtracts the mean RGB value from each pixel, computed over the whole training set.
- Used kernels of (3x3) size with a stride size of 1 pixel, this enabled them to cover the whole notion of the image.
- Spatial padding is used to preserve the spatial resolution of the image.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 4.5: Comparison between the Convet Configurations in the architecture. The ReLu layers are not shown for the sake of brevity [24].

The column E in the table above is for VGG19 (other columns are for other variants of VGG models). The difference is only in the numbers of convolutional layers in the last 3 blocks. Other elements are the same. It influences to the number of final parameters to be found.

In conclusion, while VGG19 has been a strong choice for breast cancer classification, transitioning to ResNet50 could potentially lead to better performance due to its depth, efficiency, and the use of residual learning. However, it's important to note that the choice of model should be based on the specific requirements of the task, including the complexity of the images, the size of the dataset, and the computational resources available. Therefore, it's always a good idea to test different models and choose the one that best fits the task at hand.

4.4 ResNet50

In a Convolutional Neural Network, as the number of layers increase, so does the ability of the model to fit more complex functions. However, it's not always the case that more layers lead to better performance. If a model becomes too complex, it can start to overfit, learning the noise in the training data rather than the underlying pattern. Overfit models tend to perform poorly on unseen data.
More layers in a CNN require more computational resources for both training and inference. It also increases the number of parameters in the model, leading to longer training times

and requiring more memory.
As the network gets deeper, it becomes more prone to the vanishing and exploding gradient problem, which makes the network harder to train. This issue has been partly addressed by methods such as batch normalization, residual connections (ResNets), and better weight initialization strategies.

The ResNet50 model, a deep residual network, is a popular choice for the classification of mammography images for breast cancer detection, for several reasons:

- Residual Learning: The key innovation in ResNet50 is the use of residual learning, where 'shortcut' or 'skip connections' are used to bypass layers. This approach helps to mitigate the problem of vanishing gradients, which is a common issue in training deep neural networks. As a result, ResNet50 can learn more complex features and achieve better performance than other models that do not use residual learning.
- Model Complexity and Efficiency: ResNet50 is a relatively deep network (50 layers), it has a manageable number of parameters compared to other deep models. It strikes a good balance between performance and computational efficiency, making it a practical choice for many tasks, including mammography image classification.
- Versatility: ResNet50 is a versatile model that has been used for a wide range of image classification tasks. It can handle a variety of image sizes and types.

In general CNN have a major disadvantage - **Vanishing Gradient Problem**. During **backpropagation**, the value of gradient decreases significantly, it means that hardly any change comes to weights. The needed solution appears in **Skip Connection** and takes place in the ResNet models.

The main idea of the encountered problem - the networks could not go as deep as wanted, because they started to lose generalization capability.

Backpropagation

During the process of back-propagation in a neural network, the chain rule of differentiation is used to calculate gradients and update the weights of the network. As it moves from the deeper layers to the shallower layers during back-propagation, the chain rule requires to multiply the gradients.

The chain rule states that there is a composition of functions, the derivative of the composite function is equal to the product of the derivatives of the individual functions. In the context of a neural network, each layer performs a transformation on the input using its weights and activation function. During back-propagation, the gradient of the loss function with respect to the weights of each layer is being calculating .
By multiplying the gradients as we move from deep to shallow layers, the back-propagation algorithm effectively accumulates the contributions of the deeper layers, allowing the gradients to flow back through the network and update the weights of each layer based on the overall impact of the deeper layers on the loss function.

The weights of a neural network are updated using the backpropagation algorithm. The backpropagation algorithm makes a small change to each weight in such a way that

the loss of the model decreases. It updates each weight such that it takes a step in the direction along which the loss decreases. This direction is nothing but the gradient of this weight (with respect to the loss).

Vanishing Gradient Problem

One of the problems ResNets solve is the famous known vanishing gradient. This is because when the network is too deep, the gradients from where the loss function is calculated easily shrink to zero after several applications of the chain rule. This result on the weights is never updating its values and therefore, no learning is being performed.

The Vanishing Gradient problem is a difficulty found in training artificial neural networks with gradient-based learning methods and backpropagation. In these methods, each of the neural network's weights receives an update proportional to the partial derivative of the error function with respect to the current weight in each iteration of training.

The problem occurs when the gradients of the loss function become very small and approach to zero. As the learning rate is typically a small number, the product of the learning rate and the gradient is even smaller. As a result, the weights and biases of the network are barely updated, and the learning process stalls or progresses very slowly. The issue is particularly pronounced in deep networks, where the gradients of the loss function can decrease exponentially with the depth of the network, resulting in the early layers of the network learning very slowly.
In the simple feedforward case, if we have a loss function L, and a deep network with n layers with h_i representing the hidden state at layer i, the gradient of L with respect to the hidden state at layer i is calculated by repeatedly applying the chain rule:

$$\frac{\partial L}{\partial h_i} = \frac{\partial L}{\partial h_n} \prod_{j=i+1}^{n} (\frac{\partial h_j}{\partial h_{j-1}})$$

The goal of calculating $\frac{\partial L}{\partial h_i}$, where h_i is the output of the i layer in a neural network, is to understand how a small change in the output of the i layer affects the final loss.
In the context of training a neural network, this information is crucial. As we train our model, we want to adjust the weights and biases of the network to minimize the loss function. To know how to adjust these parameters, we need to understand how small changes in their values will affect the overall loss. This understanding is provided by the gradients of the loss with respect to the weights and biases.
However, the weights and biases directly affect the output of each layer (the h_i), not the loss function itself. So, to understand how changes in the weights and biases affect the loss, we first compute $\frac{\partial L}{\partial h_i}$ for each layer.
Each factor $\frac{\partial h_j}{\partial h_{j-1}}$ corresponds to the derivative of an activation function, which in case of *sigmoid* or *tanh* is in the interval (0, 1). Multiplying many of these small numbers together can cause the gradient to diminish rapidly as we go deeper into the network, hence the term vanishing gradient.

Residual Networks help mitigate the vanishing gradient problem by introducing shortcut (or skip) connections. In a ResNet, instead of trying to learn an underlying

mapping $H(x) = y$ from an input x to an output y, we aim to learn a residual function $F(x) = H(x) - x$, where $F(x)$ is the residual. If we rearrange this, we get $H(x) = F(x) + x$, which is the formulation used in ResNets. Here, $F(x)$ corresponds to the stacked non-linear layers, and x is the identity function.

When computing gradients via backpropagation, the gradient of the loss L with respect to a hidden state h_i now becomes:

$$\frac{\partial L}{\partial h_i} = \frac{\partial L}{\partial h_n} \prod_{j=i+1}^{n} (1 + \frac{\partial F_j}{\partial h_{j-1}})$$

The term $(1 + \frac{\partial F_j}{\partial h_{j-1}})$ arises from the derivative of the layer's output with respect to its input $(F(x) + x)$. The 1 comes from the derivative of x with respect to x, and $\frac{\partial F_j}{\partial h_{j-1}}$ is the derivative of the residual function. Even if the latter becomes very small, the overall term will not vanish because of the added 1, thus alleviating the vanishing gradient problem. This approach ensures that even when the network depth increases, layers can still learn identity functions. Hence, they can theoretically learn equally well as their shallower counterparts. By allowing gradients to propagate directly through the skip connections, ResNets enable the successful training of deep networks.

Skip connection

In a typical neural network without shortcut connections, each layer feeds into the next one. For a single layer, we can represent this as:

$$y = F(x)$$

,where x is the input to the layer, $F(x)$ is the transformation performed by the layer, and y is the output.

In deep networks, during backpropagation, the gradients need to traverse all the way from the output back to the first layers. With each layer, these gradients can get smaller (or, in some cases, larger), leading to the vanishing (or exploding) gradient problem.

However, ResNets add a skip or shortcut connection that allows the gradient to propagate directly back to earlier layers. In a layer (or set of layers) with a skip connection, the output is calculated as:

$$y = F(x) + x$$

This equation means that the original input x is added to the output of the layer transformation $F(x)$. Here $F(x)$ refers to the stacked layers with their nonlinear transformations, and x is the identity function which does not change the input.

By adding x directly to the output, we create a path along which the gradient can propagate directly without being attenuated or amplified by a series of transformations. It means, even if $F(x)$ approaches zero or grows very large, the skip connection ensures a pathway for the gradients to flow.

If the dimensions of x and $F(x)$ do not match, a linear transformation (usually a 1x1 convolution) is applied to x to change its dimensions so that it can be added to $F(x)$. In such a case, the equation becomes:

$$y = F(x) + W_s x$$

,where W_s is the weight matrix of the linear transformation.

These shortcut connections allow the gradients to propagate directly through several layers without attenuation, alleviating the vanishing gradient problem. By using these connections, ResNet can efficiently train networks with 100 layers and more, which was a challenging task previously due to the vanishing gradient problem. It has demonstrated significant improvement in performance and has won several major computer vision competitions.

4.4.1 Network architecture

ResNet-50 is a 50-layer convolutional neural network (48 convolutional layers, one MaxPool layer, and one average pool layer). Residual neural networks are a type of artificial neural network (ANN) that forms networks by stacking residual blocks [10].

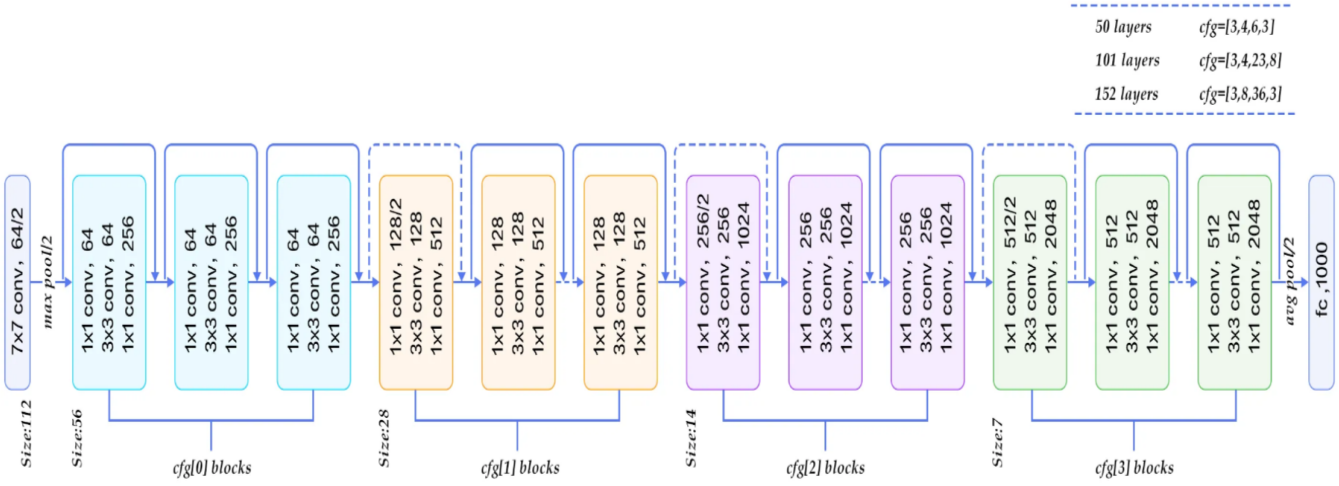

Figure 4.6: Chart of ResNet-50 architecture [10].

The ResNet-50 architecture can be broken down into 6 parts as it is illustrated above:

- Input Pre-processing
- Cfg[0] blocks
- Cfg[1] blocks
- Cfg[2] blocks
- Cfg[3] blocks
- Fully-connected layer

Different versions of the ResNet architecture use a varying number of Cfg blocks at different levels, as mentioned in the figure above. A detailed, informative listing can be found below.

ResNet-50 has an architecture based on the elements depicted above, but with one important difference. The 50-layer ResNet uses a bottleneck design for the building block. A bottleneck residual block uses 1×1 convolutions, known as a "bottleneck", which reduces the number of parameters and matrix multiplications. This enables much faster training of each layer. It uses a stack of three layers rather than two layers.

The 50-layer ResNet architecture includes the following elements, as shown in the table below.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	[3×3, 64; 3×3, 64] ×2	[3×3, 64; 3×3, 64] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×2	[3×3, 128; 3×3, 128] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×8
conv4_x	14×14	[3×3, 256; 3×3, 256] ×2	[3×3, 256; 3×3, 256] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×23	[1×1, 256; 3×3, 256; 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512; 3×3, 512] ×2	[3×3, 512; 3×3, 512] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				

Figure 4.7: Comparison between the ResNet Configurations in the architecture [10].

Analysing both figures presented in this section the ResNet50 contains:

- A 7×7 kernel convolution alongside 64 other kernels with a 2-sized stride (**1 layer**).
- A max pooling layer with a 2-sized stride (**1 layer**).
- 3×3,64 kernel convolution, another with 1×1,64 kernels, and a third with 1×1,256 kernels - layers iterated 3 times (**9 layers**).
- 1×1,128 kernels, 3×3,128 kernels, and 1×1,512 kernels - layers iterated 4 times (**12 layers**).
- 1×1,256 cores, and 2 cores 3×3,256 and 1×1,1024 - layers iterated 6 times (**18 layers**).
- 1×1,512 cores, 3×3,512 cores, and 1×1,2048 cores - layers iterated 3 times (**9 layers**).
- Average pooling, followed by a fully connected layer with 1000 nodes, using the softmax activation function.

As it is presented in the table above, the difference in the complexity architecture between the ResNet models is only about number of iterations in each block.

4.5 EfficieNet-B4

EfficientNet-B4 is a highly efficient and powerful CNN architecture that has been successfully applied to various computer vision tasks, including image classification. It can also be a

good choice for breast cancer classification due to its ability to handle complex and diverse image data.

The primary innovation of EfficientNet lies in its scaling strategy. Instead of scaling up only one aspect of the network such as depth (number of layers), width (number of channels), or resolution (input image size), EfficientNet scales up all three dimensions in a balanced way. This is based on a principled approach called compound scaling.

To find the optimal scaling strategy, EfficientNet's designers first used a small network and performed a grid search to identify the relationship between different scaling dimensions and the accuracy of the network. From this, they developed a compound scaling method which can scale the depth, width, and resolution simultaneously using a set of fixed scaling coefficients.

The EfficientNets, especially the larger variants, have demonstrated state-of-the-art performance on a variety of benchmarks, including ImageNet, while also being more resource-efficient.

The architecture of EfficientNet-B4 consists of several key elements:

- Convolutional Layers: EfficientNet-B4 incorporates a series of convolutional layers that perform convolutions on the input image. These layers extract and learn hierarchical features at different levels of abstraction.
- Depthwise Separable Convolution: This type of convolutional operation separates the standard convolution into two steps: depthwise convolution and pointwise convolution. It reduces the computational complexity while preserving the representational power of the model.
- Inverted Residual Blocks: EfficientNet-B4 employs inverted residual blocks, which consist of a bottleneck layer, an expansion layer, and a projection layer. This design reduces the number of parameters and computational cost, while still capturing and propagating important features through the network.
- Squeeze-and-Excitation (SE) Blocks: SE blocks are used to improve feature recalibration and adaptively emphasize informative features. These blocks incorporate global information by using a squeeze operation to capture channel-wise statistics and an excitation operation to recalibrate feature maps.
- Skip Connections: EfficientNet-B4 employs skip connections to improve gradient flow and enable the network to capture features at different scales. These connections help propagate information from early layers directly to deeper layers, which aids in better feature representation.
- Global Average Pooling: Instead of fully connected layers at the end of the network, EfficientNet-B4 utilizes global average pooling, which reduces the number of parameters and helps prevent overfitting. It aggregates the spatial information across feature maps into a single vector.
- Fully Connected Layer and Softmax Activation: The final fully connected layer in EfficientNet-B4 takes the output of global average pooling and maps it to the desired

number of classes for breast cancer classification. The softmax activation function is applied to produce class probabilities.

EfficientNet-B4, with its efficient architecture and advanced design choices, strikes a good balance between model size, computational complexity, and accuracy. It is known for its strong performance on image classification tasks, making it a promising choice for breast cancer classification when trained on relevant datasets.

4.5.1 Network architecture

The architecture follows the principles of compound model scaling, where depth, width, and image resolution are all uniformly increased. This approach ensures that the network stays balanced and prevents the network from becoming too large, which could potentially limit its performance.

The EfficientNet family includes models ranging from EfficientNet-B0 to EfficientNet-B7, where each subsequent model is a scaled-up version of the previous one following the compound scaling method.

The architecture will be discussed in a block approach.

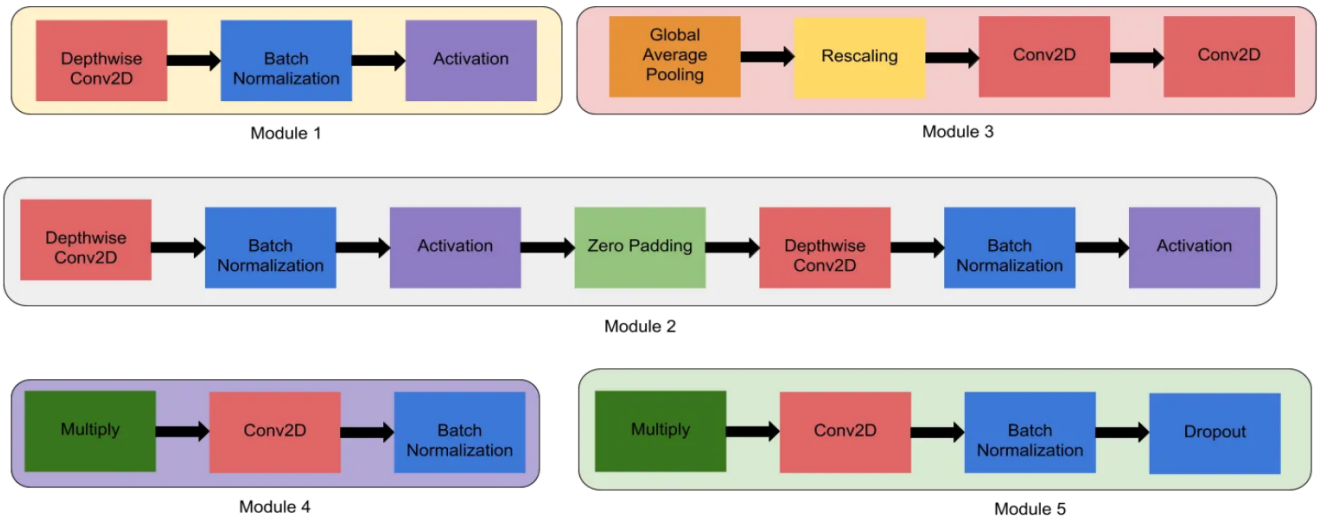

Figure 4.8: The modules diagram which are part of the EffNet architecture. Graphic source: [1]

- Module 1 is used as a starting point for the sub-blocks.
- Module 2 is used as a starting point for the first sub-block of all the 7 main blocks except the 1st one.
- Module 3 is connected as a skip connection to all the sub-blocks.
- Module 4 is used for combining the skip connection in the first sub-blocks.
- Module 5. Using this module each sub-block is connected to its previous sub-block in a skip connection and they are combined.

These modules are further combined to form sub-blocks which will be used in a certain way in the blocks.

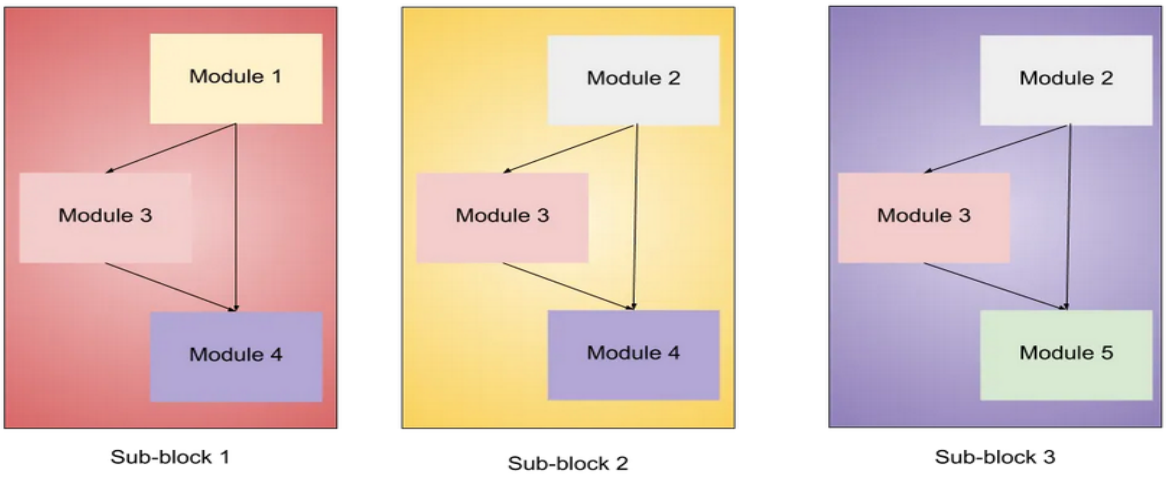

Figure 4.9: The sub-blocks diagram. Graphic source: [1]

- Sub-block 1 is used only used as the first sub-block in the first block.
- Sub-block 2 is used as the first sub-block in all the other blocks.
- Sub-block 3 is used for any sub-block except the first one in all the blocks.

EfficientNet-B4, like other EfficientNet models, is composed of a series of stages each containing one or more MBConv blocks, which are variations of the inverted residual blocks used in MobileNetV2. Each block applies operations in the following order: expand, depthwise convolve, squeeze-and-excitation (SE), and project.

- Input Layer: EfficientNet-B4 takes an input image size of 224x224 pixels.
- Stem: The stem is the initial part of the network, consisting of a single convolutional layer with a kernel size of 3x3, stride of 2, and output dimension of 48.
- MBConv Blocks: EfficientNet-B4 contains 23 MBConv blocks divided into 7 stages. The MBConv blocks include batch normalization and a Rectified Linear Unit (ReLU) as part of their structure. Depthwise separable convolutions are used within these blocks for computational efficiency.

  In more detail, an MBConv block first expands the input tensor by applying 1x1 convolutions, then applies a 3x3 or 5x5 depthwise convolution, followed by squeeze-and-excitation operation, and finally a 1x1 convolution to project the tensor back to a lower dimension. Batch normalization and ReLU are applied after the first two convolutions. The squeeze-and-excitation operation is a mechanism to recalibrate channel-wise feature responses adaptively.

- Final Layers: After the last stage of MBConv blocks, a 1x1 convolution is applied to the output tensor, followed by a global average pooling layer that reduces the spatial dimensions to 1x1. A dropout layer is then used for regularization, and finally, a fully connected layer outputs the final class probabilities using a softmax activation function.

EfficientNet-B4 utilizes skip connections and compound scaling, enhancing the flow of gradients during training and allowing the model to learn more complex patterns. This architecture has demonstrated excellent performance in various image classification tasks and can be fine-tuned for specific applications, making it a versatile tool for many computer vision problems.

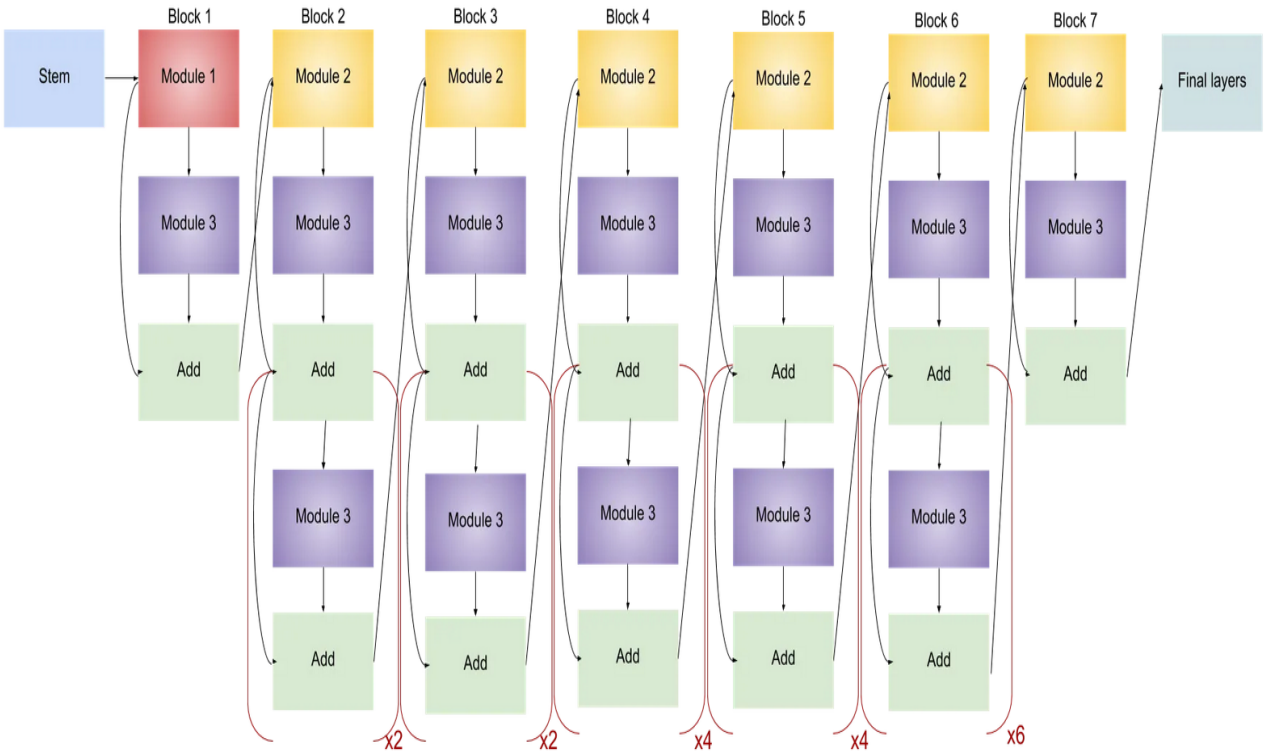

Figure 4.10: The EffNet-B4 architecture in block approach. Graphic source: [1]

Model	Top-1 Acc.	Top-5 Acc.	#Params	Ratio-to-EfficientNet	#FLOPS	Ratio-to-EfficientNet
EfficientNet-B0	**76.3%**	**93.2%**	**5.3M**	**1x**	**0.39B**	**1x**
ResNet-50 (He et al., 2016)	76.0%	93.0%	26M	4.9x	4.1B	11x
DenseNet-169 (Huang et al., 2017)	76.2%	93.2%	14M	2.6x	3.5B	8.9x
EfficientNet-B1	**78.8%**	**94.4%**	**7.8M**	**1x**	**0.70B**	**1x**
ResNet-152 (He et al., 2016)	77.8%	93.8%	60M	7.6x	11B	16x
DenseNet-264 (Huang et al., 2017)	77.9%	93.9%	34M	4.3x	6.0B	8.6x
Inception-v3 (Szegedy et al., 2016)	78.8%	94.4%	24M	3.0x	5.7B	8.1x
Xception (Chollet, 2017)	79.0%	94.5%	23M	3.0x	8.4B	12x
EfficientNet-B2	**79.8%**	**94.9%**	**9.2M**	**1x**	**1.0B**	**1x**
Inception-v4 (Szegedy et al., 2017)	80.0%	95.0%	48M	5.2x	13B	13x
Inception-resnet-v2 (Szegedy et al., 2017)	80.1%	95.1%	56M	6.1x	13B	13x
EfficientNet-B3	**81.1%**	**95.5%**	**12M**	**1x**	**1.8B**	**1x**
ResNeXt-101 (Xie et al., 2017)	80.9%	95.6%	84M	7.0x	32B	18x
PolyNet (Zhang et al., 2017)	81.3%	95.8%	92M	7.7x	35B	19x
EfficientNet-B4	**82.6%**	**96.3%**	**19M**	**1x**	**4.2B**	**1x**
SENet (Hu et al., 2018)	82.7%	96.2%	146M	7.7x	42B	10x
NASNet-A (Zoph et al., 2018)	82.7%	96.2%	89M	4.7x	24B	5.7x
AmoebaNet-A (Real et al., 2019)	82.8%	96.1%	87M	4.6x	23B	5.5x
PNASNet (Liu et al., 2018)	82.9%	96.2%	86M	4.5x	23B	6.0x
EfficientNet-B5	**83.3%**	**96.7%**	**30M**	**1x**	**9.9B**	**1x**
AmoebaNet-C (Cubuk et al., 2019)	83.5%	96.5%	155M	5.2x	41B	4.1x
EfficientNet-B6	**84.0%**	**96.9%**	**43M**	**1x**	**19B**	**1x**
EfficientNet-B7	**84.4%**	**97.1%**	**66M**	**1x**	**37B**	**1x**
GPipe (Huang et al., 2018)	84.3%	97.0%	557M	8.4x	-	-

We omit ensemble and multi-crop models (Hu et al., 2018), or models pretrained on 3.5B Instagram images (Mahajan et al., 2018).

Figure 4.11: Comparison between the EffNet configurations in the architecture. [18].

It's easy to see the difference among all the models and they gradually increased the number of sub-blocks Choosing the B4 architecture was a kind of trade off between high flop number and the accuracy as it is shown on the comparison table above.

4.6 Comparison between the CNN architectures

So far moving into conclusion after presenting three different CNN models might be said that the order of the discussed networks was from the simplest to the most complex. That might also implicate a fact that the more complex should mean better, so best accuracy. That will be tested in the next chapter.

Here's a comparison of the three models in the context of breast cancer classification:

Model	**Main Features**	**Pros**	**Cons**
VGG19	19-layer deep network with small (3x3) convolution filters.	1. VGG19's architecture is simple and uniform, which can make it easier to understand and modify. 2. The small convolution filters allow it to learn more complex features. 3. Like ResNet50, VGG19 has been pre-trained on large datasets, so it also has good feature extraction capabilities.	1. VGG19 is very computationally intensive due to its depth and the number of fully-connected nodes. 2. It may not perform as well as other models on tasks that benefit from a deeper or more complex architecture.
ResNet50	50-layer deep network with "skip connections"or "shortcuts"that mitigate the problem of vanishing gradients.	1. Skip connections allow for a more effective flow of gradients during backpropagation, which can help to train deeper networks. 2. ResNet50 has been pre-trained on large datasets and therefore has strong feature extraction capabilities. 3. It is a good balance between computational efficiency and performance for many tasks.	1. Computationally expensive. 2. Potential overfitting for small datasets.
EfficientNet-B4	Uses a compound scaling method to uniformly scale up the depth, width, and resolution of the network. It has more layers than ResNet50 and VGG19.	1. EfficientNets are designed to provide good performance while being efficient in terms of computational resources. 2. They have been pre-trained on large datasets, so they also have strong feature extraction capabilities. 3. The EfficientNet-B4 variant is particularly powerful, often outperforming other models on a variety of tasks. 4. It can handle a wide range of input sizes, making it a versatile choice for different tasks.	1. EfficientNets are more complex and may be more difficult to understand or modify. 2. They may require more computational resources and take longer to train than other models.

Table 4.1: Comparison of CNN used in the paper for breast cancer classification

These models can be fine-tuned for breast cancer classification using transfer learning. Transfer learning involves taking a pre-trained model and training it on the specific task, in this case - breast cancer classification. The idea is that the model has already learned useful features from the large dataset that can be used as a starting point for the new task, which can help to improve performance, particularly when the new dataset is relatively small.
The best model for breast cancer classification would depend on various factors such as the size and quality of your dataset, the computational resources you have available, and the specific requirements of your task. It is generally a popular approach to try out different models and choose the one that performs best on a validation set.

Chapter 5

Experimental part

Here there gonna be a look into how deep learning models are trained on mammography images, understanding the fine balance between recall and precision, and how improvements in these models can contribute significantly to early detection and prognosis.

In the experiment pytorch baseline has been used in the python version 3.9 using NVIDIA CUDA.
CUDA is a software layer that gives direct access to the GPU's virtual instruction set and parallel computational elements.

5.1 The Ensembly network

For the input after preprocessing and the proper data transformations the goal was to combine the metadata with images in a proper type of data. For the CUDA purposes the data loader stores the information about 3 components- the image, metadata and the target. All of these elements are being the form of torch library tensors. All the data collections are splitted into 3 batches. In each of them it can be found the the **image** with shape $[3, 3, 224, 224]$, the **metadata** with shape $[3, 5]$, the **target** with shape $[3, 1]$. Having all needed and properly processed elements the following networks are using transformations and manipulations inside their architectures which are based on the numbers of neutrons in the layers, which will be presented further

In the experimental part, all the simulations are running on the same data set. For the purposes of greater reliability of the fit of the predictive models to the data in each method, simulations are performed separately for the learning and test parts. During the learning phase the cross validation takes place that is why the validation samples appears.

The data is split randomly. The training sample contains 80% (stratified train-test split) of observations from the data set. The final representative sample for testing task counts 220 records from the test sample.

Table 5.1: Collection of the main parameters being used in the training and evaluation process of the CNN

Parameter name	**Value**
Folds	2
Epochs	5
Patience	3
Workers	8
Learning rate	0.0005
Weight decay	0.0
Learning rate patience	1
Learning rate factor	0.4
Batch size for learning	16
Batch size for validation	8
Batch size for testing	3

5.2 The used networks on the RSNA data

5.2.1 ResNet-50

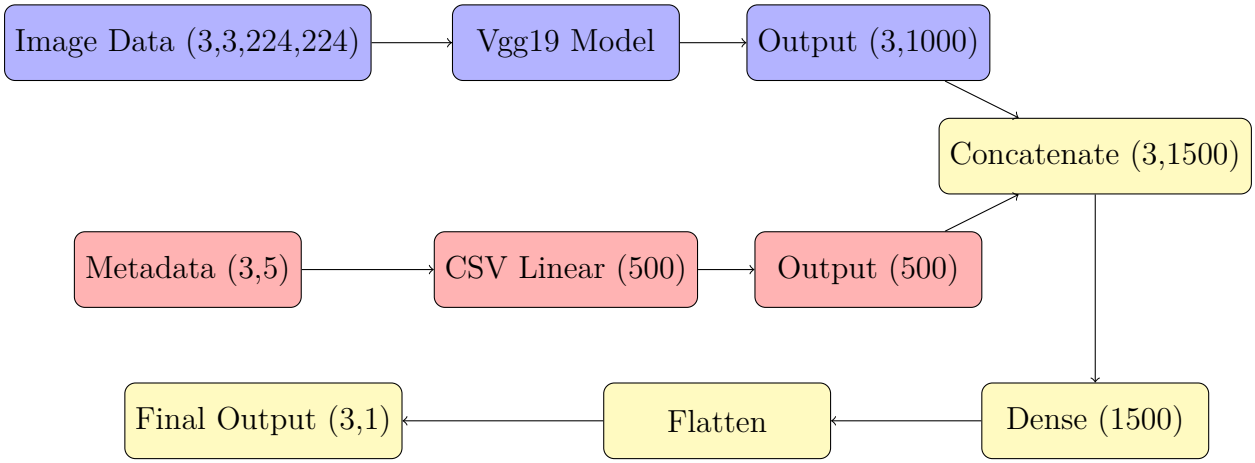

Figure 5.1: The schema of combaining networks of metada with the image data for ResNet-50.

The architecture above combines both image and metadata information for a classification task. Here's a breakdown of its components:

- ResNet50 Features (Image CNN): The ResNet50 model is used as the feature extractor for the input images. It's initialized with pre-trained weights (transfer learning) to capture high-level features from the images. The output of this part is a feature vector from the ResNet50 model.
- Metadata Processing (CSV FNN). Metadata is processed using a Feedforward Neural Network (FNN). It consists of:
  - A linear layer that takes the input metadata with 5 features and produces 500 output features.

  - Batch normalization to normalize the output of the linear layer.
  - ReLU activation function to introduce non-linearity.
  - Dropout layer with a dropout probability of 0.2 to prevent overfitting.

- Concatenation: The outputs from the ResNet50 feature extraction and the processed metadata are concatenated along the feature dimension. This combines the extracted image features with the processed metadata information.
- Classification Layer: The concatenated feature vector is then fed into a linear layer for classification. The linear layer combines both image and metadata features to make predictions. The output size of this layer matches the desired number of output classes.
- Forward Pass: The forward method defines how data flows through the network:
  - Input images (image) and metadata (meta) are passed through the ResNet50 feature extractor and metadata processing layers, respectively.
  - The resulting image features and processed metadata are concatenated.
  - The concatenated features are then fed into the classification linear layer to produce the final classification scores.

Overall, this architecture takes advantage of both image-based features extracted by a pre-trained ResNet50 model and metadata information to perform a joint classification task. The combination of these two types of information is aimed at improving classification performance.

The model's trainig process took 15 hours and 42 minutes. The Early Stopping phase has been enabled after the last epoch in the second fold, so already after the whole training (no improvement since 3 models).

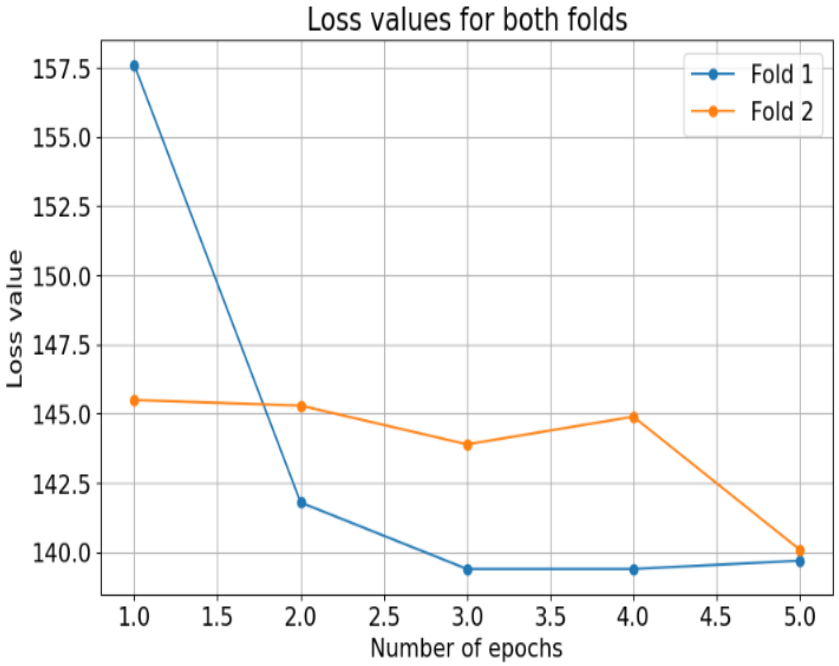

Figure 5.2: Comparison of the loss values between the folds during ResNet 50 model training

This visualization helps to understand how the loss values evolve during the training process for both folds of the ResNet-50 model. It can provide insights into the convergence and performance of the training process.

Overall, both folds show a similar trend of decreasing loss values over the first few epochs. However, Fold 1 seems to achieve a lower and more consistent loss value compared to Fold 2. This could indicate that the model's performance is better in Fold 1, possibly due to better convergence or more favorable data distribution.

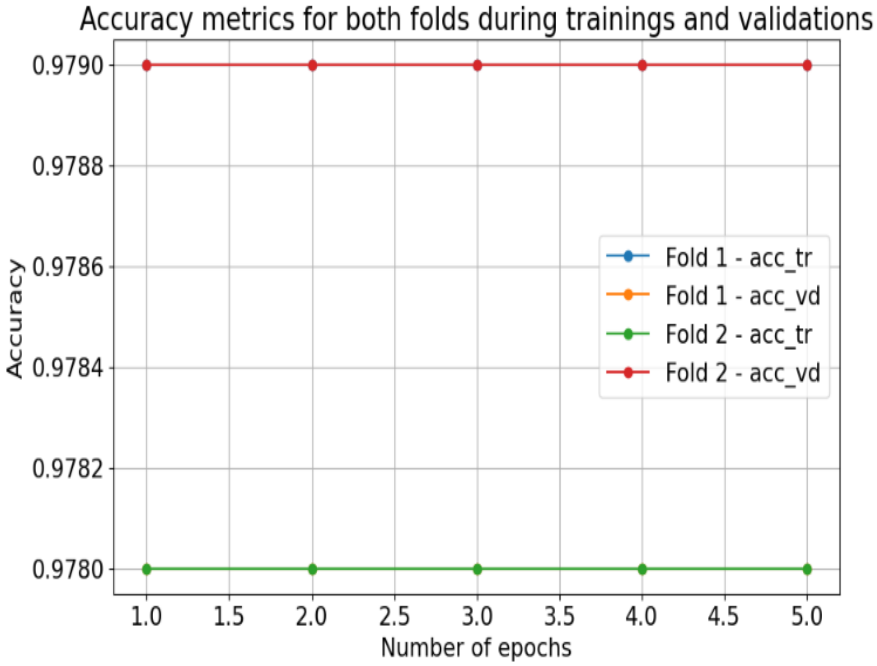

Figure 5.3: Comparison of the accuracy values between the folds during ResNet 50 model training

In order to analyze the accuracy values between the folds during the training of the ResNet-50 model for both training and validation samples. The values for all groups are constant, there is no distinction of values due to folds. The accuracy for validation are higher than for training.

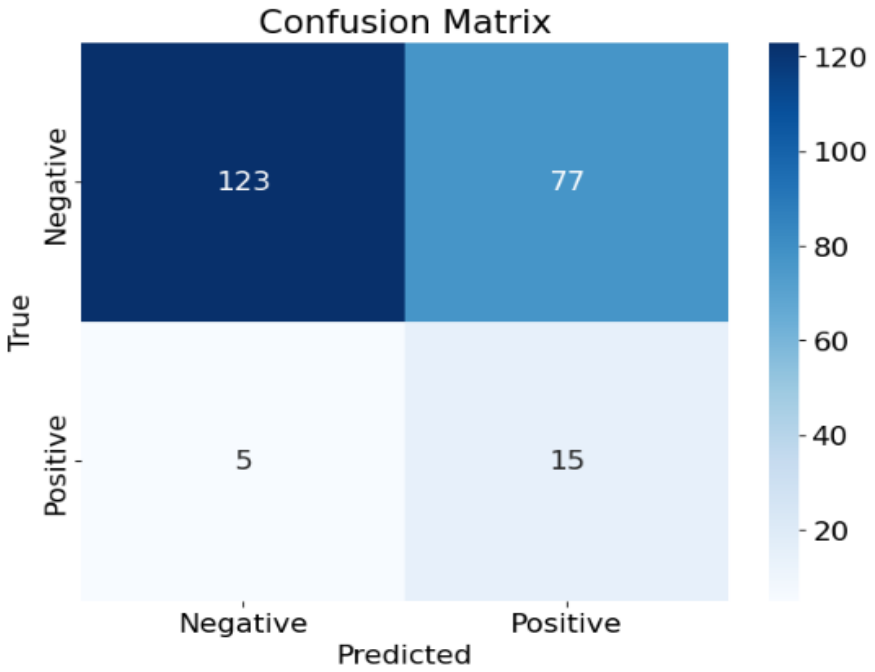

Figure 5.4: Confusion Matrix of ResNet 50 for prediction values.

- 15 from 20 positive classes from the test set have been correctly classified.
- The most significant value - the false negative value equals 5.
- Overall model has predicted in total 128 negative and 92 positive classes.

5.2.2 Vgg19

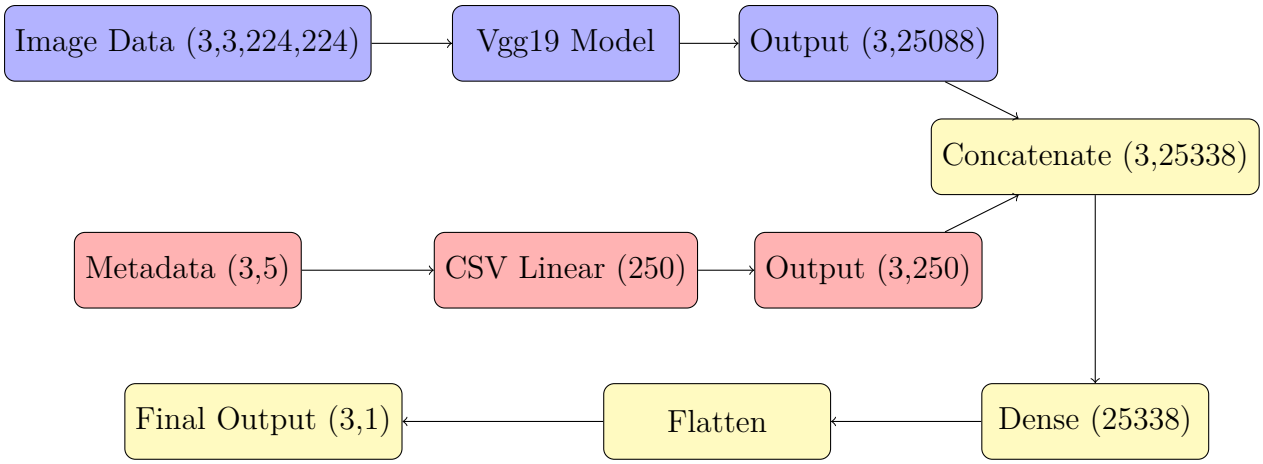

Figure 5.5: The schema of combaining networks of metada with the image data for Vgg19.

Here's a brief summary of Vgg19 components:

- Vgg19 Features (Image CNN): The VGG19 model, with its classification part removed, is utilized as the feature extractor for the input images. The pretrained VGG19 features are used to capture hierarchical features from the images.

- Metadata Processing (CSV FNN). Metadata is processed using a Feedforward Neural Network (FNN) consisting of multiple layers:
  - Two linear layers with 250 output features each.
  - Batch normalization is applied after each linear layer.
  - ReLU activation introduces non-linearity.
  - Dropout layers with a dropout probability of 0.2 help prevent overfitting.
- Concatenation: Same as in ResNet50 network.
- Classification Layer: A final classification layer is constructed using a linear layer. The concatenated feature vector, with dimension 25338 (resulting from the concatenation of VGG19 features and metadata FNN outputs), is fed into this linear layer.
- Forward Pass: The forward method defines how data flows through the network:
  - Input images (image) and metadata (meta) are passed through the ResNet50 feature extractor and metadata processing layers, respectively.
  - VGG19 features are extracted and then flattened to create a feature vector.
  - The processed metadata is obtained from the metadata FNN layers.
  - The resulting image features and processed metadata are concatenated.
  - The concatenated features are then fed into the classification linear layer to produce the final classification scores.

Overall, the VGG19Network architecture utilizes VGG19 pretrained features and processed metadata to make predictions. The combination of these features from both data sources aims to improve the model's classification performance.

The model's trainig process took 17 hours and 11 minutes. The Early Stopping phase has been enabled after the last epoch in the second fold, so already after the whole training (no improvement since 3 models).

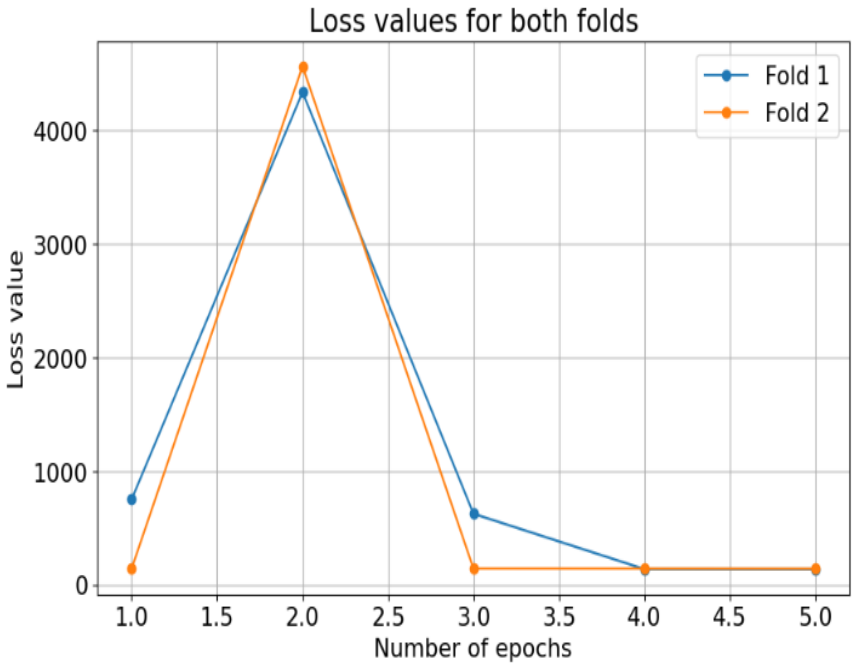

Figure 5.6: Comparison of the loss values between the folds during Vgg19 model training

Overall, both folds seem to show patterns of instability during training. The significant initial spikes in loss followed by decreases might indicate that the training process is not smooth and might require further investigation. Such behavior could result from various factors, including issues with learning rate, data preprocessing, or the training schedule. It might be beneficial to analyze the training procedure and consider adjustments to stabilize the training process and achieve better convergence.

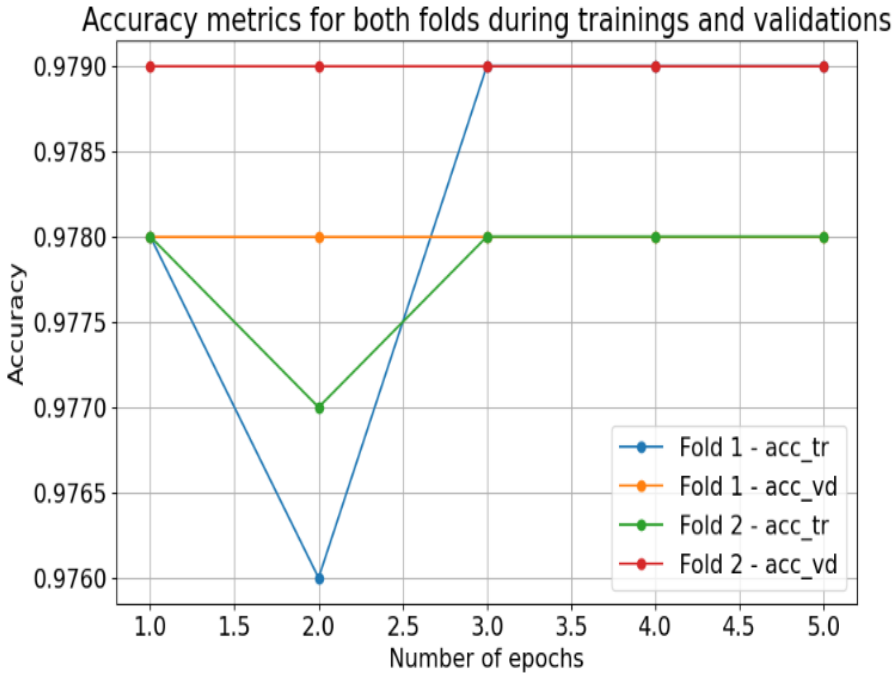

Figure 5.7: Comparison of the accuracy values between the folds during Vgg19 model training

Both folds show consistent accuracy values for both training and validation samples over the epochs. The fluctuations are minimal, indicating that the model's performance is steady and converging during training. The fact that the training and validation accuracies are closely aligned suggests that the model is not overfitting to the training data, as the performance on unseen validation data remains similar. This stability is a positive sign, indicating that the model is learning effectively and generalizing well to new data.

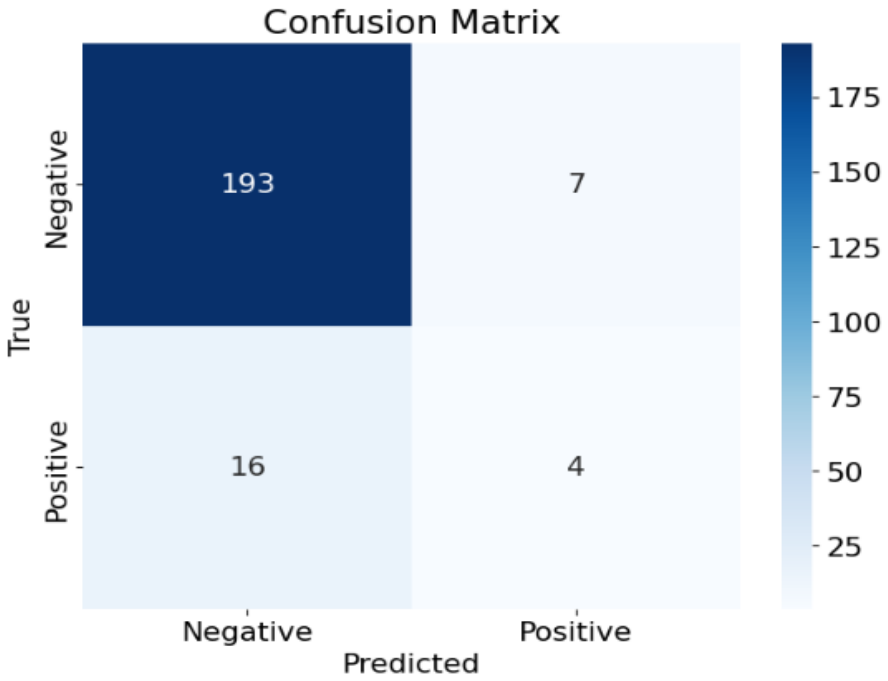

Figure 5.8: Confusion Matrix of Vgg 19 for prediction values.

- 4 from 20 positive classes from the test set have been correctly classified.
- The most significant value - the false negative value equals 16.
- Overall model has predicted in total 209 negative and 11 positive classes.

5.2.3 EfficientNet-B4

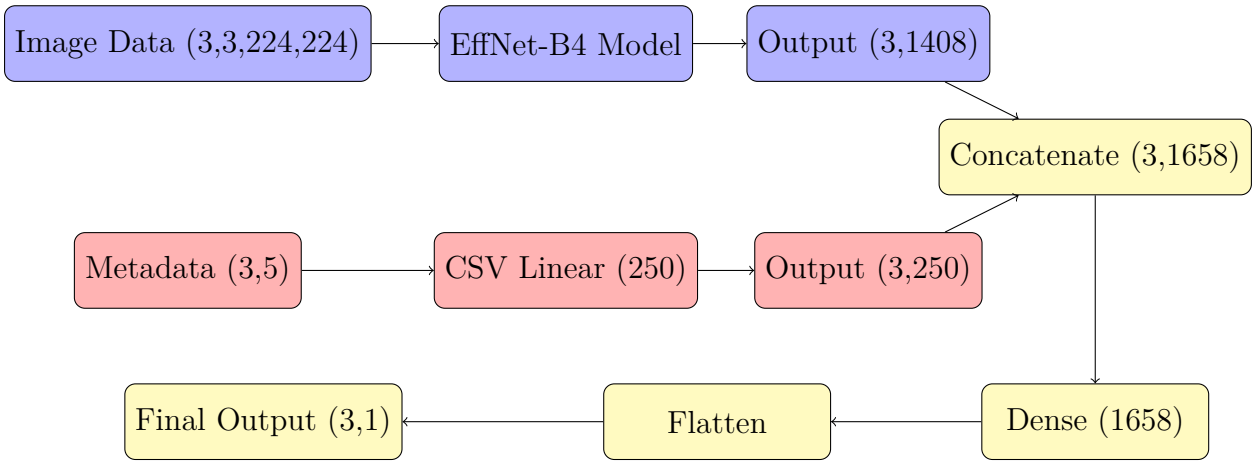

Figure 5.9: The schema of combining networks of metadata with the image data for EfficieNet-B4.

Here's a brief overview of the EffnetNetwork-B4 architecture in the described task:

- EfficientNet-B4 Features (Image CNN): The EfficientNet model, specifically the 'efficientnet-b4' variant, is employed as the feature extractor for the input images.

The model is initialized with pre-trained weights. The feature extraction process includes:

  - Extracting image features using the proper method.
  - Applying average pooling to reduce spatial dimensions.

- Metadata Processing (CSV FNN). Metadata is processed using a Feedforward Neural Network (FNN) consisting of several layers:
  - The first linear layer takes the input metadata with 5 features and produces an output of size 250.
  - Batch normalization to normalize the output of the linear layer.
  - ReLU activation introduces non-linearity.
  - Dropout layer with a dropout probability of 0.2 to prevent overfitting.
  - Another linear layer with the same architecture (linear, batch normalization, ReLU, dropout) is added for further processing.
- Concatenation: The outputs from the EfficientNet feature extraction and the processed metadata FNN are concatenated along the feature dimension. This combines the extracted image features with the processed metadata information.
- Classification Layer: A final classification layer is constructed using a linear layer. The concatenated feature vector, now with dimension 1658 (resulting from the concatenation of EfficientNet features and metadata FNN outputs), is fed into this linear layer.
- Forward Pass: The forward method defines how data flows through the network:
  - Input images (image) and metadata (meta) are passed through the EfficientNet feature extractor and metadata processing layers, respectively.
  - EfficientNet features are extracted and then pooled to create a compact representation.
  - The processed metadata is obtained from the metadata FNN layers.
  - The resulting image features and processed metadata are concatenated.
  - The concatenated features are passed through the classification linear layer to produce the final classification scores.

In summary, the EffNetNetwork architecture employs an EfficientNet model for image feature extraction and a multi-layer FNN for processing metadata. These features are combined and used for classification through a linear classification layer. The architecture is designed to handle both image and metadata inputs for a joint classification task.

The model's trainig process took 16 hours and 26 minutes. The Early Stopping phase has been enabled after the fourth epoch in the second fold - no improvement since first model in the second fold.

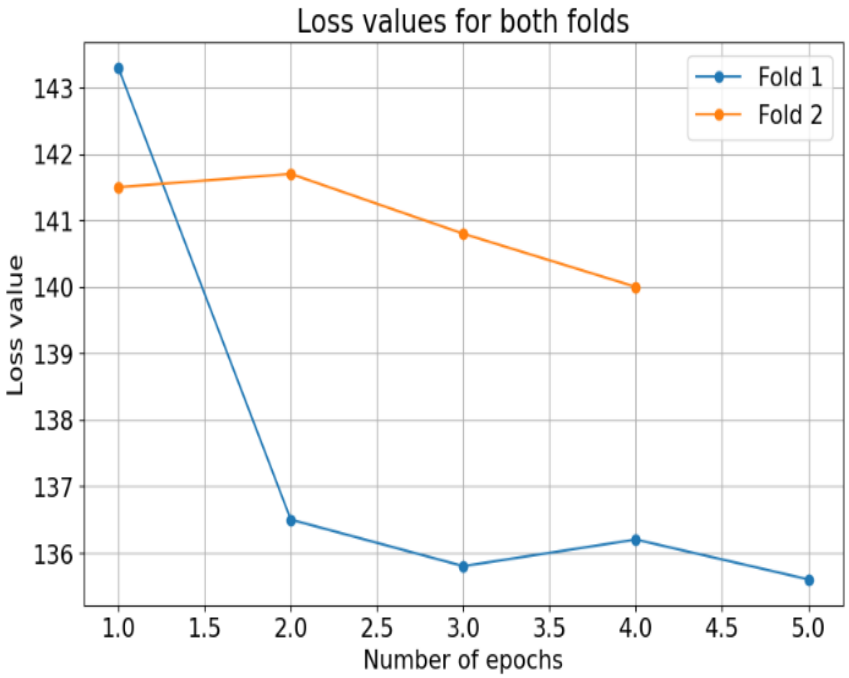

Figure 5.10: Comparison of the loss values between the folds during EffNet-B4 model training

Both folds exhibit a smooth decline in loss values, which is generally a positive sign during training. This decreasing trend suggests that the model is learning and improving its predictions as training proceeds. The relatively consistent patterns for both folds indicate that the training process is stable and well-behaved, which is important for achieving good convergence and a well-generalized model.

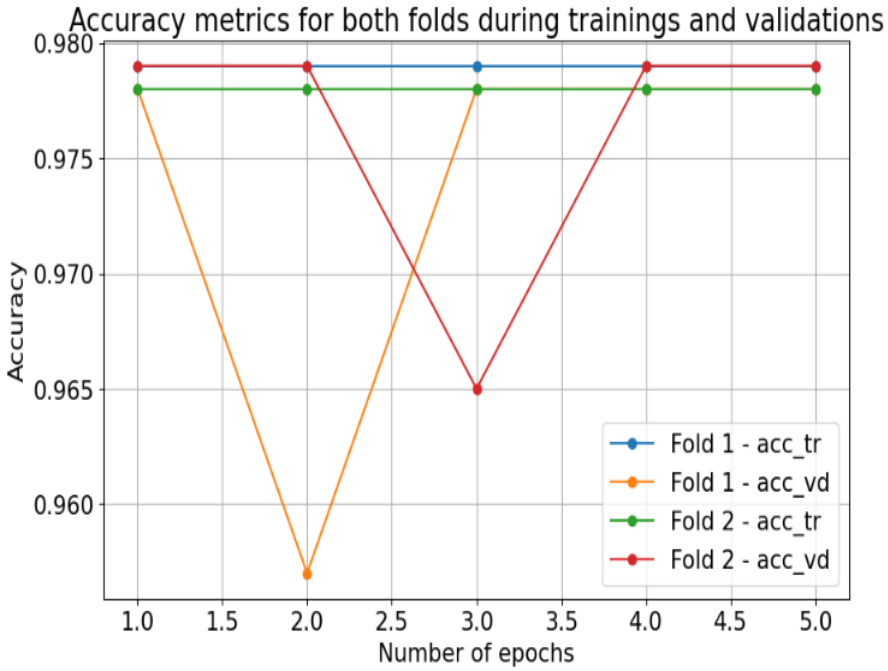

Figure 5.11: Comparison of the accuracy values between the folds during EffNet-B4 model training

Overall, both folds demonstrate consistent accuracy values for both training and validation samples over the epochs. The small fluctuations in validation accuracy, especially in Fold 1, might indicate some variability in the model's performance on unseen data. However, the accuracies are generally stable, and the training process appears to be converging effectively. The fact that the training and validation accuracies are closely aligned suggests that the model's performance on the validation set is similar to its performance on the training set, indicating good generalization.

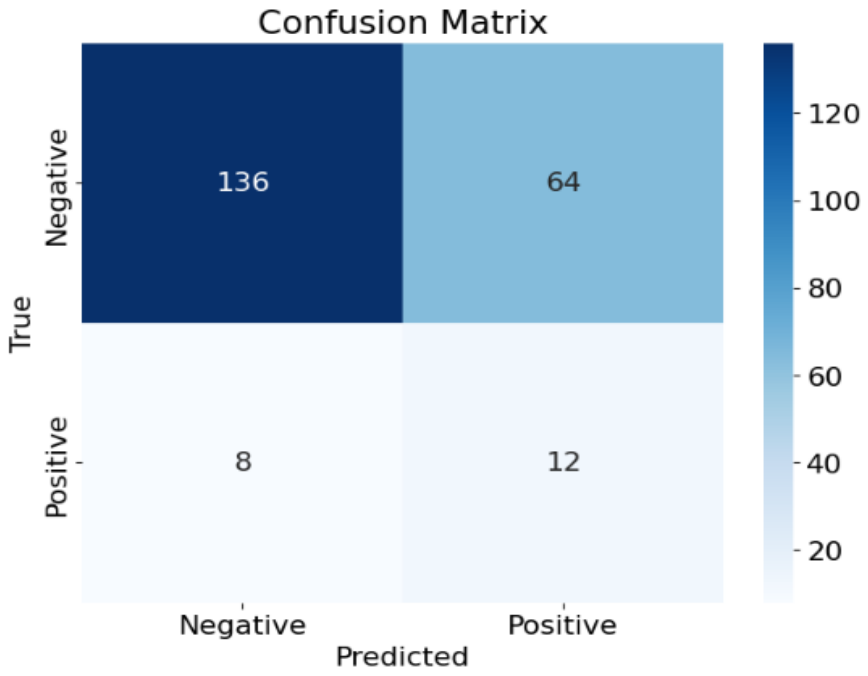

Figure 5.12: Confusion Matrix of EffNet-B4 for prediction values.

- 12 from 20 positive classes from the test set have been correctly classified.
- The most significant value - the false negative value equals 8.
- Overall model has predicted in total 144 negative and 76 positive classes.

5.3 Networks Comparison and Summary

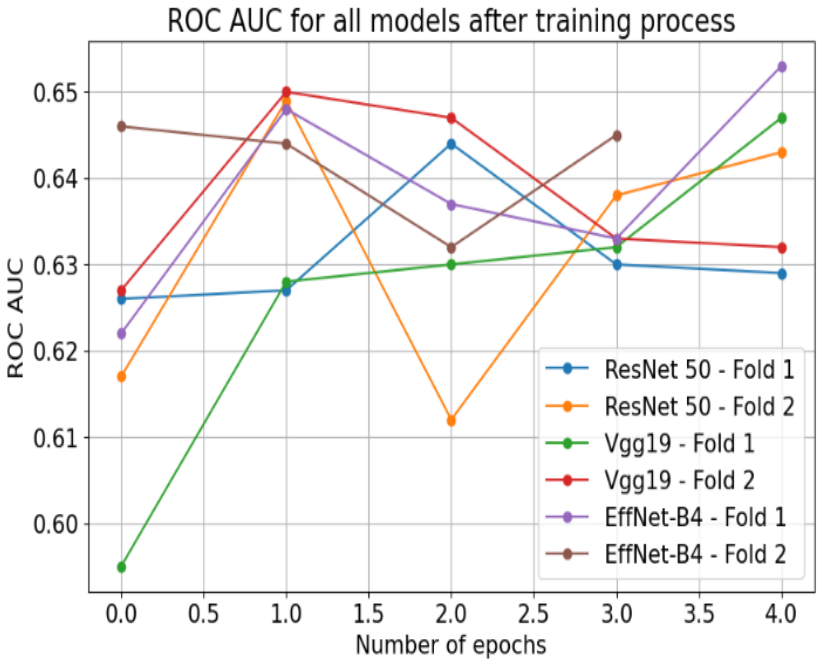

Figure 5.13: Comparison of the ROC values between the models after the training process

Overall, the ROC AUC values for all models and folds seem to remain within a relatively tight range. While there are fluctuations in the ROC AUC values across epochs, the differences are not extreme. It's important to note that ROC AUC values are influenced by the trade-off between sensitivity and specificity, and the performance might vary depending on the classification threshold.

The visualization provides insights into the comparative performance of different models on the chosen classification task, showing that EffNet-B4 tends to achieve slightly higher ROC AUC values compared to ResNet 50 and Vgg19. However, it's recommended to analyze other evaluation metrics, such as precision, recall, and F1-score, alongside ROC AUC to obtain a more comprehensive understanding of model performance.

Table 5.2: Table of comparing Classification Errors of all algorithms

Metrics	**ResNet 50**	**Vgg19**	**EffNet-B4**
AUC	0.658	0.6445	**0.675**
F1 Score	**0.268**	0.2581	0.25
Macro F1	0.509	**0.6009**	0.5203
Micro F1	**0.600**	0.2857	0.5455
Balanced Accuracy	**0.6825**	0.5825	0.64
Accuracy	0.6273	**0.8955**	0.6727
Recall	**0.7500**	0.2000	0.6000
Precision	0.1630	**0.3636**	0.1579

From the results, it can be observed that the performance of the three models varies across different metrics. Conclusions regarding the algorithms presented in the experimental part:

- Vgg19 is achieving the best results in metrics: Macro F1, Accuracy and Precision. The recall here is in a really low level - 0.2 and it is related with high accuracy. On the other hand the highest precision is connected with the lowest costs of the false positive cases consequences. The network is predicting to many negative values, which has bad impact for this kind of classification problem.

- Effnet has only AUC metric as the best result, however here for all networks these metric is tight. As the disadvantage the precision is in a really low level. The recall metric in in between Resnet and Vgg19. The training loss values looked promising in the chart, as it was said it definietely needs further investigation.

- ResNet 50 shows competitive performance in terms of AUC, Balanced Accuracy, and Recall and in F1 Score. High value for recall and low for precision are the effects of the high number of positive predictions.

The choice of the best model might depend on the specific goals and priorities of the task, considering factors like false positives/negatives and trade-offs between precision and recall. That is why the most obvious as the best representative metric would be F1 score. Unfortunately for all of the networks the metrics obtained not satisfying results.

All the discussed architectures of networks obtained similar scores, the reason might be connected with the same parameters which have been used during the training process and the fact of using identically same samples (also ImageNet). Also training times were similar. Of course there are differences, first of all the number of predictions of negative and positive classes. As it was told before, the false negative error is much more serious than the false positive errors, that is why taking into consideration all of the factors the ResNet-50 neural network has been chosen as the final one. The use of this solution would be of course most cost consuming- there would be definitely to many positive patients after the network predictions.

Chapter 6

Summary and conclusion

In recent years, the field of medical imaging has been greatly influenced by the advancements in deep learning techniques. This thesis focused on the crucial task of breast cancer classification using Convolutional Neural Networks. Breast cancer is a significant global health concern, and accurate early detection plays a vital role in improving patient outcomes. The integration of deep learning into radiology industries addresses the need for efficient and reliable diagnosis, offering transformative potential for the medical field.

The significance of this research lies in its potential to revolutionize breast cancer diagnosis. Traditional methods often require time-consuming manual analysis by radiologists, leading to potential errors and subjectivity. The utilization of CNNs provides an avenue for automated and standardized diagnosis, significantly reducing human error and variability. This has the potential to expedite diagnosis, ensure consistency, and ultimately improve patient care.

The implementation of deep learning networks in radiology introduces a powerful tool for image analysis. CNNs can recognize intricate patterns in medical images that might be elusive to the human eye. This technology has the capacity to enhance accuracy, efficiency, and scalability in medical image interpretation. Moreover, the automation facilitated by CNNs allows radiologists to focus more on complex cases, further boosting overall diagnostic capabilities.

However, the quest for optimal neural network architectures is a challenging endeavor. The architecture's design heavily influences the network's performance, and achieving the best results demands a profound understanding of both medical imaging and deep learning. This challenge is compounded by the necessity for substantial computational resources and time for training and fine-tuning the CNN. Proper expertise and access to powerful hardware are essential prerequisites.

In conclusion, the utilization of CNNs for breast cancer classification marks a pivotal advancement in medical imaging. The application of deep learning networks addresses the urgent need for accurate, efficient, and automated diagnosis. This technology not only enhances diagnostic accuracy but also empowers medical professionals to deliver prompt and effective care to patients. As the radiology industry continues to embrace artificial intelligence, it is imperative to acknowledge the ongoing research required to perfect network architectures and harness their full potential for the betterment of medical diagnostics.

Bibliography

[1] Agarwal, V. Complete architectural details of all efficientnet models. *Medium* (2020).

[2] Anastasiadi, Z., Lianos, G. D., Ignatiadou, E., Harissis, H. V., Mitsis, M. Breast cancer in young women: an overview. *Updates in surgery 69* (2017), 313–317.

[3] Antropova, N., Abe, H., Giger, M. L. Use of clinical mri maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. *Journal of Medical Imaging 5*, 1 (2018), 014503–014503.

[4] Beghi, G. A decade of research on thermochemical hydrogen at the joint research centre-ispra. *Hydrogen Systems* (1986), 153–171.

[5] Bentley, J., Ford, J., Taylor, L., Irvine, K., Roberts, C. Investigating linkage rates among probabilistically linked birth and hospitalization records. *BMC medical research methodology 12* (09 2012), 149.

[6] Bhalerao, K. Characterizing the reliability of a biomems-based cantilever sensor.

[7] Boyd, N. F., Guo, H., Martin, L. J., Sun, L., Stone, J., Fishell, E., Jong, R. A., Hislop, G., Chiarelli, A., Minkin, S., et al. Mammographic density and the risk and detection of breast cancer. *New England journal of medicine 356*, 3 (2007), 227–236.

[8] Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., Haworth, A. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology 65*, 5 (2021), 545–563.

[9] Gierach, G. L., Ichikawa, L., Kerlikowske, K., Brinton, L. A., Farhat, G. N., Vacek, P. M., Weaver, D. L., Schairer, C., Taplin, S. H., Sherman, M. E. Relationship between mammographic density and breast cancer death in the breast cancer surveillance consortium. *Journal of the National Cancer Institute 104*, 16 (2012), 1218–1227.

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

[11] Hijazi, S., Kumar, R., Rowen, C., et al. Using convolutional neural networks for image recognition. *Cadence Design Systems Inc.: San Jose, CA, USA 9* (2015), 1.

[12] Khan, S. H., Hayat, M., Porikli, F. Regularization of deep neural networks with spectral dropout. *Neural Networks 110* (2019), 82–90.

[13] Lin, Y., Wu, J., et al. A novel multichannel dilated convolution neural network for human activity recognition. *Mathematical Problems in Engineering 2020* (2020).

[14] Mamczur, M. Jak działają konwolucyjne sieci neuronowe (cnn)?

[15] Marmot, M. G., Altman, D., Cameron, D., Dewar, J., Thompson, S., Wilcox, M. The benefits and harms of breast cancer screening: an independent review. *British journal of cancer 108*, 11 (2013), 2205–2240.

[16] Morid, M. A., Borjali, A., Del Fiol, G. A scoping review of transfer learning research on medical image analysis using imagenet. *Computers in biology and medicine 128* (2021), 104115.

[17] Moyer, V. A., Force*, U. P. S. T. Screening for prostate cancer: Us preventive services task force recommendation statement. *Annals of internal medicine 157*, 2 (2012), 120–134.

[18] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. Imagenet large scale visual recognition challenge. *International journal of computer vision 115* (2015), 211–252.

[19] Schwartz, L. M., Woloshin, S., Fowler Jr, F. J., Welch, H. G. Enthusiasm for cancer screening in the united states. *Jama 291*, 1 (2004), 71–78.

[20] Sprague, B. L., Conant, E. F., Onega, T., Garcia, M. P., Beaber, E. F., Herschorn, S. D., Lehman, C. D., Tosteson, A. N., Lacson, R., Schnall, M. D., et al. Variation in mammographic breast density assessments among radiologists in clinical practice: a multicenter observational study. *Annals of internal medicine 165*, 7 (2016), 457–464.

[21] Stephens, K. These doctors are using ai to screen for breast cancer. *AXIS Imaging News* (2021).

[22] Stephens, K. Rsna announces launch of screening mammography breast cancer detection ai challenge. *AXIS Imaging News* (2022).

[23] Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., Li, L. Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Scientific reports 6*, 1 (2016), 27327.

[24] Zhang, X., Zou, J., He, K., Sun, J. Accelerating very deep convolutional networks for classification and detection. *IEEE transactions on pattern analysis and machine intelligence 38*, 10 (2015), 1943–1955.

[25] Zhang, Y., Zhu, H., Meng, Z., Koniusz, P., King, I. Graph-adaptive rectified linear unit for graph neural networks. In *Proceedings of the ACM Web Conference 2022* (2022), pp. 1331–1339.

[26] Zheng, Y., Yang, C., Merkulov, A. Breast cancer screening using convolutional neural network and follow-up digital mammography. 4.

[27] Zhou, X., Li, C., Rahaman, M. M., Yao, Y., Ai, S., Sun, C., Wang, Q., Zhang, Y., Li, M., Li, X., et al. A comprehensive review for breast histopathology image analysis using classical and deep neural networks. *IEEE Access 8* (2020), 90931–90956.