REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

BREAST CANCER DETECTION IN X-RAY IMAGES USING FEATURE PYRAMIDS AND CNN ALGORITHM

Ahmed Abdulkarim Kado AL-TAHHAN

Master's Thesis

Supervisor

Ass. Prof. Dr. Ayça Kurnaz TÜRKBEN

Istanbul, 2022

BREAST CANCER DETECTION IN X-RAY IMAGES USING FEATURE PYRAMIDS AND CNN ALGORITHM

Ahmed Abdulkarim Kado AL-TAHHAN

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled OPTIMIZING ROUTING AND SECURITY MECHANISM IN AD HOC NETWORKS prepared by AHMED ABDULKARIM KADO AL TAHHAN and submitted on 14/12/2022 has been **accepted unanimously** for the degree of Master of Science in Electrical and Computer Engineering.

_____

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN

the Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Ayça Kurnaz TÜRKBEN	Department of Software Engineering, Altınbaş University	_____
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altınbaş University	_____
Asst. Prof. Dr. Serdar KARGIN	Department of Biomedical Engineering, Arel University	_____

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.
Submission data of the thesis to Institute of Graduate Studies: ____/____/____

I hereby declare that all information and data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Ahmed Abdulkarim Kado AL-TAHHAN

Signature

ACKNOWLEDGEMENTS

First and foremost, praises and thanks to ALLAH, the Almighty, for His showers of blessings throughout my research work to complete the research successfully.I would like to express my deep and sincere gratitude to my research supervisor, Ass. Prof. Dr. Ayca Kurnaz TÜRKBEN and Head of Department of computer engineering in altinbas university Prof. Dr. Mesut Çevik, for giving me the opportunity to do research and providing invaluable guidance throughout this research. His dynamism, vision, sincerity and motivation have deeply inspired me. He has taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under his guidance. I am extremely grateful for what he has offered me. I would also like to thank him for his friendship, empathy, and great sense of humor. I am extending my heartfelt thanks to his wife, family for their acceptance and patience during the discussion I had with him on research work and thesis preparation.

I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future. I am very much thankful to my wife and my kids for their love, understanding, prayers and continuing support to complete this research work. Also I express my thanks to my sisters. Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

ABSTRACT

BREAST CANCER DETECTION IN X-RAY IMAGES USING FEATURE PYRAMIDS AND CNN ALGORITHM

AL-TAHHAN, Ahmed Abdulkarim Kado

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Dr. Ayca Kurnaz TÜRKBEN

Date: 12/2022

Pages: 73

In the majority of cases, to detect breast cancer, doctors do regular mammograms of patients at risk. This is called screening. When a patient is found to have breast cancer, pathologists perform histopathology. This technique involves looking for breast cancer metastases in the lymphatic system. If metastases are present, more aggressive treatment is required. Histopathology is a difficult task for a human. It appears that in competition, a pathologist makes only 3.5% of false predictions, while in ordinary situations on very small tumours, the error rate can go up to 42%, or almost a one in two chance. To try to improve the quality of cancer detection, researchers have tried to develop algorithms to help pathologists find tumours. They have programmed AIs for this. By giving them images of cells, these algorithms try to find patterns allowing them to affirm the presence or absence of cancer The objective of this work is to develop an approach that identifies patients, through images, that present the disease, that is, the system should be able to classify patients as normal (without disease) and with cancer. For this, CNR are used in a database of thermographic images (infrared) of patients who had the disease as well as patients without the disease. The objective of this project is to respond to the motivation by developing a Matlab Project that will be inspired by the existing solution.

Keywords: CNR, CNN, CAD, MRI.

TABLE OF CONTENTS

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

AI	:	Artificial Intelligence
FSK	:	Frequency Shift Keying
IoT	:	Internet of Things
PAN	:	Personal Area Network
OFDM	:	Orthogonal Frequency Division Multiplexing

1. INTRODUCTION

1.1 BACKGROUND

Early detection of breast cancer significantly increases the patient's chances of being cured and avoids the complications and sequelae caused by the more advanced stages of the disease, which reduces the mortality rate due to this type of cancer. The proposal and development of more effective, accurate and less harmful techniques for patients, for the diagnosis of the disease, especially in its initial stage, is extremely important. With this in mind, the work proposes an intelligent system for the detection of breast cancer, based on the analysis of thermographic images and DL algorithms (RNC), which can be an alternative to conventional diagnostic tests for the disease. In the end, the results indicated whether or not this approach is a candidate to replace traditional exams. The results of the work are of interest to the group of people who are targets of the disease (women), to oncology specialists, in addition to researchers in the area of image processing and DL.

1.1.1 Existing Solutions

The medical images used for this project are so-called WSI images. Bandi P. et al. explain that these are very large images, often exceeding 100,000 X 100,000 pixels. Due to their size they cannot be treated as a standard image.

a. BIGTIFF

Regarding WSIs, several image standards exist. The images processed during this project are in BigTIFF format. Conventional players cannot read such large images, so a special player is required to display them. This reader requires multi-scale, pyramid-structured images, as shown in Figure 2. The IIPImage tool was used to transform the BigTIFF images into zoomable images.

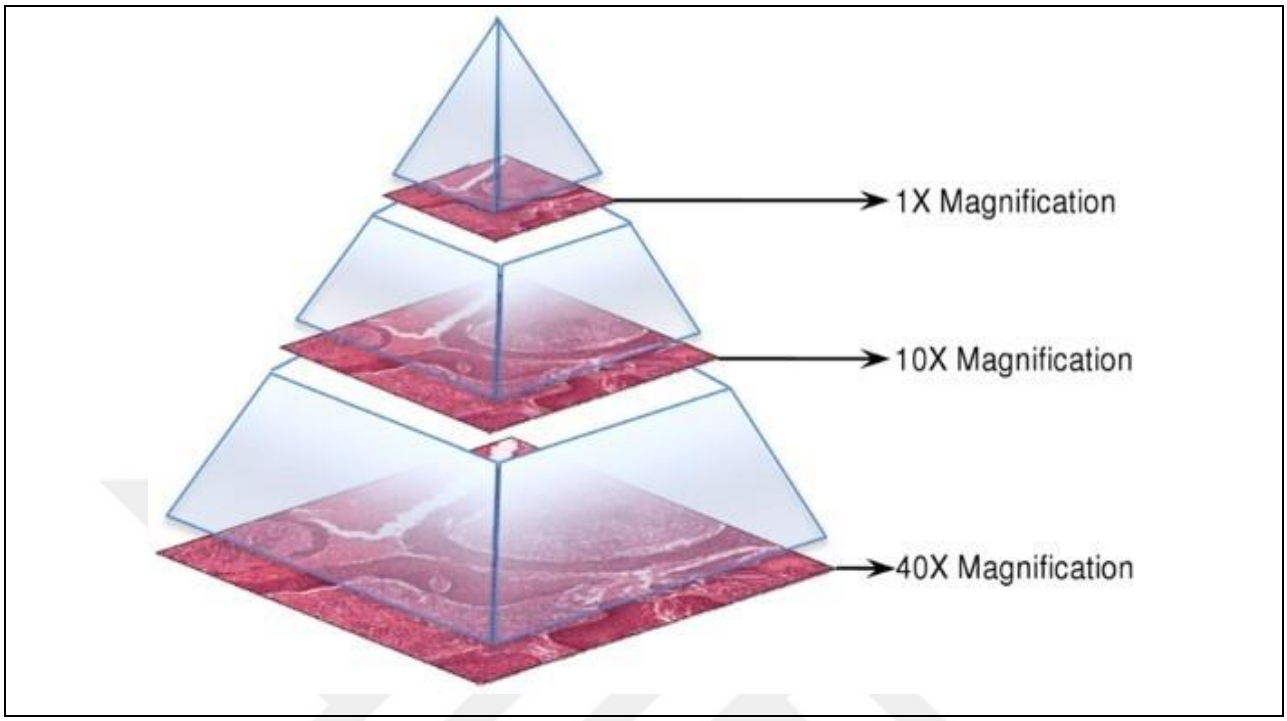

Figure 1.1: Illustration of pyramid image

b. Image bank

The images used in this work are taken from the image bank of the Camelyon17 challenge. It is a scientific competition organized by the Diagnostic Image Analysis Group (DIAG) and the Department of Pathology at Radboud University Medical Center (Radboudumc) in Nijmegen, the Netherlands. The objective of the challenge is to evaluate algorithms for the automatic detection and classification of breast cancer metastases in images of histological sections of lymph nodes.

c. OpenSeadragon

is a web-oriented, open-source zoomable image viewer for high-resolution zoomable images. It is implemented in pure JavaScript.

The principle is simple. The image displayed by this player is in low resolution compared to the original one. It is then possible to zoom in the image. When maximized, only the visible

part is loaded. This means that you never need to load the whole image and thus maintain its quality. This player will be implemented in this project.

1.1.2 DL Algorithms Related to Images

DL is a branch of ML as shown in Figure 1.3.

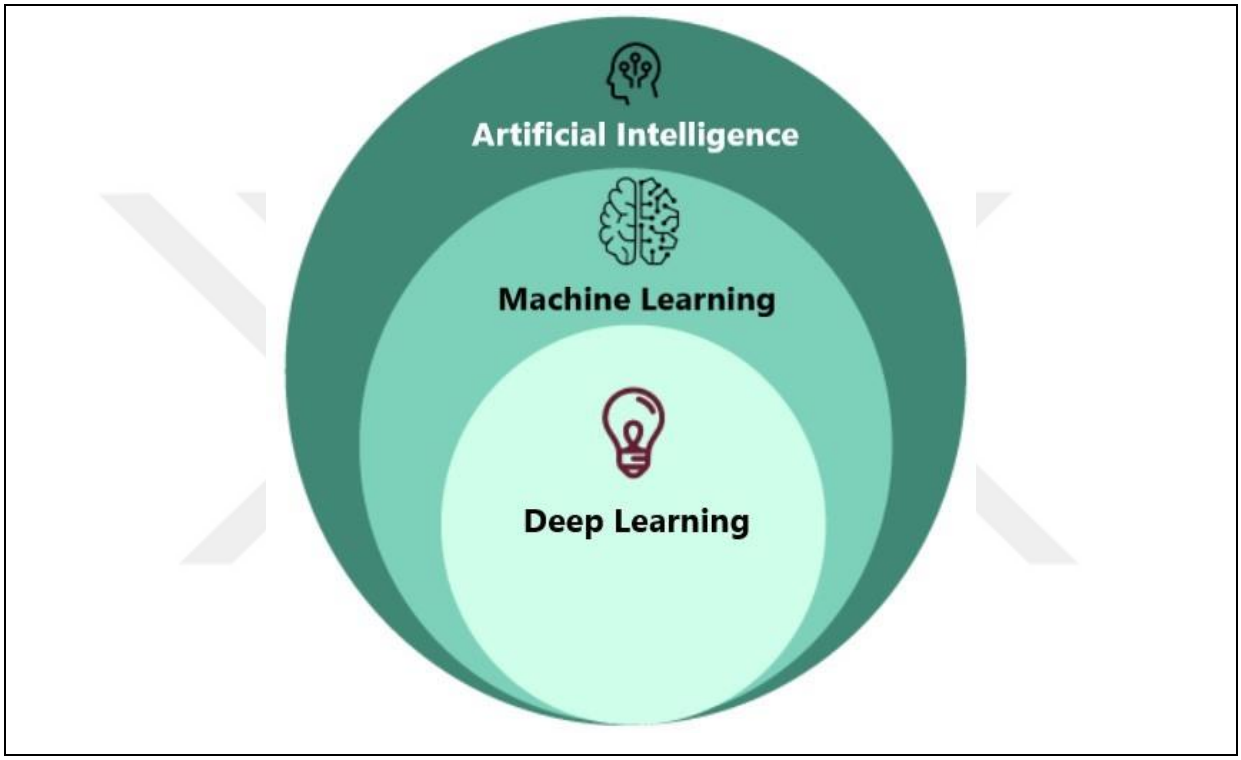

Figure 1.2: Illustration of DL as a branch of ML – Inspired by javapoint

ML is the study of computer algorithms that allow programs to improve automatically by gaining experience

Thanks to ML, a machine is able to learn without having been explicitly programmed for it. DL goes one step further, being able to extract patterns from data using neural networks.

As shown in Figure 1.3, DL algorithms also scale much better than classical ML algorithms with large volumes of data.

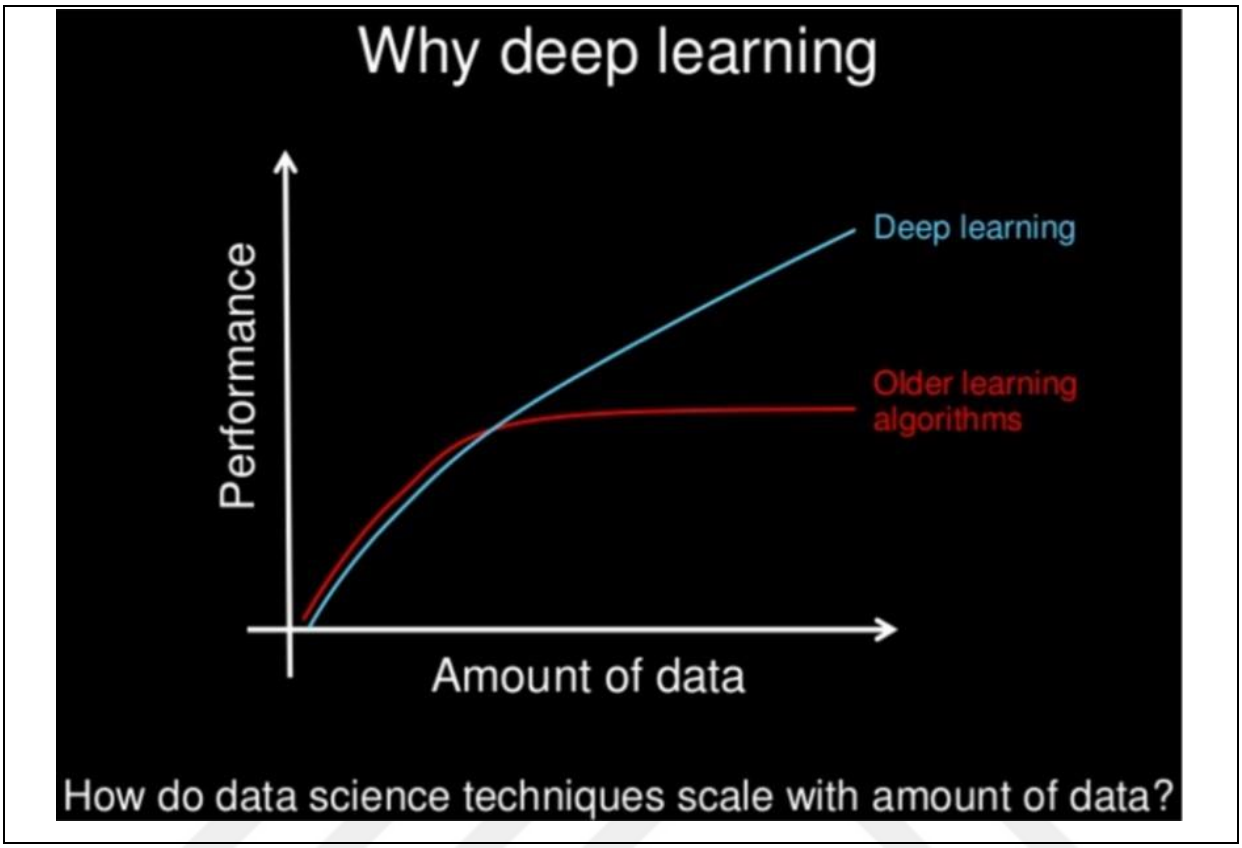

Figure 1.3: Explanation of better performance of DL compared to old ML methods

The DL basically works like a black box. We know the data that enters the algorithm as well as their path, but we do not know exactly what happens there.

1.1.3 CAD For Breast Cancer

In the majority of cases, to detect breast cancer, doctors do regular mammograms of patients at risk. This is called screening. When a patient is found to have breast cancer, pathologists perform histopathology. This technique involves looking for breast cancer metastases in the lymphatic system. If metastases are present, more aggressive treatment is required.

Histopathology is a difficult task for a human. Figure 6 shows the rate of errors made by a pathologist in various situations. It appears that in competition, a pathologist makes only 3.5% of false predictions, while in ordinary situations on very small tumours, the error rate can go up to 42%, or almost a one in two chance. To try to improve the quality of cancer detection,

researchers have tried to develop algorithms to help pathologists find tumours. They have programmed AIs for this. By giving them images of cells, these algorithms try to find patterns allowing them to affirm the presence or absence of cancer. In Figure 1.4,

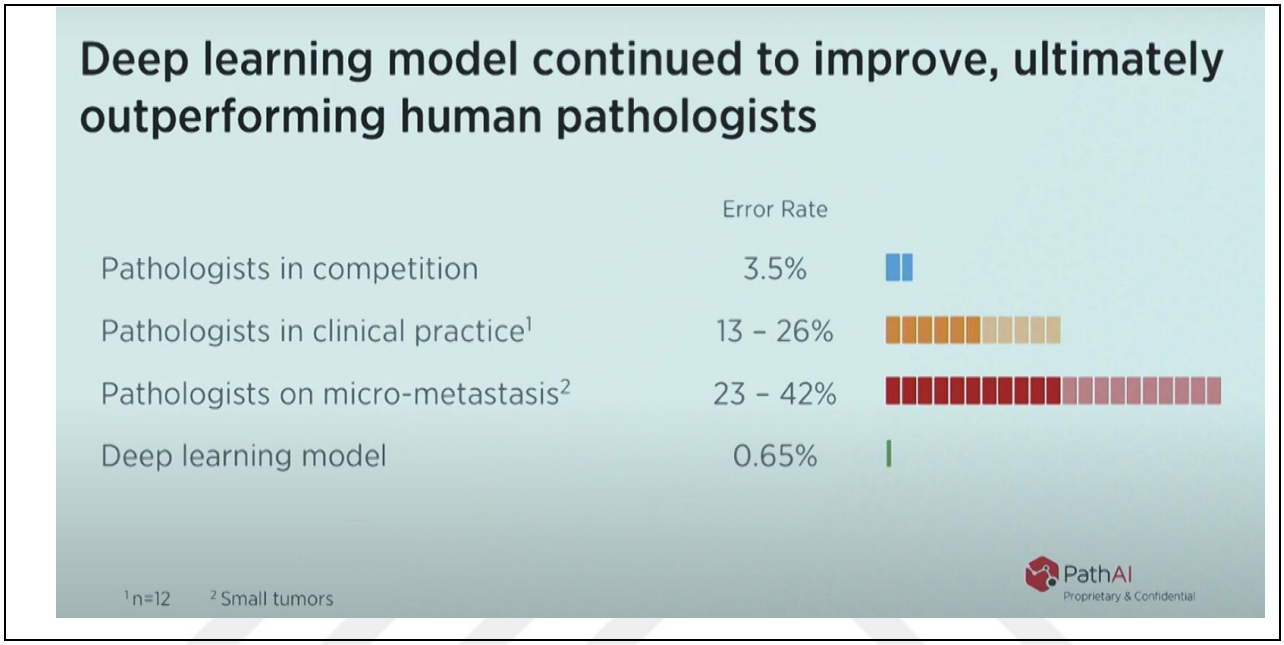

Figure 1.4: Comparison of human versus machine analysis performance

However, this remains only a tool available to pathologists. A computer tool cannot replace a human when making the final decision to apply a treatment. When a pathologist makes a diagnosis, he takes full responsibility for it. If he decides to follow the decision of an AI, he is the one who assumes all the risks.

Dr. Henning Müller explains that currently image-based computer tools are not commonly used. They are tested in many hospitals for specific applications such as mammograms or pulmonary nodules. Regarding cancer detection, most hospitals are not yet equipped to allow the use of computer-assisted solutions. Concerning the Valais hospital, the first scanners for images will be installed this autumn.

1.2 MOTIVATION

Breast cancer is the most common type of cancer in Brazil and worldwide. Among women, it is the most recurrent type and the second leading cause of death in the Americas, in addition to

accounting for about 25% of new cases each year Early detection of breast cancer is essential to increase the chances of treatment and cure effectively, which reduces the mortality rate caused by this disease Mammography is the most common test for detecting breast cancer. This is an X-ray of the breasts made by an X-ray device called a mammography unit. Mammography, however, has some negative points such as problems in detecting the disease in the case of dense breasts (in the case of young patients), exposure to X-rays and discomfort for patients during the exam. Infrared thermography is a technique that measures the infrared radiation emitted by a body/object to determine its temperature through some device, such as an infrared camera. Cancer cells produce marked heat, due to their greater metabolic activity and vascularization compared to healthy cells. This heat reaches the surface through thermal conduction and can be detected by an infrared camera. Breast cancer diagnoses using only thermographic images are not yet accepted by the medical community, being used as a secondary exam. However, infrared thermography applied in the medical field is a fertile field for research, since this technique does not cause harmful damage to the patient in addition to having a low cost. In order to increase the reliability of these diagnoses, several image analysis techniques can and are being used. The techniques that have the best results are those based on Artificial Intelligence (AI). Deep Learning (DL) is a sub-area of AI and Machine Learning, which uses algorithms inspired by the structure and functioning of brain cells (neurons), called Artificial Neural Networks (ANNs), in order to teach the software to recognize patterns in digital representations of images, sounds, and other types of data. Convolutional Neural Networks (CNN) are networks that deal specifically with image analysis. These algorithms have contributed in several areas, including the medical area, as they present great performance and flexibility in environments that are not very restricted.

1.3 OBJECTIVES

The objective of this work is to develop an approach that identifies patients, through images, that present the disease, that is, the system should be able to classify patients as normal (without disease) and with cancer. For this, CNR are used in a database of thermographic images (infrared) of patients who had the disease as well as patients without the disease. The objective

of this project is to respond to the motivation by developing a Matlab Project that will be inspired by the existing solution, it should have the following features:

a. Display Whole Side Images [MRI] images
b. Select an area of the displayed image
c. Choose the displayed image, the desired DL model and the type of explanation sought
d. Show model prediction
e. Show requested explanations

1.4 CONTRIBUTIONS

The methodology used during this work is inspired by the CNN methodology. The latter emphasizes communication and collaboration between development teams and the client. The most important point of the Agile methodology is communication. During the realization of this project, during this, two points are addressed. First of all, the work carried out during the week is presented and the difficulties encountered are discussed if necessary. Once this part is completed, the modifications to be made and the additions desired for the following week are defined. The method applied in this work consists of the following steps:

a. A literature review was carried out on the use of thermographic images for the detection of breast cancer.
b. Study of the state of the art in neural networks.
c. Construction of a database of thermographic images of patients with and without breast cancer.
d. Use of feature pyramid to classify thermographic images on the basis used
e. Experiments were carried out on different CNN architectures considered by the literature and proven to be highly powerful in the task of image classification.

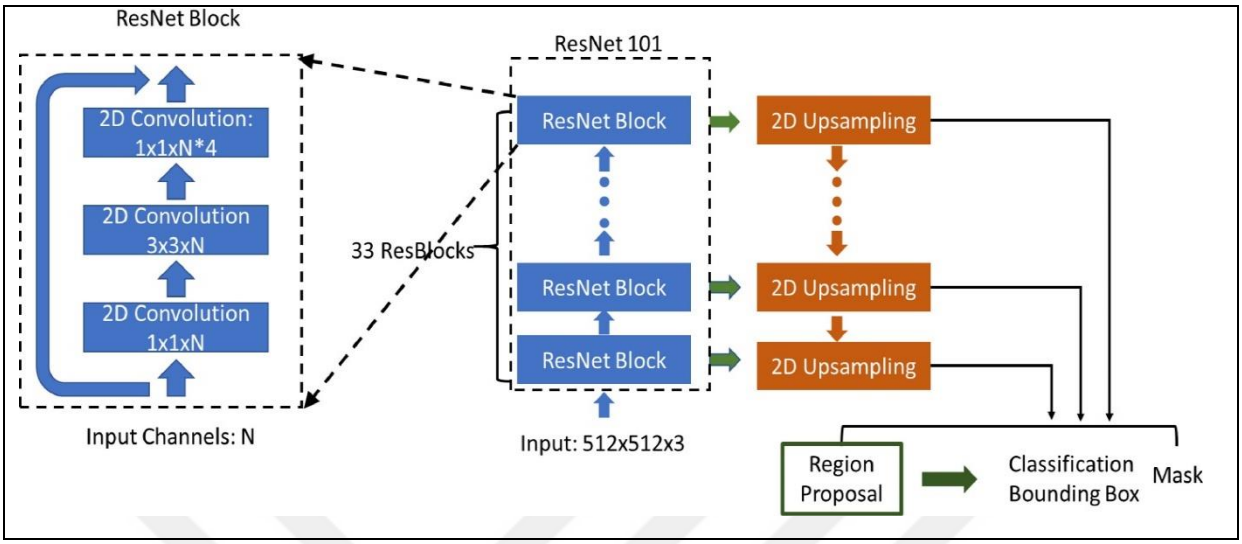

Figure 1.5: Outline of the Proposed network

2. LITERATURE REVIEW

2.1 CHAPTER OUTLINE

Several methods have been applied in the literature to attempt to identify exact and efficient border curves of breast tumors in medical imaging. These models may be classified into three main categories.

Machine learning algorithms make extensive use of the unique characteristics of handmade goods in order to overcome these challenges (or pre-defined features) A functional extraction technique is employed to extract crucial information from the input picture, and then a discriminating model is trained to distinguish malignancies from normal tissues, as shown in the following figure. The development of machine learning systems often relies on manually built features, such as random forests10 and vector support vector machines11-12, which are both used in data mining applications. Extraction algorithms must comb through the processes and functions that have been supplied in order to locate the needed characteristics, edge details, and other relevant information. The success rate of these treatments reduces further when the difference between healthy and malignant tissues is unclear.

2.2 RELATED WORKS

It is possible to combine a large number of normal breast atlases into a single new image mode using the Multi-atlas (MAS) algorithms. Due to the problems associated with registering common breast atlases, as well as the need for a large number of atlases, existing MAS algorithms have proved ineffective in applications demanding speed 3. Deep learning algorithms may be used to determine the most critical aspects of a dataset automatically. These techniques have shown astounding results in a number of areas, including pedestrian detection (15), speech recognition and comprehension (17), and tumor segmentation (18). (18). (19).

According to Heang-Ping Chan, a technique called Convolutional Networks (CN) may be utilized to classify tissues and tissues containing masses obtained from digital mammograms. This technique is referred to as convolutional networks (CN). The extension and enhancement of the study reported here is a primary objective, which is realized via the use of a more complex

architecture and unique methodologies previously used to CNN in other contexts. CNNs are used to change pictures via the application of filters (kernels), and then subsampling neural networks are used to do classification using the convolutional neural network information. As seen in the illustration, conventional sampling entails splitting the image into non-overlapping rectangles and then calculating the average of each of those rectangles. For instance

according to a 2019 research by Jianchao Yang and colleagues [24], it is possible to increase classification performance for a broad variety of objects and contexts by swapping the maximum operation for the average operation during the classification process. Numerous contemporary deep learning models handle classification problems using a multinomial logistic regression neural network, often known as a conventional neural network with softmax activation function.

Yichuan Tang's paper [13] demonstrates that combining a neural network with a support vector machine (SVM) increases performance on a variety of popular datasets, including MNIST (a public database of handwritten figures), CIFAR-10 (10 classes of diverse online pictures), and ICML 2013. In this instance, it is about recognizing facial expressions. The structure of task-modality and task-task structure were both examined in the construction of Zhang et al.21's TSBTS network. When constructing a task modality structure that detects many tumor sites, the pathological implications of the various modality data are considered. By finding the most distinct region of a given tumor section in the surrounding area and then detecting that area inside the tumor using the task modality structure, it is feasible to discover more tumor sites. the creation of a method for articulating critical components of knowledge transfer across a variety of distinct data sources in various modes The researchers were able to ease the flow of information across several modalities by using a generative opposing network (GAN). OM-Net was established in response to Zhou et aluneven .'s breast volume data collection.23 (one-pass multi-task network). OM-Net teaches itself discriminating and joint properties, as well as joint features, using general and task-specific parameters. Through the use of online training and data transfer technologies, the OM-performance Net's is enhanced. The predictions' results are also transferred across tasks with the usage of a guided task (CGA) module (see below). According to Havaei et al.24, the Deep CNN architecture enables the simultaneous extraction of local and global contextual data. They use a simple but successful technique to function extraction in their model. Coupé et al.25 created the AssemblyNet model, a 3D MRI segmentation model for the whole breast, based on

the parliamentary decision-making technique. Through this legislative network, it is possible to resolve previously unresolved situations, render complex verdicts, and reach an agreement on mutually incompatible subjects. The AssemblyNet achieves a majority vote by information exchange with other U-Nets in the vicinity.

2.2.1 Deep Learning Based Methods

This network may be used to overcome the issue of having a limited quantity of training data. Rehman et al. [25] proposed using a Deep Convolutional Neural Network (DCNN) to separate vertebrae in order to identify spinal cancer. A map was used to include the deep network training level into the recommended approach. Additional datasets were used to validate the U-Net design and to gather parametric levels for tumor patients receiving treatment. The framework has a vast number of algorithms at its disposal. Due to its strength and durability, it was used in biological applications requiring both strength and resilience, such as picture segmentation. Wong et al.[26] pioneered the use of finite element analysis (FEA) for biomechanical lumbar breast mechanics computation. The whole lumbar disk body unit was designed using the FEA approach. Additionally to the huge tetrahedral core and cortical shell models, a FEA vertebral body model is included. Additionally, the disk is formed of a nonlinear circular fiber layer of intrusive nucleus. This study demonstrated the feasibility of developing modeling goals for fracture evaluation, pedicle turnout, and bone remodeling. The FEA model was subjected to a variety of loading circumstances, and nonlinear tests were conducted as a consequence of these conditions. To evaluate the correctness of these tests, the results of these tests were compared to data from previously published tests. The development of the FEA model and the load conditions were examined. Khan et al.[27] began their investigation of the lumbar breast's aging effects with a PCFA. Both the human breast and the degenerative pattern that occurs in it were created using major component analysis. To explore the effects of natural aging on the lumbar breast, this approach was used to determine the height of vertebrae and disks, as well as the intensity of paraspinal muscles and subcutaneous fat, and the volume of cerebrospinal fluid. Magnetic resonance imaging of the lumbar breast identified 61 individuals ranging in age from two to ninety-three years with comparable characteristics. Quantitative analysis was performed utilizing two-dimensional visualization and previously acquired information of the vertebral alterations

generated by the PCA transformation. To get a better understanding of the relationships between spinal features and changes in spinal characteristics as a result of aging, factor analysis was used. Gaonkar et al. [28] demonstrated the use of Multi-Parameter Ensembl Learning (MPEL) for segmentation of lumbar vertebral bodies in magnetic resonance imaging (MRI) breast imaging. This approach made use of column clinics to assess patients with a wide variety of scanning features. This demonstrates the resilience of protocol scanning. The strategy, as well as the technology itself, performed well while using a single, manually selected MPE system for autonomous segmentation of vertebral bodies into scans with echo lengths ranging from 24 ms to 147 ms and moderation durations ranging from 1500 ms to 7810 ms. Fallah et al.[29] developed the Hierarchical Conditional Random Forest (HCRF) for volumetric division of vertebral bodies and intervertebral disks utilizing fat water magnetic resonance imaging, and it has received general acceptance thereafter. Classification used a multivariate linear discrimination technique to select acceptable labels for sparse and imbalanced training data using spatial and hierarchical consistency. According to the results of these tests, this approach's prediction ratio is lower than that of identifying the segmented component. Wang et al. [30] argued for the usage of a Deep Siamese Neural Network (DSNN) to identify multi-resolution spinal metastatic lesions. To account for the considerable variations in the size of metastatic lesions that occurred during the trial, a Siamese approach to the deep neural network was used. Analytical sub-networks for the identification of spinal metastases are composed of three identical multi-resolution networks. By using three distinct picture patches for each point of interest, a network input was generated for each of the three locations of interest. It has been observed that by using a weighted average to summarize the findings of Siamese neural networks in local magnetic resonance imaging diagrams, the number of false positives may be minimized (FPs). The data were evaluated and interpreted in 27 cases by using the receiver's free-functioning functionality (FROC).

2.2.2 Metaheuristic Based Methods

provide an explanation for the previously reported deployment of extra heuristic hock transformation and ratification technologies in the spinal study of the Internet of Medical Things (IoMT) platform [2], [31]. To address these issues, the authors propose in this study an ATS-CDM model for the detection and exact segmentation of spinal tumors. The relative politeness distance

between each voxel and the data set features is calculated by comparing its categorization, training, patch labeling, and patch extraction attributes to the data set features. Additionally, it is converted to an absolute location for the purpose of performing exact segmentation of the spinal tumor region using robust ROC modeling, allowing for more accurate tumor identification. When a photograph is segmented, a certain cluster of pixels inside the image is recognized as being connected with a particular subject. When there are several items of interest and it is essential to construct a class, semantic segmentation is required. In medical imaging, a method called segmentation is utilized to detect anatomical features such as tumors and organs, allowing for the collection of quantitative data for clinical research, diagnosis, and surgery. Segmentation is a vital element in developing any computer support system, as it has a direct impact on the system's overall performance. The border between anatomical structures is a critical piece of information that cannot be overlooked while classifying anatomical structures. The establishment of the boundary, which may be difficult to perceive with the naked eye, is one of the most difficult difficulties in segmentation. Numerous strategies have been reported in the literature for recognizing and segmenting bone structures in MRI data. CVs and DIVs are used extensively in these research. It is quite challenging to locate validation for the thirty scoliosis photos included in this research. This knowledge gap is primarily the result of the difficulties in obtaining data on a particular disease. There are just a few methods available for segmenting both structures concurrently. Due to the two structures' rectangular form, they are easier to detect in the sagittal plane, where the first method focuses its efforts. 3D segmentation has become a reality as a consequence of the advancement of traditional approaches mixed with deep learning. Automated and semi-automatic techniques for DIV segmentation are available in the published literature. Graph segmentation is a semi-automatic technique for segmenting data. Ben Ayed et al. (2011) proposed an interactive technique to segment DIVs using a graphical user interface by using the graph cut approach (Boykov & Funka-Lea, 2006). The technique attempts to optimize a cost function by using previous knowledge on the intensity and shape of the DIVs in the picture. The authors were especially interested in the effect that prior information had on their technique throughout their inquiry. To have access to this data, the user needs first create an ellipse for each DIV. According to the research, the method taken to the item had a significant impact on the final decision. Additionally, we encounter techniques that include the usage of an atlas. Michopoulou

et al. describe a semi-automatic segmentation of healthy and degenerative lumbar IVDs using MRI in their work (2009). An atlas of a DIV was constructed by recording landmarks in a series of locations centered around DIVs and then using those locations to generate the atlas. The authors provided a priori knowledge of IVD anatomy by using three distinct segmentation algorithms based on the atlas. Each pixel's intensity is utilized to classify it, which is done via the use of fuzzy grouping. The rate of each tissue type inside a voxel is calculated by examining it using the FCM approach devised by Bezdek (Hathaway & Bezdek, 1988). The Pharm-RFCM algorithm is used in lieu of Bezdek's FCM algorithm in the second procedure (Cannon et al., 1986). The third method is relatively similar to the second, except that when the atlas is registered in the picture, it is given an elastic deformation. Comparative tests revealed that the third approach, which was followed by the second, was the most stable and dependable of the three. While this is a time-consuming job, it is well worth the effort. Additionally, because of the high computational cost of this strategy, the author opts for the second approach, which sacrifices segmentation performance for a more manageable execution time, in order to get the desired results (see Figure 1). Before segmentation can begin, the atlas must be prepared manually, and this stage is critical to the final product's quality. As a result, using this strategy in a clinic environment would be very challenging. Additionally, techniques of segmentation based on geographical areas are feasible. Segmentation of IVDs should be addressed, according to Law et al. (2013), in order to minimize human engagement. To estimate the vertebrae's position, an anisotropic directed flow representation is used, which necessitates manual labeling of the first cervical vertebra to verify that the vertebrae are accurately recognized. Utilizing the DIV descriptors generated in the preceding phase, a maximum directional distance transform is conducted to provide an energy function that can be decreased using an active contour. Despite the fact that MRI requires little human intervention, the noise present in the image is apparent with this approach. According to the findings, the ligaments that surround the DIVs in the test participants were segmented. On the one hand, the use of classifiers and manually selected and extracted attributes are instances of automated procedures. Chevrefils et al. (2009a) suggested utilizing MRI to segment IVDs for healthy and scoliosis patients, respectively. The watersheds technique is used for the most basic division of DIVs. To fine-tune the output, a KNN classifier is used, which is taught the classification rules using a vector of statistical and spectral texture information. The authors

performed a comparative study in their experimental research, comparing and contrasting a number of textural characteristics.

2.2.3 IVD Based Methods

The results indicated that by including statistical and spectral texture information into the feature vector, overall performance was greatly enhanced. Despite the excellent accuracy, the recall rate for validation images is poor, according to the study's authors. They explain this by stating that due to the method's proclivity for over-segmentation, the authors have reservations about its capacity to accurately depict the outlines of IVDs. Oktay and Akgul (2013) predicted that an IVD detection technology will be developed in the future. The HOGs properties are used to train an SVM classifier, which is subsequently assisted in detection by a probabilistic graphical model. Numerous people have expressed their happiness with the outcome. The HOGs features of an item determine its overall look (DIVs in this study). The form and structure of a healthy DIV are almost similar. As a consequence, they may be classified and classified using HOG features. On the other hand, the research did not evaluate IVDs with deformations (by scoliosis for example). Segmentation is accomplished in a variety of methods, depending on the geographic location. Chen et al. described techniques for localizing and segmenting intravascular devices (IVDs). After training, the images are localized by developing a data-driven regression model that estimates the displacement vectors of numerous randomly picked points in the center of each DIV. Along with the same detection strategy, a new algorithm substitutes the regression model by classifying each individual pixel as a component of the DIV or an inconspicuous component of the surrounding background. When it comes to segmenting DIVs, the authors first demonstrated that their approach outperforms a decision tree forest algorithm in an experimental evaluation. On the other hand, there have been some documented issues. During the detecting phase, there is a significant effect on the subsequent segmentation. The latter is more obvious in the volume's core slices, where the trunk may be seen clearly, due to the trunk's prominence in those slices. This method may be used to detect spinal abnormalities, whether they are caused by illness or are just more difficult to identify because the breast is more difficult to view. According to Huang et al., there are three processes involved in the identification and segmentation of CVs (2009). The first step is to train a CV detector using the adaBoost algorithm (Freund et al., 1999). To choose the

properties of the training images, the wavelet representation of the image and the image's intensity are used in combination. The detector output is refined post-processing using a curve fitting approach developed by Vrtovec et al. (2005). It is hoped that by using this curve, false positives would be decreased and any previously unknown CVs will be detected. Finally, after the CVs have been discovered, they are segmented using the normalized cutting approach (Shi & Malik, 1997). As shown by our testing, post-processing is vital for the detection of CVs. The same identification and segmentation approach reported by Kelm and colleagues may be used to identify IVDs as well as the cardiovascular system (CVS) (2013). The detection approach is implemented in line with Zheng et alintroduction .'s of marginal space learning (2008). The graph cut approach is then used to determine the graph data set's anatomical structural contours, and the resulting graph is shown. The flexibility of the research in terms of medical image structure and modality makes a major addition to the area of study. As a consequence of these findings, the technique has been shown to be resistant to IVD degeneration. Despite the graph cut approach's apparent robustness in terms of following outlines, qualitative assessments clearly indicate overlaps between the two topologies. The approach fails due to its inability to segment both structures concurrently.

Rather than genuine segmentation, we observe a near detection (ie more precise than a centered window) with this technique. DIVs and CVs may be segmented simultaneously by Neubert et al. (2012). A Canny filter and an intensity profile are used to locate the vertebrae's centers. The segmentation is then carried out by placing models of average form on these spots. The approach's performance was found to be excellent, as shown by the findings. However, it is believed that it takes 35 minutes to segment a single vertebra. Because the identification of vertebrae isn't particularly reliable, researchers have limited their testing to the mid-slices of the volume where IVDs and CVs are clearly visible. The DIVs and CVs take on a distinct appearance on the slices where the breast becomes more apparent and on those where it is completely obscured. Detection and, thus, segmentation would be a failure in this circumstance. According to reports, DIV segmentation is also claimed to have an issue with under-segmentation of the breast on T2-weighted images. Traditional segmentation techniques, like the ones discussed above, have significant drawbacks that make them unsuitable for use in the clinical setting. Execution time or human involvement is required for a successful strategy. In spite of the fact that this method does

not need any kind of training, it is very susceptible to noise in the images. Because segmentation performance is so tightly tied to detection in systems needing a learning mechanism, this step is virtually a requirement in and of itself. To make matters worse, even narrowing down the features to be extracted is a challenge since doing so limits the applicability of the technique to just a subset of possible photos. When anatomical structures are distorted, however, this procedure cannot be used. Sophisticated semantic segmentation challenges can be solved using deep learning approaches since their success in picture classification has allowed them to be used to address more difficult tasks like regression or discriminating. While unusual, it is feasible that the first scenario may occur at some point. Indeed, Suzani et al. (2015) proposed a technique for detecting, localizing, and segmenting CVs using this approach. There are two stages to the strategy. A deep neural network is used to tackle the regression issue of localization. It is taught to determine the distance between each CV's center and each voxel in the same manner that Chen et al. used (2015). A statistical model is registered to do the segmentation. It was shown that, in contrast to Chen et al. (2015), neural networks could identify CVs across the whole volume range, not only the center. Nevertheless, the qualitative findings suggest an over-segmentation of the CVs, with certain portions of the resumes being seen as the image's backdrop. Discrimination is a typical way to solve a segmentation issue. CNN-based segmentation methods used pixel-by-pixel classification, in which the value of neighboring pixels was used to determine the categorization of a given pixel. The whole picture is filtered via a sliding window centered on one pixel. An picture's pixels are categorised one at a time, with the process being repeated until the whole image has been processed. In the 2012 Electron microscopy segmentation challenge at the IEEE International Symposium on Biomedical Imaging (ISBI), Ciresan et al. (2012) applied this approach and came in first place. This sliding window technique, even though CNNs have surpassed other methods to segmentation jobs and it seems to be straightforward, has certain drawbacks. In order to save time and resources, the windows are usually quite narrow compared to the size of the input pictures, which limits the amount of information the classifier can use. In addition, the procedure is hampered by unnecessary computations, which causes the inference to take much longer than it should.

2.2.4 CNN Based Methods

Similarly, Jianhua et al. (2016) employed the similar method to segregate IVDs from an MRI. However, in this investigation as well, no validation on MRIs of people with malformed breasts was done. To accomplish segmentation, CNNs evolved to use convolution and dot products interchangeably, as well as to use convolution layers to replace fully linked layers. It was Long et al. (2015) that proposed this notion for the first time by providing what would serve as the foundation for the most current CNN designs for semantic segmentation: fully convolutional networks (FCNs). After Lecun et al. (1998) sketched out the broad strokes of a pyramidal architecture, CNNs followed suit, with the spatial resolution of images decreasing across layers and the depth increasing until it reaches completely connected layers, which are linked to the layer output. Oversampling techniques reduce the depth of a pyramidal structure and increase its spatial resolution by removing the entirely linked layers, which are replaced with a pyramidal structure with reduced depth and increased spatial resolution. The output layer's dimensions match those of the image's input layer. Similarly, each channel corresponds to the segmentation mask of the class linked with it in terms of depth and number of classes. Using completely linked layers for segmentation is not a good idea for two reasons: they compel all input pictures to be the same size, and they cause a significant loss of local information in segmentation. 3D FCN segmentation of CVs from MRI was suggested by Korez et al. (2016) and used a 3D FCN trained to build a probability map of volume CVs. The probability map used to direct a deformable model to the CVs' contours provides the basis for the post-processing. Deep learning-based segmentation of DIVs in MRI has been proposed by Chen et al. (2016) utilizing a 3D FCN architecture. The authors conducted an experiment to find out whether segmenting on 2D or 3D photos is more effective. FCN 3D's use of volumetric contextual information resulted in outcomes that were closer to those achieved by FCN 3D. Pyramid representation of the original FCN presented by Long et al. (2015) finishes with just one up-sampling operation and the output layer. Precision in localisation and categorization was much improved as a result of this. But a major issue developed, namely, the poor quality of the produced photographs A number of novel forms of FCN have been suggested, however we will only be interested in encoder-decoders because of our methodological approach. In 2015, Ronneberger et al. announced an NCF-based architecture for biomedical data

segmentation that has now become a standard in the area. An encoder/decoder architecture is known as the u-net. In the 2015 Electron microscopy segmentation challenge at the IEEE International Symposium on Biomedical Imaging, the u-net surpassed all other approaches. (ISBI)).

3. MATERIALS AND METHODS

This chapter presents the theoretical foundation on computer vision and neural networks, developing the basic concepts used in the work.

3.1 COMPUTER VISION

The area of computer vision aims to generate algorithms capable of processing, analyzing and identifying images in a similar way to the process performed by human beings.

Computer vision work goes beyond image processing, aiming not only to modify the characteristics of the image, but also to understand its content. The tasks performed involve the interpretation of the signal, acquiring and processing relevant information, thus producing a high-level characterization of the original image.

The main tasks of interest performed in the area of computer vision can be divided into:

Classification: categorizing the image into a set of possible known classes.

Object detection: find object instances in an image. In addition to classifying objects, it is also necessary to locate them within the image.

Semantic segmentation: classifying the pixels of an image individually among the set of possible known classes.

Semantic segmentation of instances: semantic segmentation that separates each occurrence of objects within a class. Each instance of a class is independently segmented.

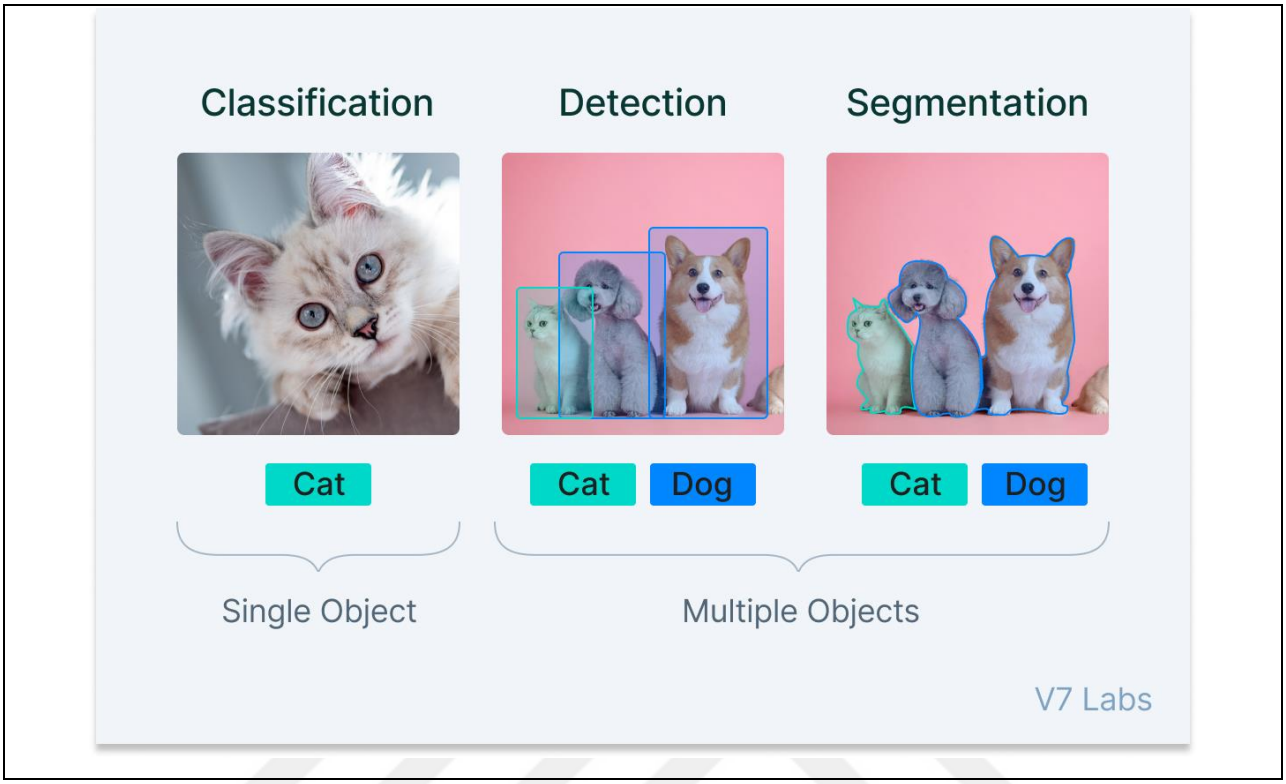

Figure 3.1: Basic types of tasks performed in the area of computer vision. a) Classification of the image as a whole within a set of classes. b) Detection of objects between different classes

Each object is individually approximated by a rectangular region of the image. c) Segmentation of all objects of a class, done pixel by pixel for the entire image. d) Segmentation of objects individually, so that the segmentation separates the various occurrences of objects.

These techniques can be applied in several areas, generating a system capable of interpreting the real world and allowing the interaction between computers and the environment. Some of the application areas of computer vision are:

Automation (machine vision, industrial control, product and merchandise inspection).

Navigation (autonomous cars, accident prevention).

Identification of people (traffic control, surveillance).

Automatic reading of texts and signs.

Medical diagnosis (anomaly detection, exam classification, automated screening).

Each of these tasks consists of interpreting the information contained in the original image in a useful way. This requires that the original variable (the image) be condensed into a higher-level

variable. This new variable will be a vector with a much smaller dimension than the image, but which is representative of its main characteristics.
This condensation of information can be performed using descriptors, which are functions specifically designed to obtain variables that characterize the original image.

3.2 DESCRIPTORS

One of the ways to implement computer vision methods is through the use of descriptors. This technique allows generating a set of variables that describe one or more aspects of the original image in a simplified way.

The use of descriptors has been quite common in recent decades, mainly due to the computational efficiency of the algorithms. The descriptor project generated very interesting results in the area of computer vision, such as face detection [22] and object classification.

Descriptors are image processing algorithms that extract information such as color, texture, shape, color momentum, histogram, etc. from an image. The success of this technique was driven by the development of more efficient descriptors, such as HOG (Histogram of Oriented Gradients) [24], SIFT (Scale-Invariant Feature Transform) [25] and ORB (Oriented FAST and Rotated BRIEF) [26], which allow you to efficiently obtain more complex information about the image.

By obtaining the appropriate descriptors, it is possible to make decisions about the image (classification, detection, segmentation, registration, etc).

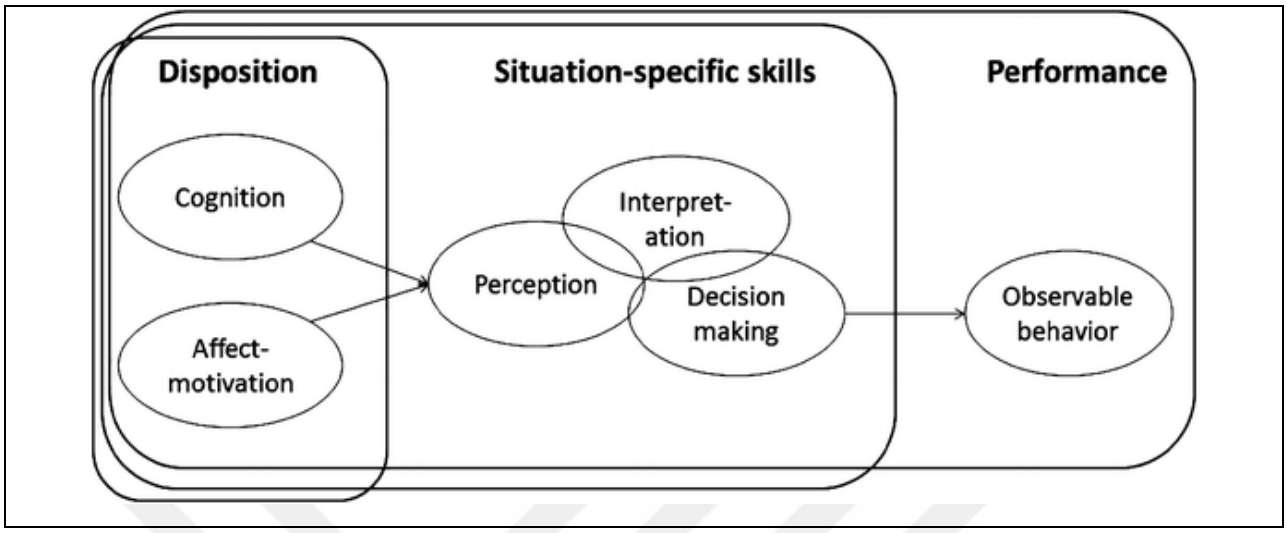

Figure 3.2: Image interpretation (decision making) is based on the descriptors obtained

For the interpretation to be done correctly, it is necessary that the descriptors used reflect the characteristics of interest of the image. Multiple descriptors can be used, forming a vector of variables that represents the original image.

As the most relevant information set of an image is application dependent, the descriptors need to be chosen or adapted specifically for each task. In general, descriptors are defined based on the characteristics of the original image and developer heuristics. Thus, the main difficulty in using descriptors lies in the complexity of developing specific descriptors for each task to be performed.

The problem related to the development and choice of specific descriptors can be solved with the use of an adaptable tool, capable of learning the desired function from experience and analysis of examples. This type of method belongs to the domain of machine learning (machine learning), from which the techniques of neural networks and convolutional networks arise, capable of solving several problems in the area of computer vision.

3.3 NEURAL NETWORKS

Artificial neural networks are computational models with a biological inspiration based on the human brain, capable of performing complex functions by combining a large number of basic elements.

This structure is assembled from the combination of multiple nodes, where each node performs a simple mathematical function. The end result is an extremely parameterizable model, capable of shaping and learn to perform specific tasks from the appropriate adjustment of parameters.

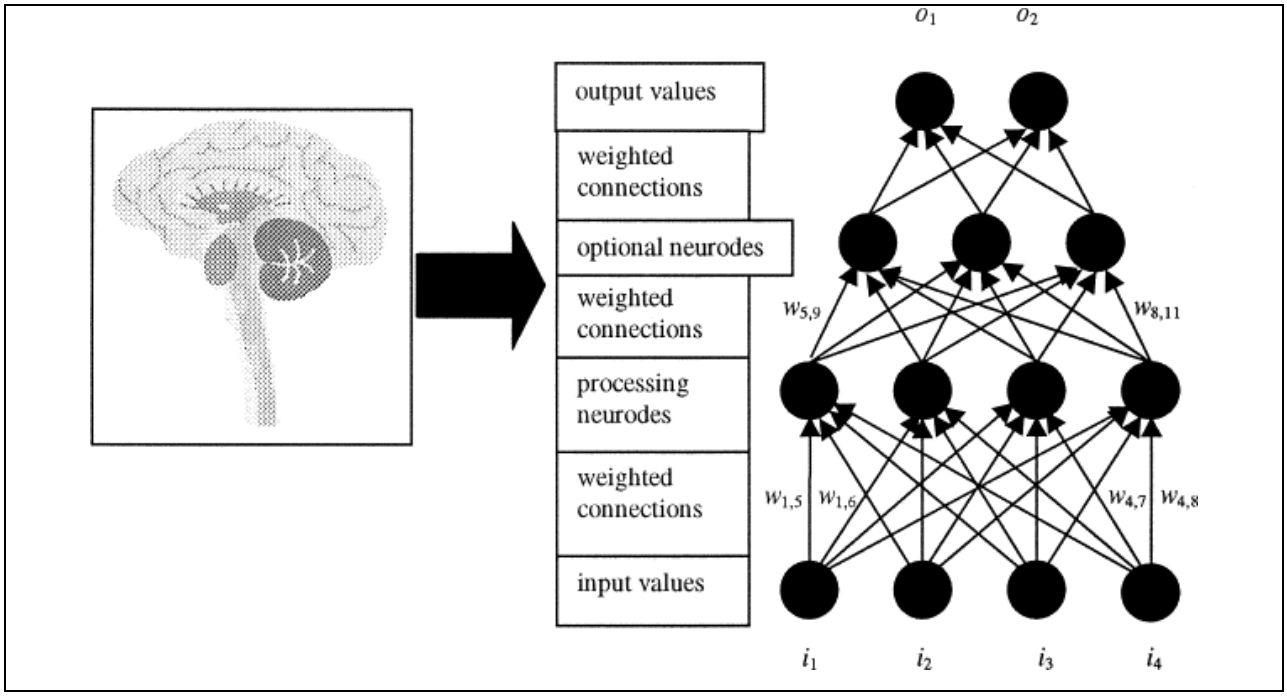

Figure 3.3: Neural networks function performed by the descriptors

The following procedures, such as classification and detection, can be kept the same or eventually absorbed into the neural network itself.

The operation performed by each node of a neural network is basically the junction of a linear combination and a non-linear activation function. The simplicity of a node allows it only to perform functions of low complexity, but a sufficiently large set of nodes is capable of approximating any mathematical function [27, 28].

Figure 3.4 is a schematic representation of a node in a neural network. The input x passes through a linear combination defined by the weight vector W, producing the scalar akj. The final output zjk of the node is the value akj after passing through an activation function g(a).

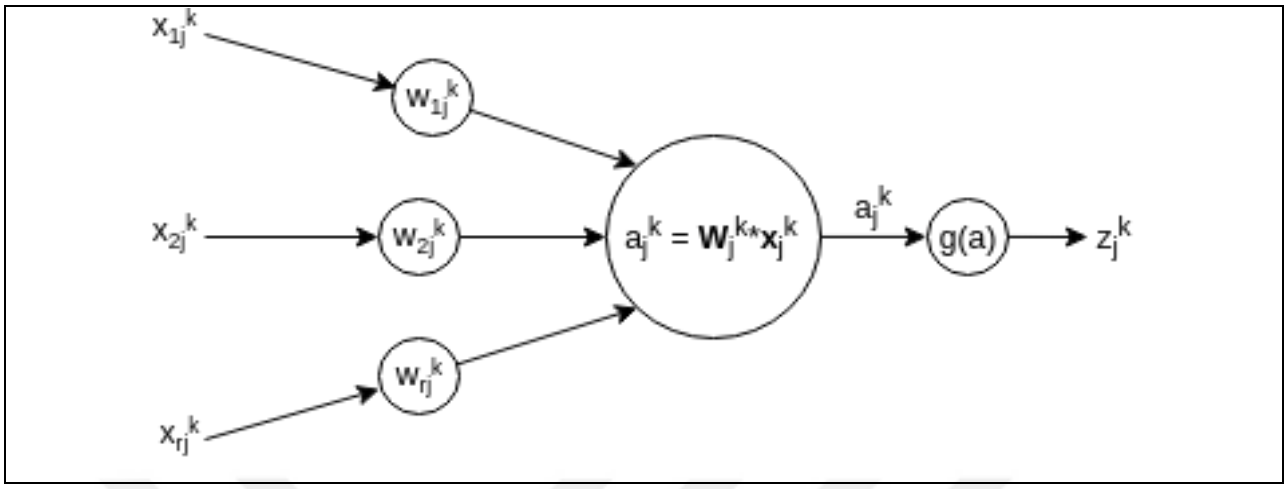

Figure 3.4: Basic unit of a neural network

Representation of node j, belonging to layer k and with an input vector with r components. Note that it is common to have a bias variable in the linear combination, so that akj = Wjkxkj + bkj, however this can be simplified by assuming that one of the components of the input vector x is constant The activation function g(a) is responsible for adding a non-linear component to the behavior of the node. Without this activation function, all nodes would only perform linear functions that are not capable of producing more complex approximations.

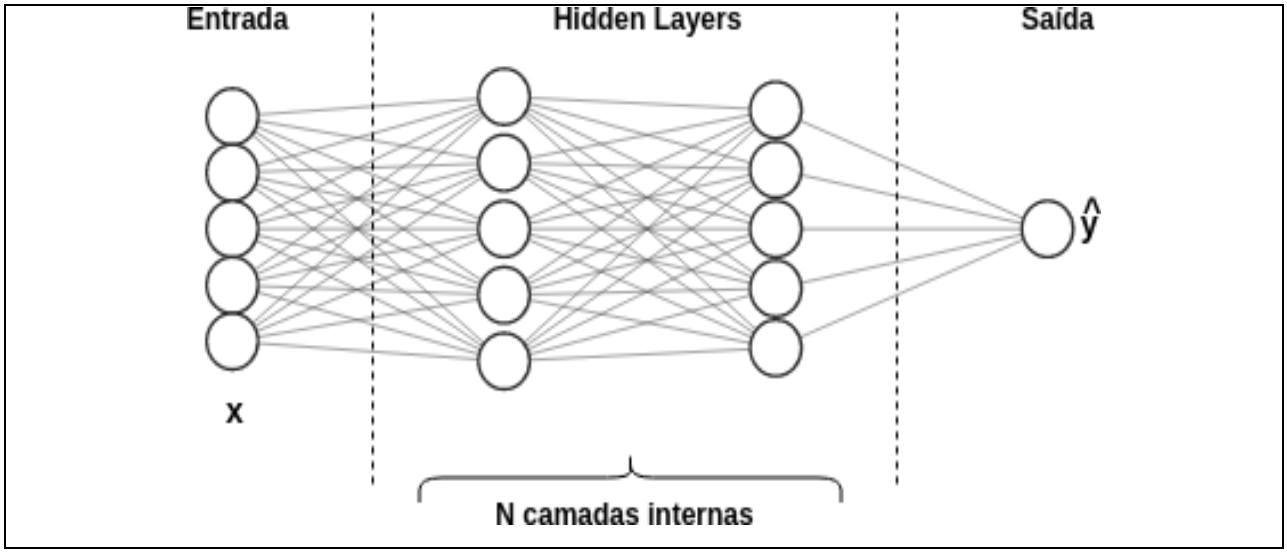

Figure 3.5: Example of a basic neural network (Fully Connected Neural Network) with input, output and inner layers.

The output y is the result of the network, obtained by the composition of operations of each node in relation to the input vector x Figure 3.5 shows an example of a neural network, a structure known as a Fully Connected Neural Network. This network is assembled from the combination of nodes in parallel (several units in the same layer) and in series. (several layers operating in sequence). The structural characteristics of the network, such as the number of nodes, number of layers, activation functions, etc., are known as network hyper-parameters.

Regardless of the specific structure, the objective is that the function f̂(x) performed by the neural network approximates any function f(x) with the smallest possible error.

The quality of the approximation performed by the neural network depends on the adjustment of the internal parameters of each node. The main method of adjusting these parameters is formed by an iterative optimization process, where the weights are modified from the analysis of multiple reference points of the original f(x) function. From the observation of each of these samples, the network is optimized in order to minimize the difference between the ŷ and y values, using algorithms such as gradient descent and backpropagation.

3.4 DESCENDING GRADIENT AND BACKPROPAGATION

Backpropagation is the main method used to train neural networks. This algorithm aims to optimize the initial weights of a neural network, minimizing a loss function that characterizes the quality of the current predictions of the network.

This technique is an efficient way to implement gradient descent optimization in multilayer neural networks [29], using the chain rule to propagate the gradient value to all network elements. Without the use of backpropagation, training neural networks with multiple layers becomes computationally expensive, preventing the construction of larger structures and therefore limiting their ability to represent more complex functions.

3.4.1 Descending Gradient

Assuming a neural network with weights W, we obtain the output ŷ from the propagation of the input signal x through the network. The quality of the forecast is measured from the error function

L(y,yˆ), which measures a characteristic distance (such as root mean square error, cross entropy, etc.) between the expected value y and the predicted value yˆ.

The main difficulty in using this algorithm is in calculating the partial derivative of the error with respect to each weight of the network. Backpropagation emerges as an efficient method to calculate the required gradients within a multilayer neural network.

3.4.2 Backpropagation

The backpropagation algorithm is the method used to efficiently propagate the value of ∂L ——— $\partial w(y,ky\hat{})$to the inner layers of the neural network [30]. This method allows calculating the gradient of the L layer from the gradient calculated for the L + 1 layer, using the structural characteristics of the network and the chain rule for differentiation. The gradient is therefore propagated sequentially from the layers closest to the output towards the input.

The functioning of this algorithm strongly depends on the structure of the neural network used, so that changes in the architecture and activation functions alter the efficiency of the optimization process. In addition to being a way to implement gradient descent in neural networks, backpropagation can be used in conjunction with other methods, such as normalization, momentum, batch learning, etc.

The backpropagation equations obtain a very simplified and recursive form, which allows the use of different methods to optimize the computational implementation of the algorithm [29]. The main factor for the high performance training of neural networks is the use of graphics processing boards (GPUs), which allow solving both the direct equations (feedforward) and the backpropagation equations in a much more efficient way.

3.5 CNN NETWORKS

One of the main disadvantages of fully connected neural networks is the difficulty of working with signals with coherence, such as audio signals (sequential coherence, in time) and images (spatial coherence, in 2D). This type of neural network does not take into account the relationship between the parts of the input signal, losing very precious information about the variables.

A convolutional neural network (CNN) is a type of network that allows you to process images and still maintain the spatial coherence of the signal. This is accomplished by using convolutional filters within their base units, replacing conventional linear blending with an operation more appropriate for use in imaging.

3.5.1 Convolutional Filter

Each layer of a convolutional network is composed of a set of convolutional filters, where each filter operates on the output of the previous layer.

Similar to fully connected neural networks, each filter in a convolutional network performs a simple mathematical function, and the total set of filters results in a more complex function This distinction is not very important, since the only modification

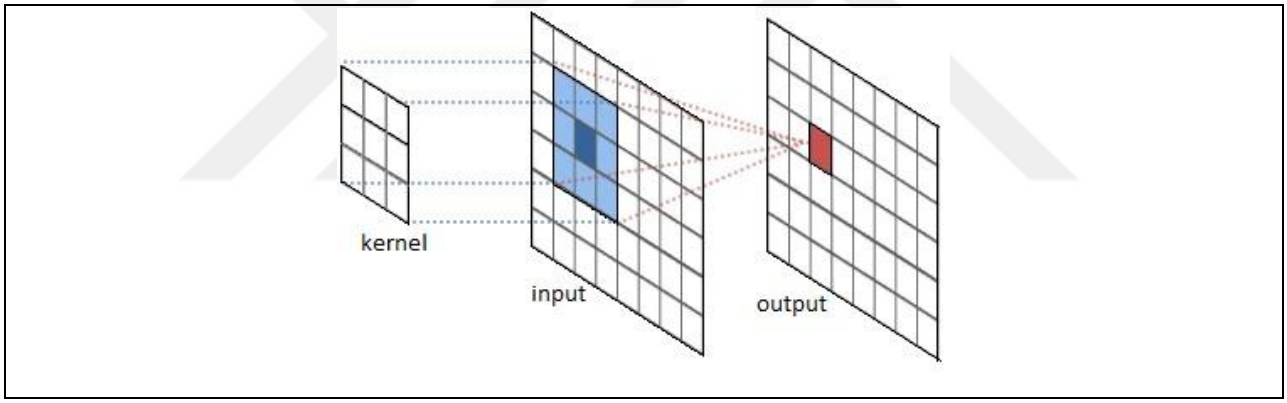

Figure 3.6: Visualization of the convolution operation [1] on an image (input).

Each point of the resulting activation (output) corresponds to the application of the kernel over a different region of the image. Applying the filter across the entire image results in a map of activations is the inversion or not of the kernel before its application to the image. As filter weights are not fixed parameters, but values that will be learned during training, this distinction will be irrelevant after network optimization.

The convolution operation can be visualized in Figure 3.6. Three important properties distinguish convolutional networks from traditional neural networks [29] and allow working with images more efficiently: Sparse interactions: Unlike traditional neural networks, convolutional networks work

with sparse connections between the input and output of a layer. A unit at the output of a convolutional layer only interacts with a region of size W × H (size of the kernel used) of the original image, known as the receptive field. This feature allows for more efficient data processing and makes it easier to obtain local information (around a region of the image).

Parameter sharing: during convolution, the same kernel is applied over the entire input signal sequentially so the function parameters are used multiple times. This is efficient for image processing, as a pattern of interest can occur independently at any position in the image, and therefore kernel reuse allows for the optimization of a single function. For example, the detection of a texture does not depend on its position in the image, but on the local characteristics, so that the same filter can be applied to the entire signal.

Translational invariance: the convolution operation is translation invariant, forming an activation that is able to map the occurrence of a feature over the entire length of the input signal. This is very interesting for object detection, as the translation of the object in the original image will result in an equivalent translation of activations. It should be noted that convolutional networks are not invariant to rotation, scale changes and other geometric transformations, so different techniques must be used to solve these problems.

A basic convolutional network is structured in the same way as an ordinary neural network, from the concatenation of filters in a series of layers. Each layer contains a known number of filters, each of which operates on the output of the previous layer.

Training a convolutional network is also done using backpropagation. The principle of using the algorithm is the same, for all lattice weight The main difference is in the calculation of δk as a function of $\delta k+1$, because here the calculation of akj is done from the convolution operation.

A convolutional network is defined by several hyper-parameters, which describe the general format of its construction. For example, the number of layers, number of filters per layer and total number of parameters are basic hyper-parameters, generally used as a criterion for comparing different structures.

There are several case studies of convolutional networks that have become the basis for the development of future structures. These networks are recognized for presenting innovative techniques that have advanced research in the area.

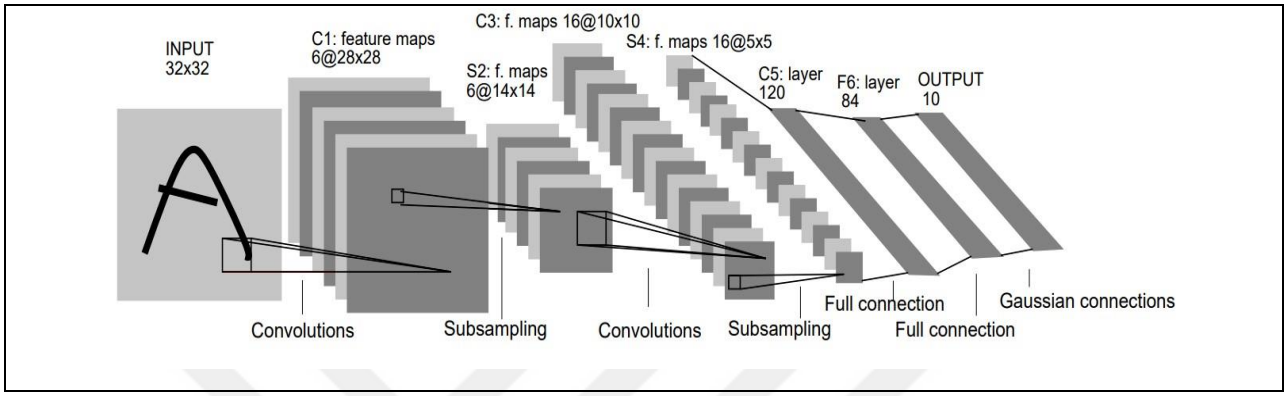

Figure 3.7: One of the first successful applications of Convolutional Networks, made in 1998 [2].

The LeNet network was developed to perform character recognition using convolutional filters, showing a successful alternative to the use of descriptors.

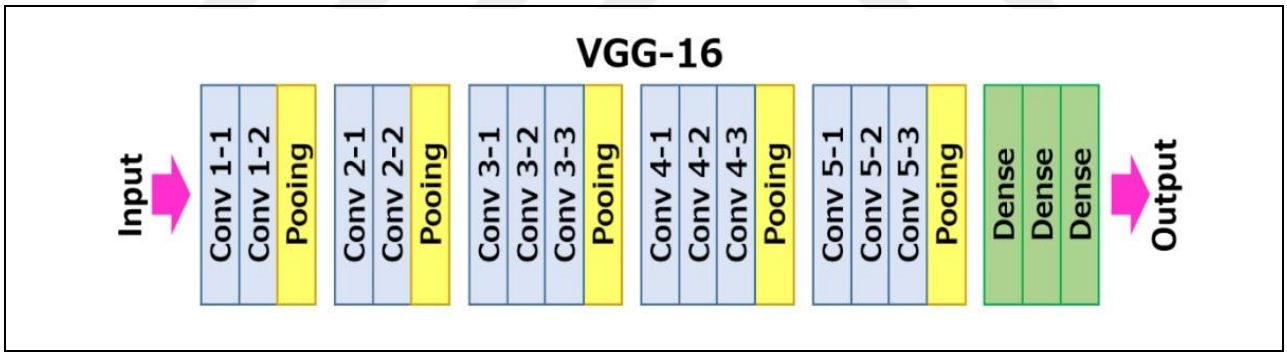

Figure 3.8: The VGG16 network [3] was presented at the ILSVRC2014 competition, by the Visual Geometry Group at the University of Oxford

This network became the basis for the development of several other structures, popularizing the use of multiple 3x3 filters in sequence in place of larger filters, allowing the development of deeper networks with fewer parameters.

3.5.2 RETINANET

RETINANET is a one-stage detection network, which seeks to achieve the performance of two-stage networks while still keeping computational complexity down

This structure shares several characteristics with other one-stage networks (anchors, Pyramid Feature Network), but implements an innovative loss function capable of solving the class imbalance problem.

The performance of the RetinaNet structure is even superior to that found in other two-stage detection networks. The execution time, as shown in Figure 3.1, is shorter than that of other one-stage networks for similar precision values.

The main contribution of the RetinaNet network is in the development of the Focal Loss function [7]. This function aims to reduce the weight of the unbalance between the foreground and background anchors in the training of the network, introducing an exponential factor that reduces the loss of already well classified anchors.

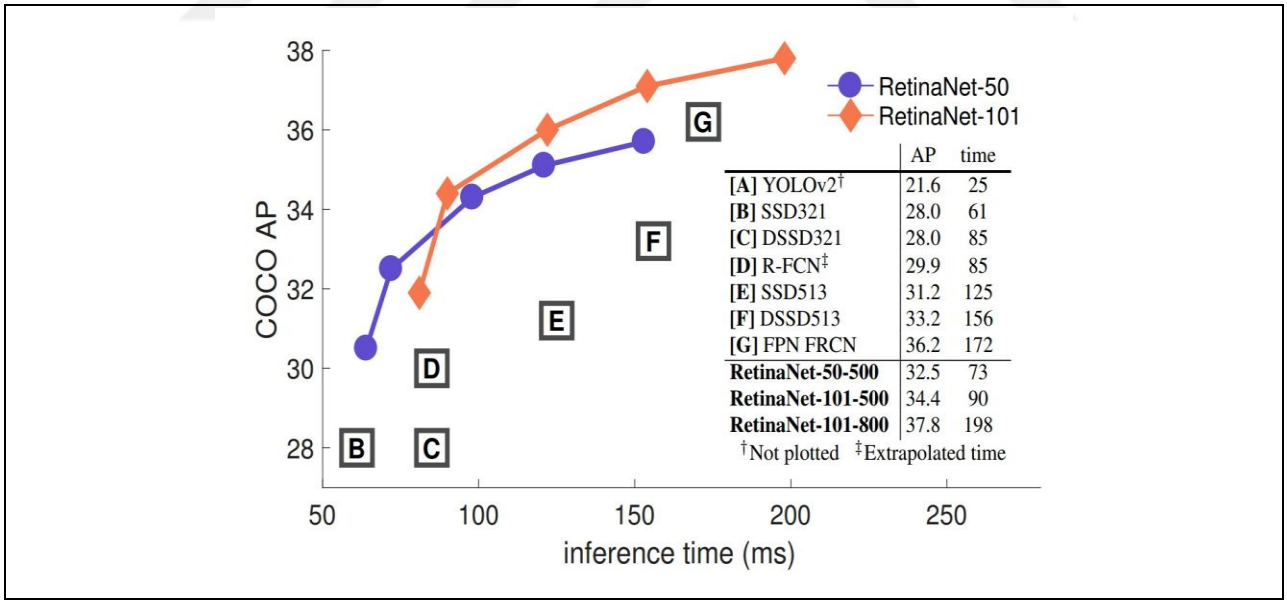

Figure 3.9: Execution time (inference time) versus precision (AP) in the COCO database

The RetinaNet network achieves superior performance to all other structures, including the two-stage Faster R-CNN network [6]. The curve points are generated from the evaluation of images in different resolutions (five scales between 400 and 800 pixels).

3.6 RETINANET

For example, suppose an application that contains 10,000 anchors, where 99% of them correspond to the background and 1% to the foreground. If we classify all the anchors equally, with probability P = 0.1 (considering 0 as the background), we will be getting the prediction right for the entire background and wrong all the predictions of the objects of interest. However, as the class distribution is extremely unbalanced, the resulting average error is low, although the result is useless for a detector.

The result of this is that the training of a one-stage network using cross-entropy loss is completely controlled by the anchors that correspond to the background, while the anchors of interest become less relevant for training. On the other hand, with the use of Focal loss, the classification of the anchors of interest continues to be influential for the training of the network, since the loss referring to trivially classified background anchors becomes less relevant. The RetinaNet network is constructed as shown in Figure 3.3. It is basically composed of 4 fundamental components, described below: backbone: the feedforward convolutional network responsible for processing the original image. Learn to remove the descriptors necessary to describe the image. It can be performed with different types of structures, such as residual networks [4], dense networks [47] and structures such as InceptionNet [48] and EfficientNet [49], etc. Feature Pyramid Network (FPN): component responsible for increasing the network's ability to work with objects at different scales. In the original implementation [7], 5 levels of resolution are used, identified as {P3,P4,P5,P6,P7}, where each level receives anchors with basic resolutions {322,642,1282,2562,5122}, respectively. For example, the P3 level, which works with the lowest resolution, receives anchors with size $32 * 32$ and all its variations. classification subnet: uses the activations of each FPN level to classify each of the corresponding anchors among the possible classes. The same classification subnet is used for the different levels of the FPN, where each level corresponds to a set of anchors with a specific size. refinement subnet: performs a function similar

to the classification subnet, but the result is the regression of a 4-element vector. Each of the anchors receives 4 refinement coordinates, which predict the offset between this anchor and the true bounding box. Note that this refinement is independent of the classes, that is, it is performed globally, and the distinction between the types of objects is made only by the classification network. Classification and regression subnets are convolutional, typically composed of convolutions with common 3x3 filters. This results in a completely convolutional and trainable network with the backpropagation algorithm. The loss used for the classification subnet is the Focal loss, described earlier. The loss used for the regression subnet is the smooth L1 loss (Eq. 3.2) applied to each coordinate, but other losses could be used,

3.6.1 Loss Networks

Much of the study in the area of neural networks is in the development of more efficient structures. The publication of convolutional networks such as LeNet [2], AlexNet [31] and VGGNet [3] corresponds to the development of new techniques that allow the generation of more complex networks with more layers.

The ResNet network [4] is a very powerful structure that achieves

some of the best results in the area of convolutional networks. With the publication of this new technique, the construction and optimization of much deeper networks was allowed, achieving performance superior to the constructions available until then.

One of the motivations for the development of residual networks is the issue of gradient fading.

3.6.2 Gradient Descent

One of the difficulties in developing deeper networks is the vanishing gradient effect [29]. This is a phenomenon that occurs in the training of networks during the use of the backpropagation algorithm, and prevents the proper propagation of the error to the internal layers of the network.

Activation functions such as sigmoid and hyperbolic tangent are used due to the effect of saturation of the output, which keeps the values produced by the network units controlled and prevents the explosion of the internal values of the network. On the other hand, the derivative of these functions

is bounded (always less than a known value) and converges to zero when the input moves away from zero (Figure 3.10).

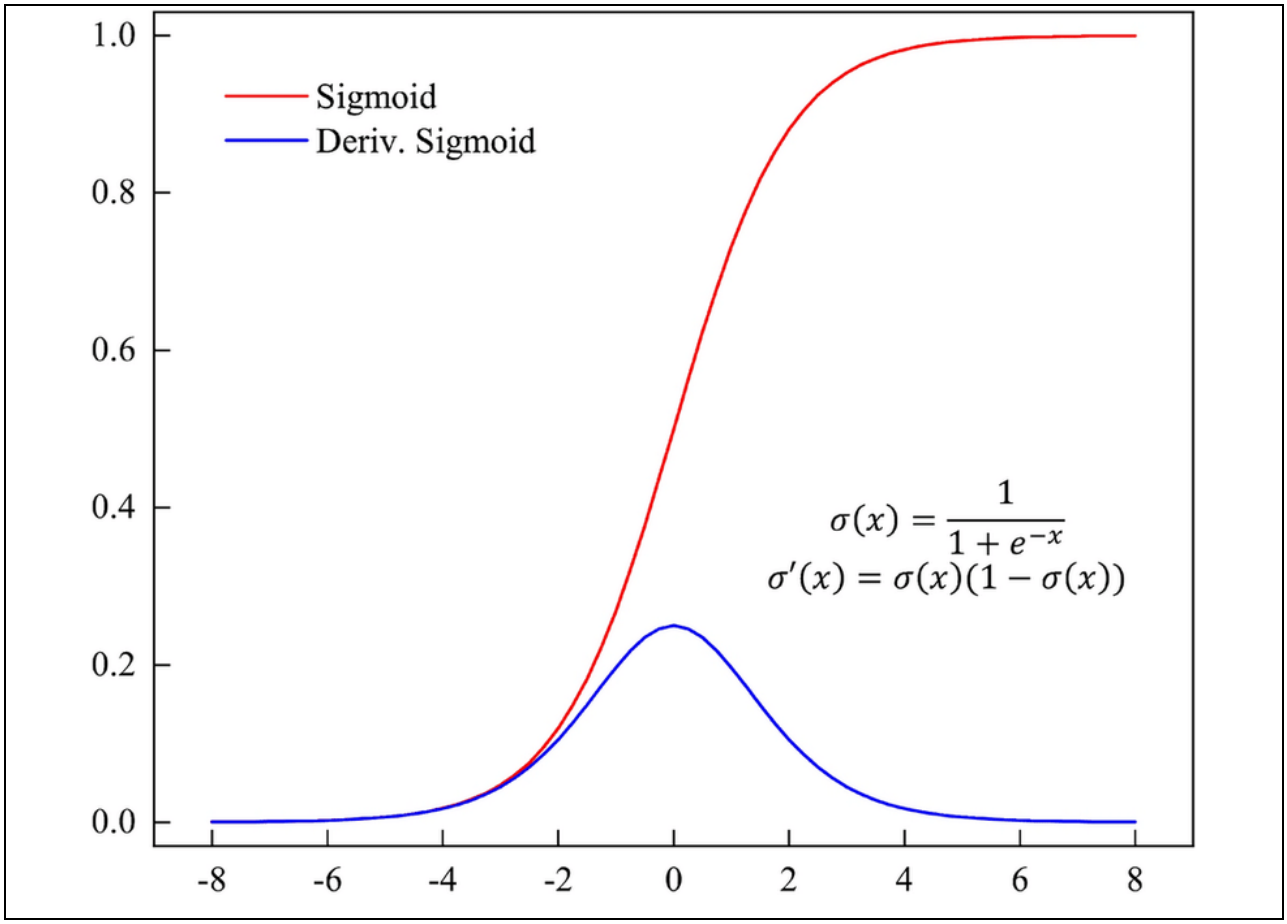

Figure 3.10: Sigmoid activation function and its derivative. It is observed that the sigmoid function saturates for values of | x | large, so that its derivative approaches zero

As seen before, the backpropagation algorithm performs the propagation of the gradient through the layers of the neural network. This process is sequential, and each step requires the multiplication of the previous result by the derivative of the activation function used between the layers. In this way, the use of activation functions such as the sigmoid can generate gradually smaller values of the gradient within the network, so that eventually the value received by the first layers of the network is insufficient to perform any significant optimization in the network, effectively interrupting the training.

There are different ways to mitigate this phenomenon, which mainly involve changes in activation functions and network structure. Some of them are: Using different activation functions. The activation function ReLU [33, 34], for example, maintains a constant derivative value for positive inputs, avoiding the problem of gradient saturation for a larger input range.

Data normalization [34]. We can normalize the activations of each layer of the neural network, placing the signals in an interval closer to zero, and therefore inside a window where the activation function maintains a larger derivative.

Use of auxiliary loss functions [35], placed in intermediate layers of the network, in order to include an additional loss term for the first layers of the network.

Use of convolutional blocks that use skip connections or shortcuts between the layers of the network, creating new paths for propagation of the gradient.

3.6.3 Residual Blocks

A residual block consists of the joining of a set of traditional layers plus a shortcut connection in parallel. Figure 3.10 shows an example, where the shortcut connection is the identity function.

The use of residual blocks generates a direct path between y and x, which always allows the existence of a significant gradient regardless of the function performed by the other layers. The use of this technique allows the construction of much deeper networks without loss of performance.

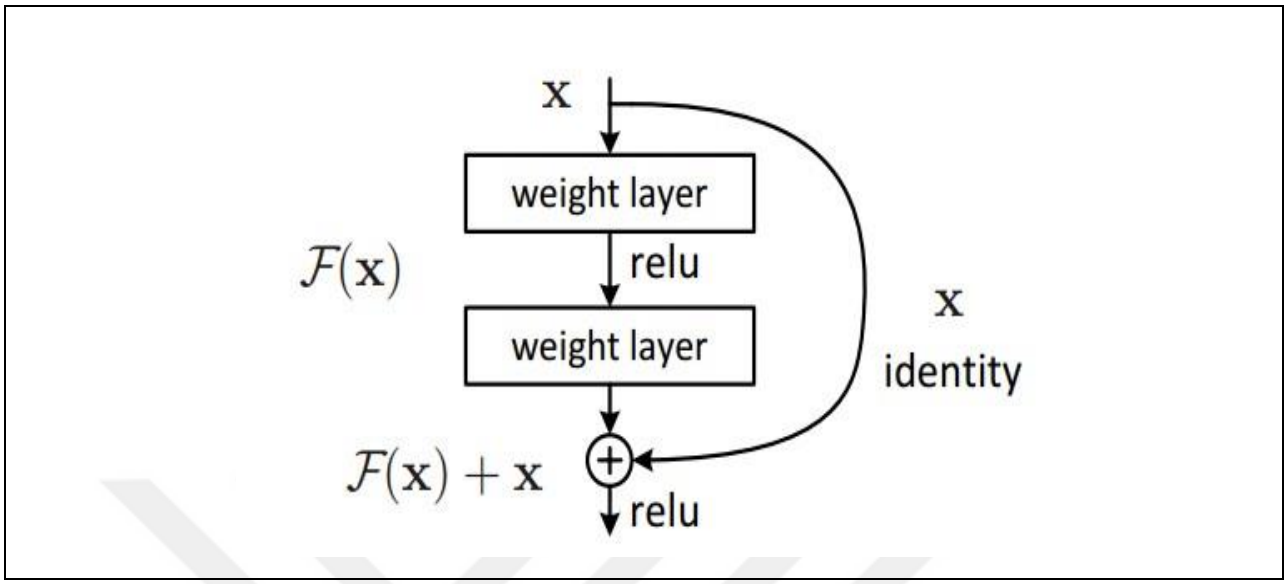

Figure 3.11: Basic block of a residual network [4]

There are several variations of this construction, modifying both the function performed in the shortcut connection and in traditional convolutional layers.

The increase in error between the ResNet110 and ResNet1202 networks is attributed to the overfitting of the network, given that the capacity of the second network is much higher and the training error of both is reported to be similar.

3.7 OVERFITTING, UNDERFITTING AND AUGMENTATION

Within the context of machine learning, it is desired that the model produced is capable of generalizing the knowledge obtained in training to new data samples [29]. From this definition, two situations arise that define the behavior of the model for new samples: overfitting: situation where the trained model does not learn the internal structure of the dataset, but rather memorizes the local characteristics of the training examples. In this way, the generalization capacity of the model is low, and the performance of the algorithm for new samples is inferior to the performance obtained during training.

underfitting: situation where the model does not reach the desired performance both during training and in new samples. This happens in cases where the capacity of the model is not sufficient to represent the dataset, or that the model has not been sufficiently trained.

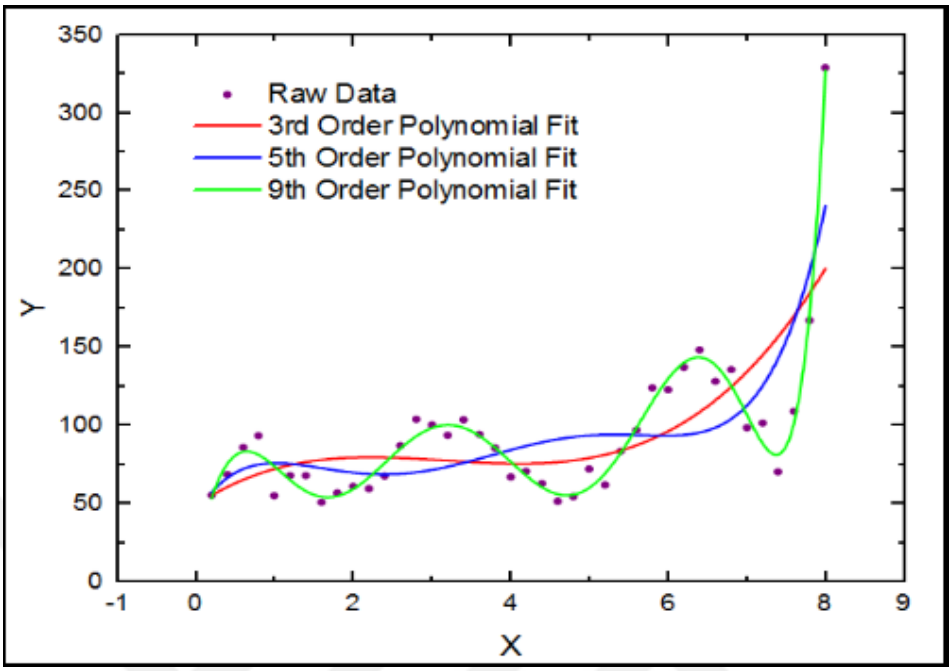

Figure 3.12: Regression of the same data set, using polynomials of different degrees

We see that a low degree polynomial does not have the ability to adequately approximate the function, while very high degree polynomials tend to diverge to values intermediate to the training samples. The optimal result can be obtained using an intermediate solution.

The appearance of overfitting in neural networks is quite common in practice, due to the high parameterization of modern models and the finite number of training samples. The result is the development of a model that makes predictions based on local characteristics of each sample, basically memorizing the data set, while the generalization ability is reduced.

A common way to alleviate this problem is to use augmentation at training time. This technique aims to artificially extend the dataset, creating new samples from random variations of the original samples.

The effect of this is that the model is not trained on the same image multiple times, but on distinct random variations of the same image. The tendency is for the basic structural characteristics of the image to remain constant (assuming that the processing used does not compromise the image content), while noise becomes an inconsistent variable for the model.

This technique is especially powerful for image classification and detection applications, due to the high dimensionality of the variables and the often small number of samples.

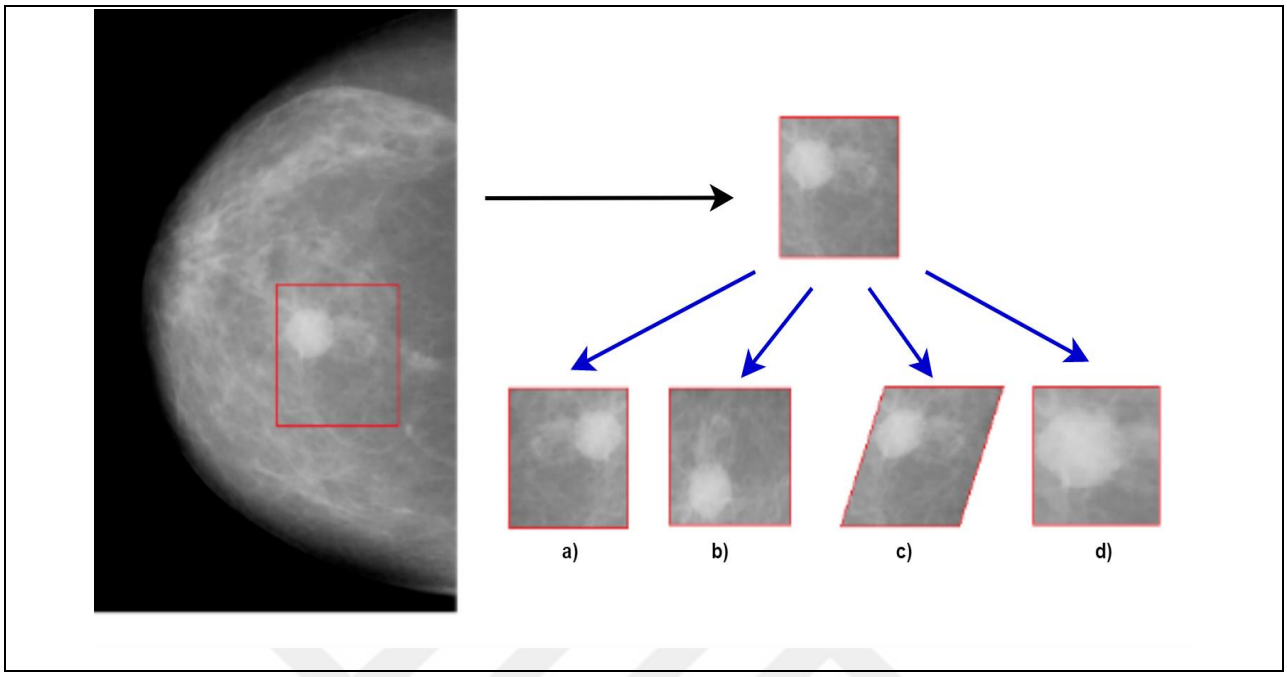

Figure 3.13: Examples of augmentation over an image region. a) and b) consist of a horizontal rotation and rotation of the image, respectively, c) consists of distortion (skew) and d) represents a change of scale (zoom) of the image.

Many other types of augmentation exist, not just limited to affine transformations, such as adding white noise, changing the contrast and histogram, etc.

3.8 FEATURE PYRAMID NETWORK

One of the challenges present in the area of object detection is the recognition of objects at different scales. The same object can come in different sizes, depending on the relationship between the object and the camera, which makes it difficult to develop algorithms capable of detecting objects in any situation.

This problem is even more pronounced in situations where you work with objects of different classes, and different classes naturally contain different sizes. The resulting resolution in a convolutional network may be adequate for evaluating some classes of objects, but inadequate for others. It is therefore interesting to develop detection techniques that allow working with images and objects subject to scale variations.

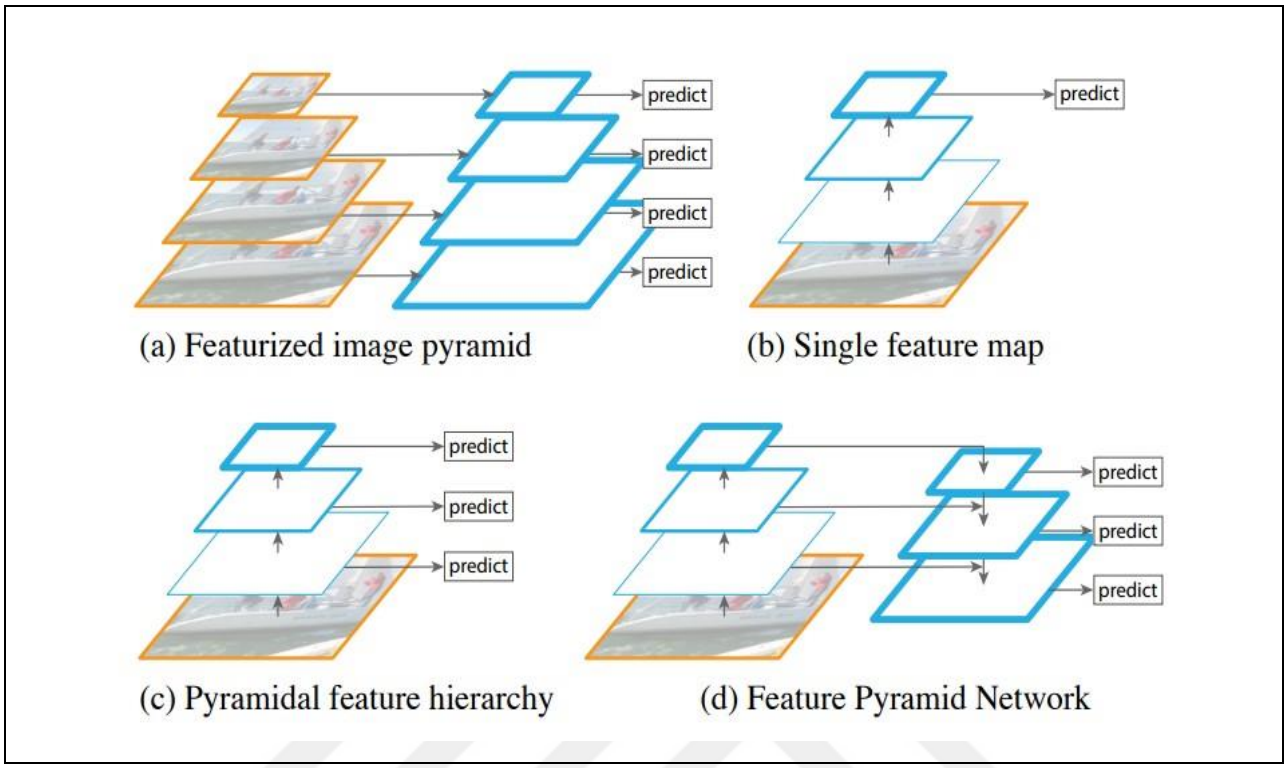

Figure 3.14: a) Using an image pyramid, b) Using only the last convolutional layer allows for much faster processing, c) The use of intermediate layers produced by the convolutional network allows working with multiple scales at no additional cost, d) The Feature Pyramid Network (FPN) proposal allows a structure with performance equivalent

The layers produced allow working with different scales, still maintaining activations with high-level information, Featurized Image Pyramid [36] (Figure 3.13(a)) is the basis of current solutions, being widely used both with detectors based on descriptors and convolutional networks. This method consists of making predictions for different scales of the original image, where each level of the pyramid is analyzed independently of the others. This ensures that each level of the pyramid maintains a high level of information, but at different scales. The main disadvantage of this method is the need to perform all the processing for each level of the pyramid, significantly increasing memory usage and processing time.

Single Feature Map(Figure 3.14(b)) is the basic method used in convolutional networks, where detection is done only in the last activation layer of the convolutional network. Originally used in networks such as [37, 6], this method obtained interesting results, but the inclusion of multi-scale techniques results in more accurate structures.

Pyramidal Feature Hierarchy promises to solve the problem of multiscale and computational cost, using layers already naturally produced by the convolutional network for prediction. Used by networks such as SSD [38], this technique mainly fails in the detection of small objects, since the layers with higher resolution are not processed enough to allow the detection of objects.

Finally, we have the Feature Pyramid Network (FPN) [36] structure that allows us to produce layers with different scales, all of which contain high-level information capable of enabling object detection.

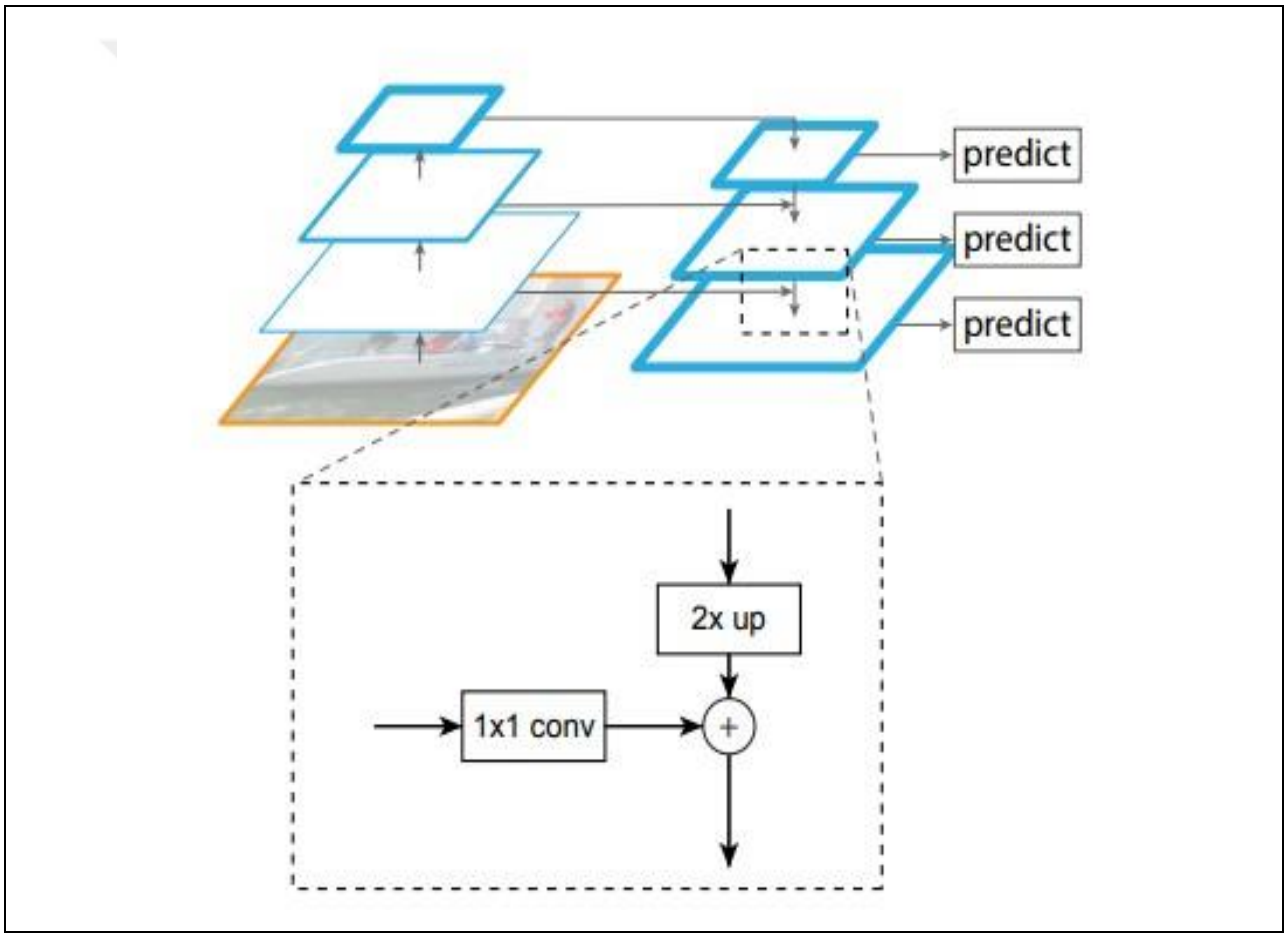

Figure 3.15: Feature Pyramid Network

Multi-scale layers are produced from a top-down structure and by the use of side connections. Building high resolution layers from low resolution layers (using 2x upscaling) allows propagation of high-level information throughout the pyramid, while lateral connections (1x1 convolutions) help to reconstruct the spatial relationships of the original image The FPN framework allows

creating multi-scale layers from a basic network, called a backbone, with minimal extra processing. Figure 3.15 shows an example of construction done from a residual backbone (ResNet), resulting in 4 different layers {P2, P3, P4, P5}.

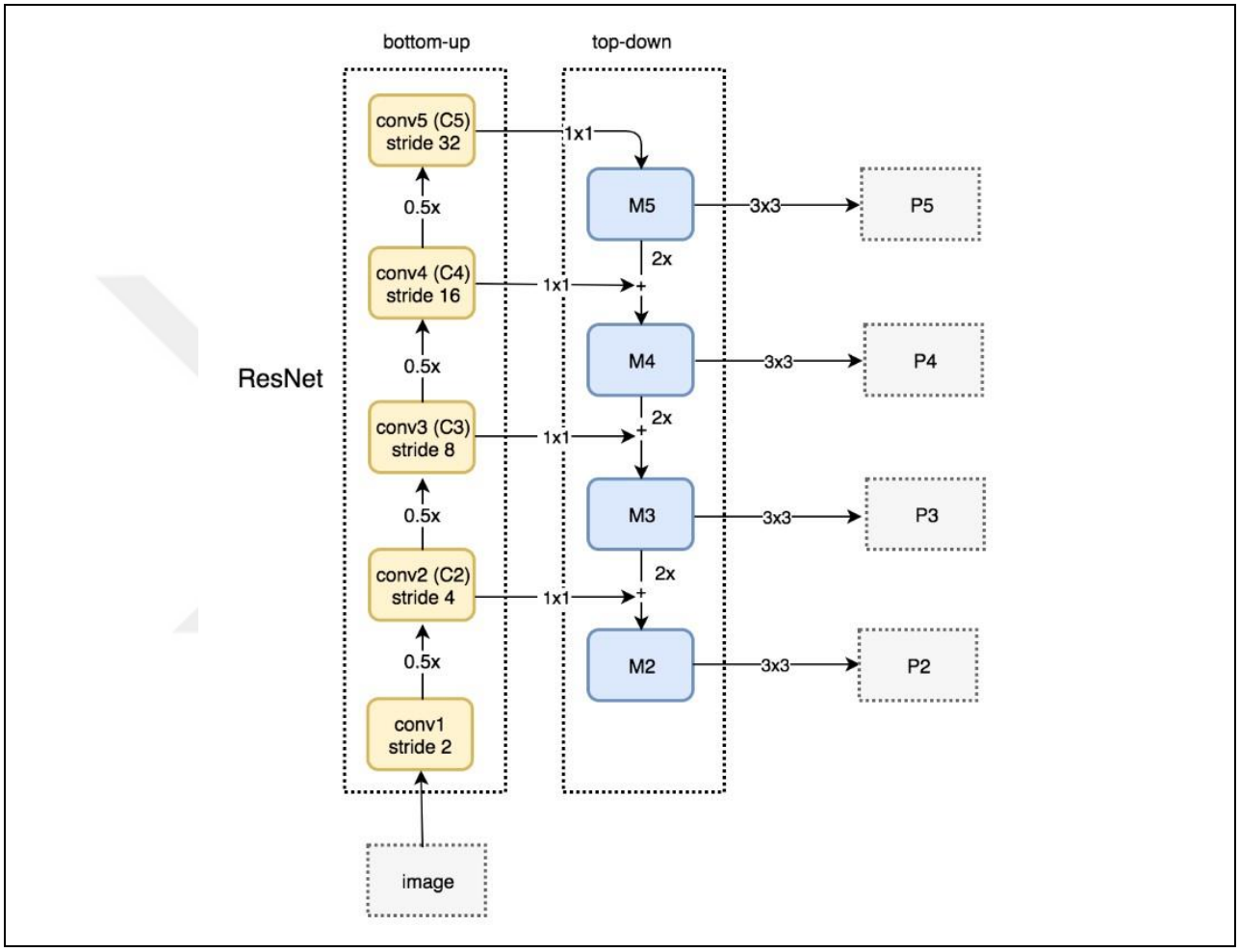

Figure 3.16: Feature Pyramid Network built from a ResNet backbone [5]

The bottom-up structure refers to the convolutional network, where convolutional layers (conv) and pooling layers (0.5x) are interleaved. The top-down construction builds the activation pyramid, using lateral connections (1x1 convolutions to modify the number of activation channels) and upsampling (to reverse the pooling process) and finally the element-by-element sum of the two branches. The Pn output of the layers is produced by a last convolution of the obtained values.

It is observed that the Pn levels are created from the activation layers of the original backbone. Each level is produced sequentially, forming outputs with increasing resolutions and a high level

of semantic information. In this way, object detection can be done at all levels, making it possible to detect objects at multiple scales efficiently and accurately.

3.9 ANCHORS

Object detection algorithms consist of evaluating multiple regions of an image, determining whether or not a given region contains the objects of interest. Using anchors allows you to perform this task quickly and accurately for a large number of regions.

Anchors are a set of bounding boxes with pre-defined height and width, placed on a regular basis over the original image. Object detection is performed by classifying and fine-tuning each anchor, indicating that the region of the original image corresponding to this anchor contains the object of interest.

The detection process is done by generating a relationship between each anchor in the original image and the network output. Each anchor defined on the original image is considered a reference region, and the purpose of the network is to classify (define if the anchor corresponds to an object) and adjust (correct the position of the anchor in relation to the object). Here is a set of 9 anchors centered on the same point, with different sizes and shapes. This set is repeated throughout the image, so that the resulting anchor set covers the entire image.

The predictions made by the network are performed in relation to the anchors arranged, as in Figure 3.16. For each anchor within a set of possible N, the following values are predicted: Classification: each anchor is classified among the K possible classes, resulting in a vector with $N \times K$ probabilities.

Regression: the position and size of each anchor are refined, in order to adjust their positioning in relation to the real object in the image. The result is an offset for each of the four coordinates of the anchor, forming a vector with $N \times 4$ elements.

The classification and regression of the anchors is done based on the activations of the convolutional network [36, 6], as shown in Figure 3.17. The activation layer shown is the result of a single convolutional network, but the same principle can be applied to multiple outputs, as in the case of an FPN structure.

Classification and regression are performed using two subnets that evaluate the activation layer. The implementation of these subnets is generally convolutional, with a much smaller number of layers than the main network.

Assuming an activation layer with W × H resolution (and any number of channels), W × H regions in the original image will be evaluated, each with A corresponding anchors. In this way, the results of the classification and regression subnets are vectors of size W × H × A × K and W × H × A × 4 respectively.

It is noticed that the total number of anchors in the image can be different for different applications. The main parameters that modify the number of anchors refer to the resolution of the activation layer (W × H), which regulates the number of regions evaluated, and A, which defines the number of anchors per region. In this chapter, the concept of one- and two-stage detection networks will be presented. We will see the characteristics of these two methods, and finally the RetinaNet detection structure will be presented, which will be used in the simulations of this work.

4. PROPOSED METHODOLOGY

4.1 SYSTEM OUTLINE

Artificial Neural Networks (CNN) are computational techniques that present a mathematical model inspired by the central nervous system of living and intelligent organisms, that acquire knowledge through experience, and that are capable of performing machine learning, pattern recognition and classification of data and can be used in a wide range of applications. A large artificial neural network can have hundreds or thousands of processing units, or neurons, and its operation is relatively simple.

An CNN performs operations based on attributes, accessed by the input chCNNels, and certain weights, which indicate their influence on the unit's output. This is typically organized in the form of layers, with units that can be connected to the units of the previous layer. A neural network is mainly specified by its topology, the scattering of cells in the layer, and the training rules. The intelligent behavior of an CNN comes from the interactions between the processing units of the network.

Most neural network models have some training rule, where the weights of their connections are adjusted according to the presented patterns, that is, they learn through training offered based on previously labeled data.sides.

4.1.1 Problem Addressed

Among the applications of CNNs, we highlight their use as pattern recognition, that is, to identify and classify information into categories. With that in mind, this work presents the modeling and development of an CNN to classify a database with information related to breast cancer. Breast cancer is the second most common cancer that occurs in women compared to all other types of cancer. Breast cancer is fatal in less than half of all cases and is the leading cause of cancer death in women, accounting for 16% of all cancer deaths worldwide.

Automated processes allow you to improve the efficiency of the breast cancer detection and classification processa, allowing doctors to diagnose the problem, assess the expense needed to solve it and even define the time needed to treat the disease in the fastest and most accurate way.

4.1.2 Database

The database chosen to be classified is Breast Cancer Wisconsin (Original), it was obtained from the hospitals of the University of Wisconsin, in Madison, United States, in cases diagnosed by Dr. William H. Wolberg [3]. It is composed of a dataset that records the characteristics for breast cancer cases. There are two classes, benign and malignant. The malignant class of this dataset is considered to be outliers (outliers, which present a large deviation from the others), while the points in the benign class are considered to be inliers (instances considered normal, within the standard).

The database has a multivariable feature, where a random variable is based on the behavior of other variables, whether it is within the standard sample or not.The data are divided into 699 instances, of which 458 are benign (65.5%) and 241 are malignant (34.5%), with 0 identified as benign and 1 as malignant. Each instance has 9 attributes with numeric integer value between 1 and 10. These attributes and their ranges are:

a. Group thickness: (1 - 10)
b. Cell size uniformity: (1 - 10)
c. Cell shape uniformity: (1 - 10)
d. Marginal adhesion: (1 - 10)
e. Single epithelial cell size: (1 - 10)
f. Bare cores: (1 - 10)
g. Soft Chromatin: (1 - 10)
h. Normal cores: (1 - 10)
i. Mitoses: (1 - 10)

For this work, 683 instances were considered, divided into 70% for the training set, 15% for validation and 15% for tests, 16 instancess were disregarded due to some discrepancy in the data.

To illustrate the database, a Parallel Coordinates chart was used (Figure 4.1) which allows viewing the values of each instance of different classes, each row showingpresents an instance where the values on the X axis represent the attributes and on the Y axis the values of the attributes. In Figure 1 it is possible to identify that most class 0 data have low values, whereas class 1 values have dispersed values in the domain.

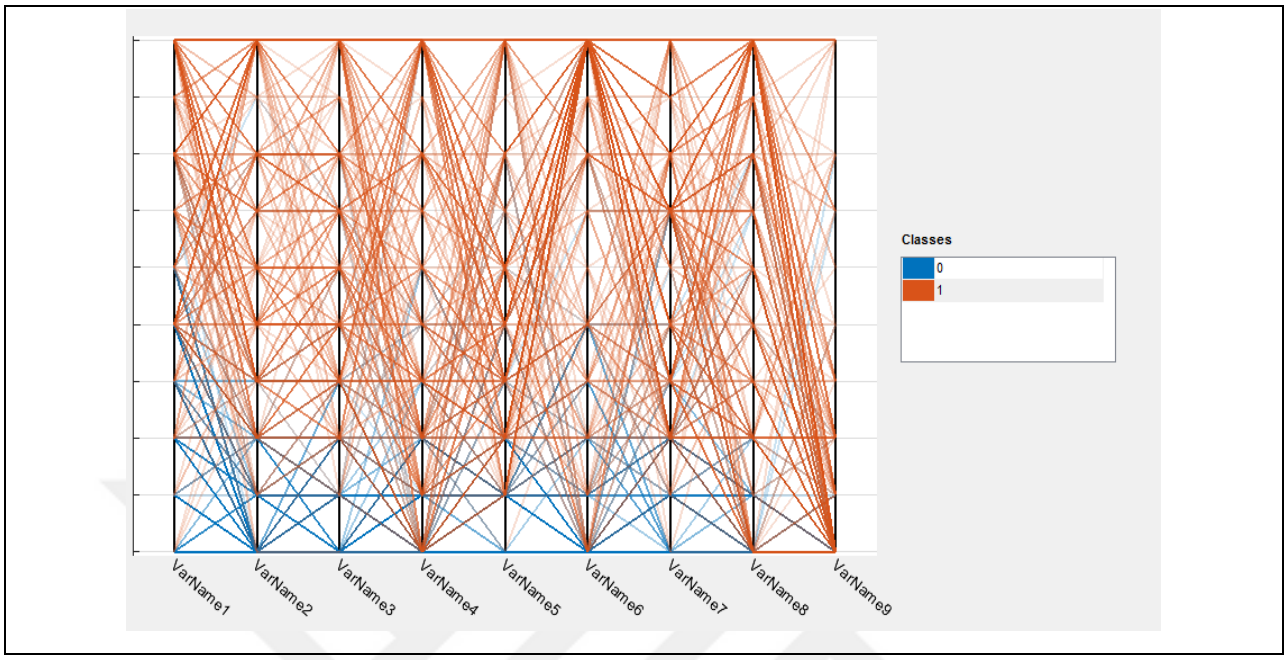

Figure 4.1: Parallel Coordinate Chart

4.2 DEVELOPMENT

After processing the database, a check was carried out to identify whether the data were linearly separable, for which the Scatter Plot chart of Classification Learner [4] of variable 9 was used. Figure 4.2 demonstrates the dispersion of attribute nine of class 1 , Figure 4.3 shows the dispersion of attribute nine of class 0. To verify if the classes are linearly separable, attributes 9 of class 1 and 0 were placed on the graph, Figure 4 shows that it is not possible to draw a line and separate these data linearly, it is soon ruled out using a perceptron neural network, requiring a more complex neural network.

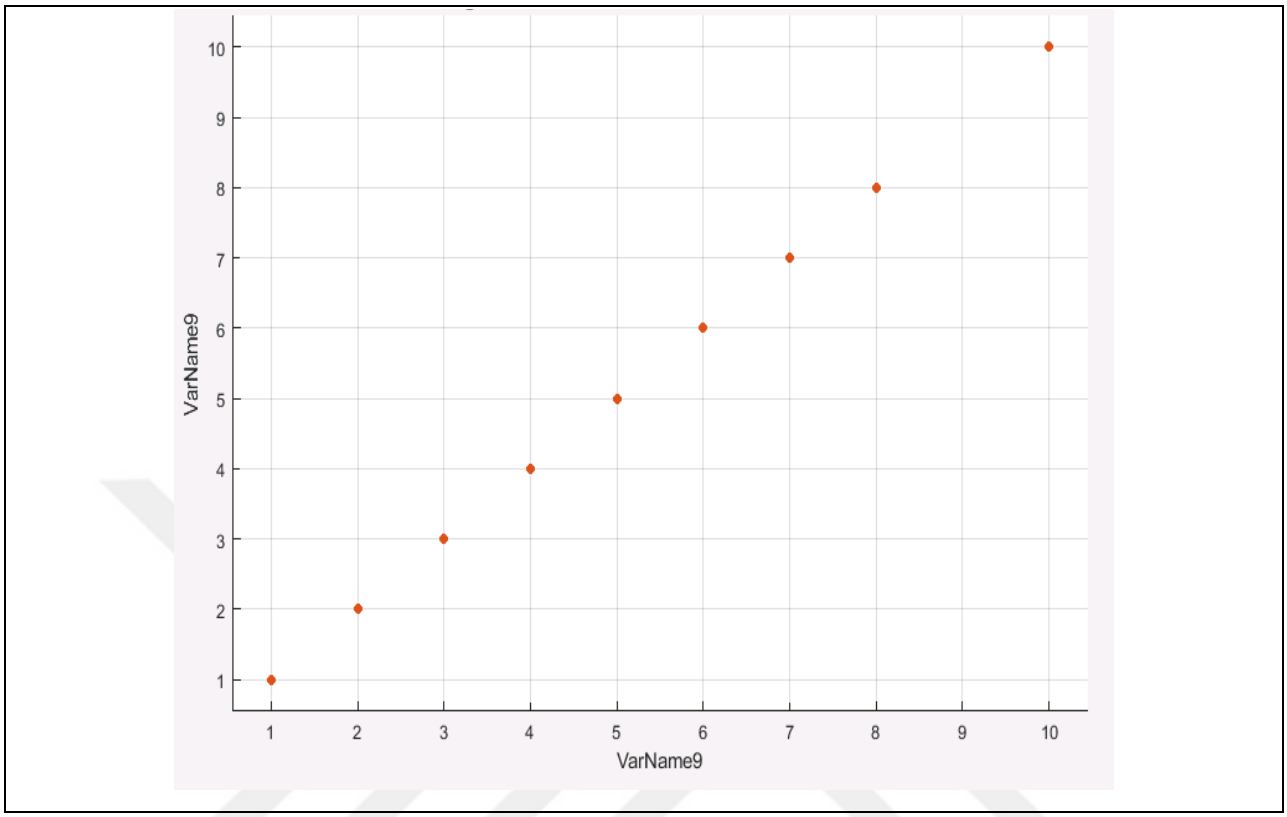

Figure 4.2: Dispersion of the Malignant class

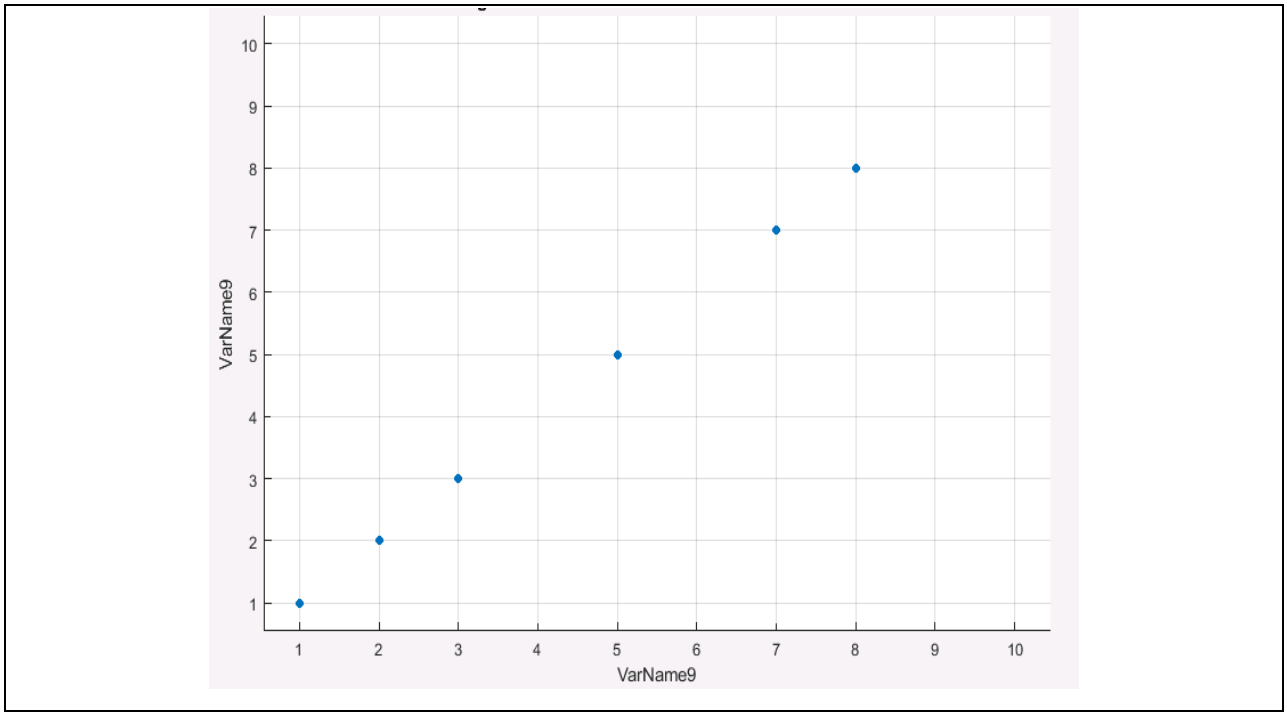

Figure 4.3: Dispersion of the Benign class

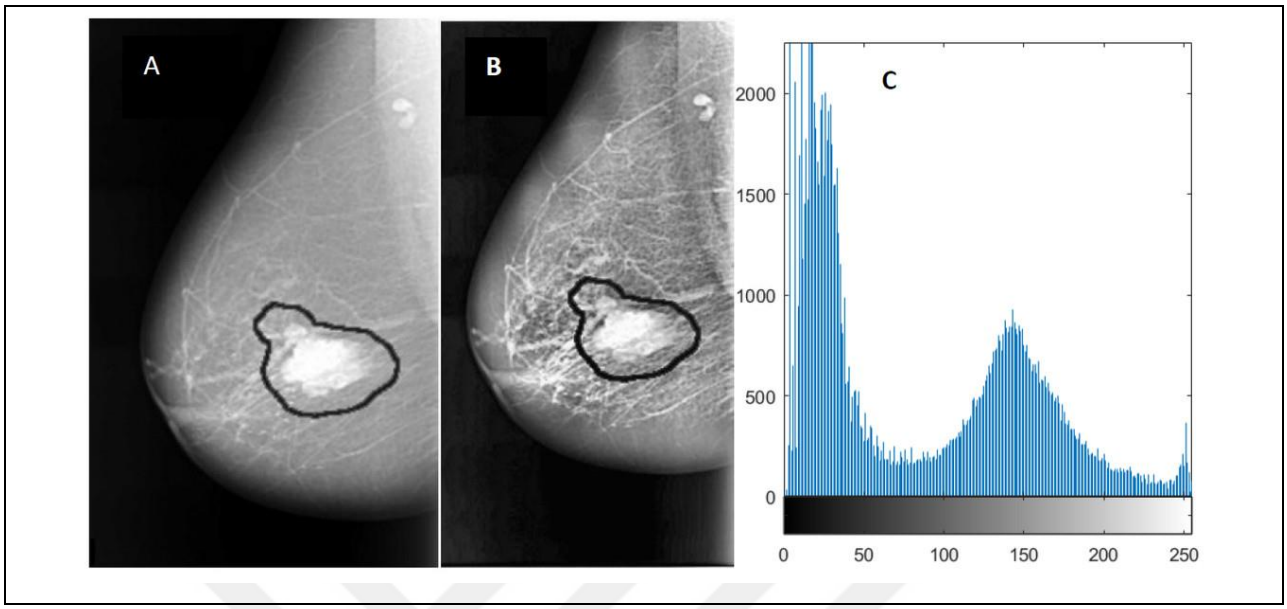

Figure 4.4: Dispersion of the superimposed Malignant and Benign classes

For the development of the CNN, Matlab and the Neural Net Pattern Recognition Toolbox were used to generate an initial architecture of a Feedforward artificial neural network.(most common type of neural network in practical applications). It has no cycles and is often represented in multiple layers. In the first layer, the input layer, the neurons receive excitation signals and in the output layer, the last layer, the result of the CNN processing is sent. Between the input and output layers are the hidden layers (hidden layers, intermediate or hidden layers).

The architecture of a Feedforward Artificial Neural Network is defined by the way in which artificial neurons can be clustered. Figure 5 presents the structure of a Feedforward multilayer network, containing an input layer (Input Layer), a hidden layer (Hidden Layer) and an output layer (Output Layer).

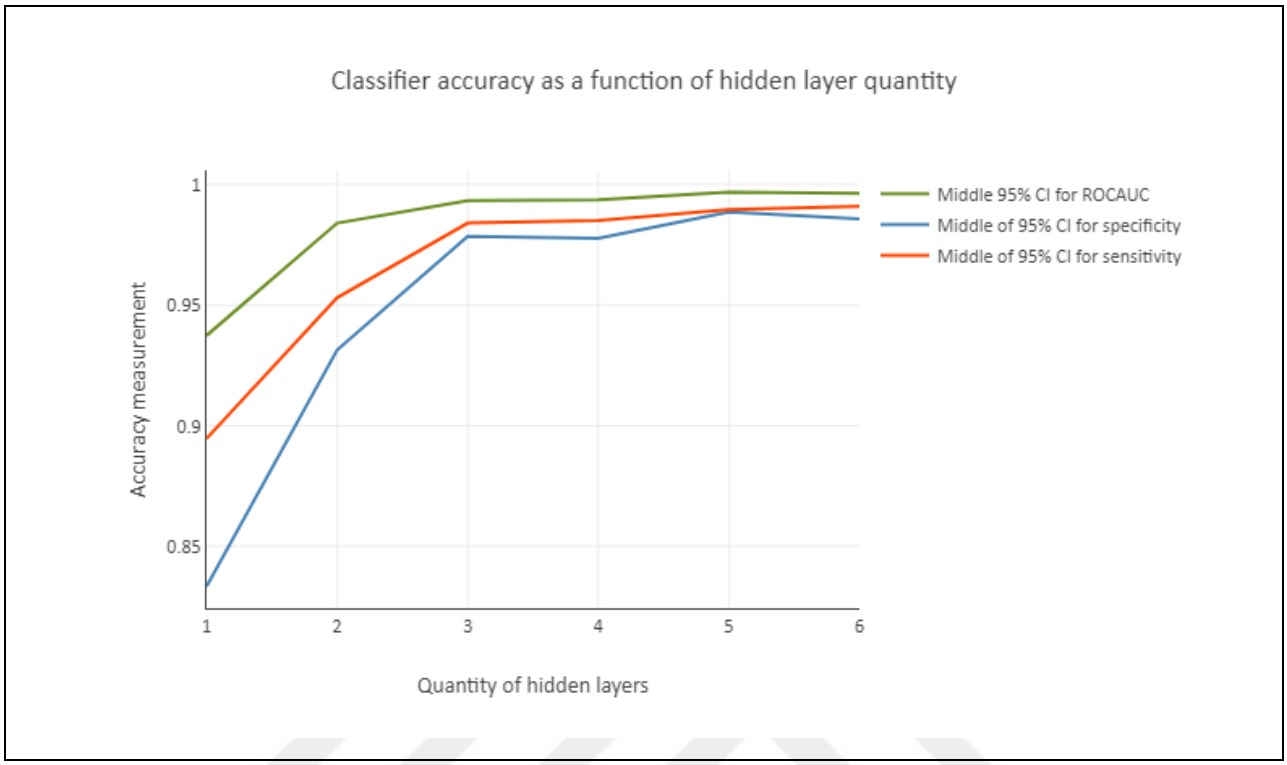

Figure 4.5: Training and validation of a Feedforward CNN

the type of network feed forward implemented is a Pattern Recognition Network [6] that is trained to classify inputs according to target classes. The configuration parameters used were the size and amount of hidden layers and the training function. Figure 4.6 shows the structure of the neural network implemented using Pattern Recognition Network.

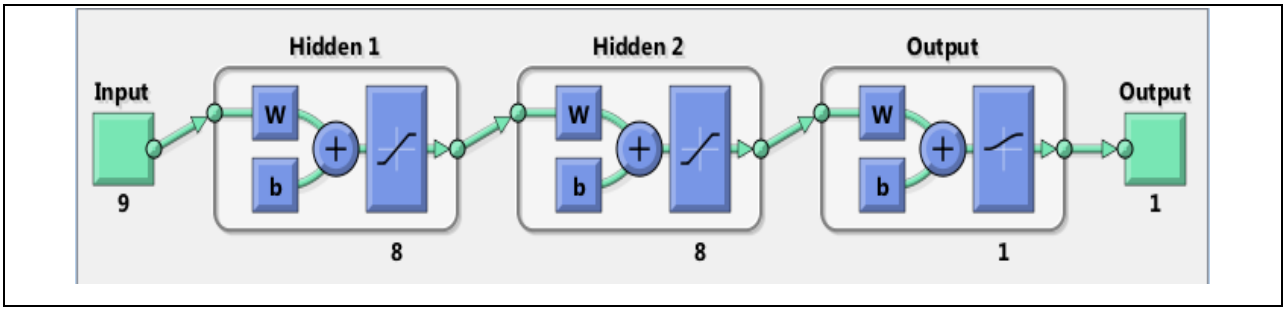

Figure 4.6: Structure of the implemented CNN

4.3 TESTS

The tests were performed according to the parameters below:

a. Input layer: 9 neurons.

b. Output layer: 1 neuron.

c. Hidden layers (hidden output): 1 or 2.

d. Neurons in each hidden layer: 6, 7, 8 or 9.

e. Layer topology function: tritop.

f. Training algorithm used: trainbr.

g. Activation function used: satlins, purelin, poslin, softmax.

h. Execution division: training, validation and testing.

Initially, the neurons of the input layer were defined as nine neurons due to the database having nine attributes, and the neurons of the output layer being a single neuron with the softmax activation function [9]. To define the training algorithm used, tests were carried out in order to find out which would be the best training, for which a Feedforward neural network with two hidden layers was used. The training algorithm selected was trainbr (Bayesian Regularization Backpropagation) which takes longer, but is better for solving complex problems. The algorithm updates the weight and bias according to the Levenberg-Marquardt optimization, determining the best combination to produce a very generalized network. The Marquardt adjustment parameter value was the default value corresponding to 0.005.

To define the numbers of hidden layers as well as the number of neurons in each hidden layer, the article by Jeff Heaton [8] was consulted, which aims to help define these parameters. According to Heaton, the number of hidden layers should be one for continuous and finite space mappings and two for problems that require greater precision and more than two layers for more complex representations of data. Based on this, we used hidden one- and two-layer variations in the tests.

According to [8] a calculation is used to estimate the amount of initial cells in each hidden layer, below are the three guidelines to be followed:

a. The number of hidden neurons must be between the size of the input layer and the size of the output layer.
b. The number of hidden neurons should be 2/3 the size of the input layer plus the size of the output layer.
c. The number of hidden neurons must be less than twice the size of the input layer.

Therefore, the number of neurons calculated is seven, however, in the present work, variations in the number of neurons in each hidden layer were tested, being 6, 7, 8 and 9 neurons. To connect the neurons we used a tritop topology function that calculates the positions of neurons for the layers which are arranged in an N-dimensional triangular grid, the dispersions used were: 2x3, 3x2, 1x7, 7x1, 2x4, 4x2, 2x2x2 and 3x3. These distributions have been described and can be viewed in Appendix A.

In the article by Snehal Gharat [10] it is explained how the different types of activation functions work, as well as what is the best activation function to be used for a specific case. classification, linear functions are the most used and currently produce excellent results being only used in hidden layers. Based on these statements and because the linear functions allow the neural network to converge very quickly, the linear activation functions satlins, poslin and purelin were tested. described below:

d. Satlins: is the saturated symmetric linear function, which calculates the output of a layer from its input. A = satlins(N) receives N and where N is the Input Matrix.

Algorithm:

$$\text{satlins}(N) = -1, \text{ if } N <= -1 \quad (4.1)$$

$$N, \text{ if } -1 <= N <= 1 \quad (4.2)$$

$$1, \text{ if } 1 <= N \quad (4.3)$$

Graphic:

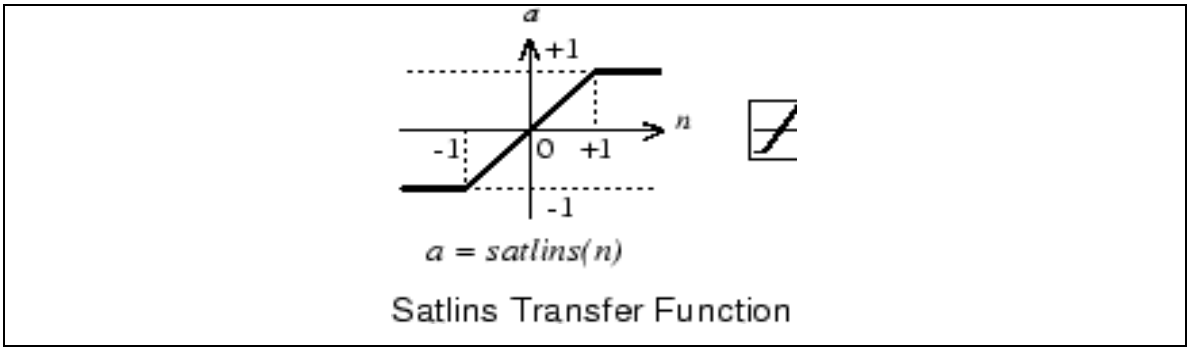

Figure 4.7: Test results using satlins can be found in Appendix B

e. Purelin: is a linear transfer function that calculates the output of a layer from its input. N is the Input Matrix.

Algorithm:

$$a = purelin(n) = n \qquad (4.4)$$

Graphic:

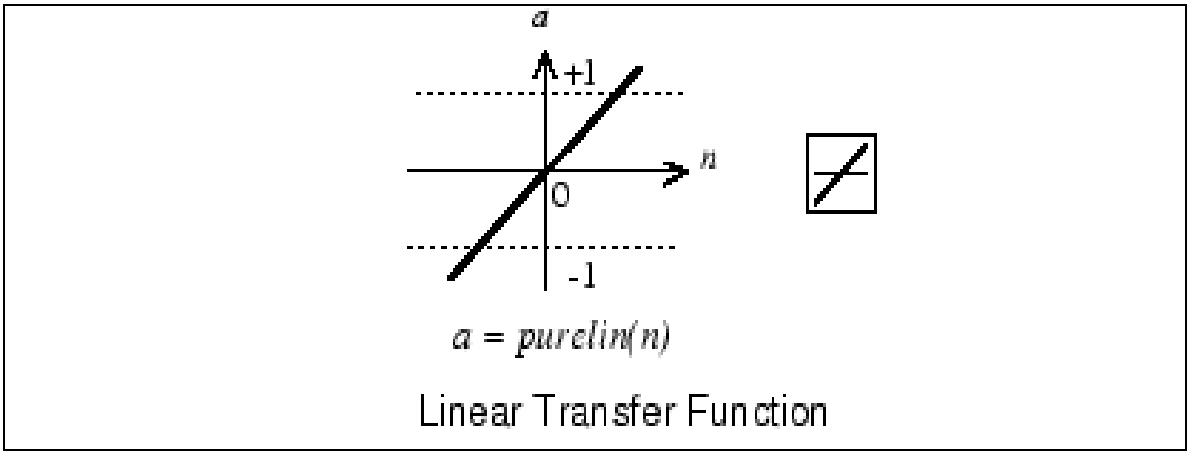

Figure 4.8: Test results using purelin can be found in Appendix C

f. Poslin: it is the positive linear activation function ie the function only corresponds to positive values, calculates the output of a layer from its input, poslin(N) receives N and where N is the input matrix.

Algorithm:

$$poslin(n) = n, \text{ if } n >= 0 \qquad (4.5)$$

= 0, if n <= 0 (4.6)

Graphic:

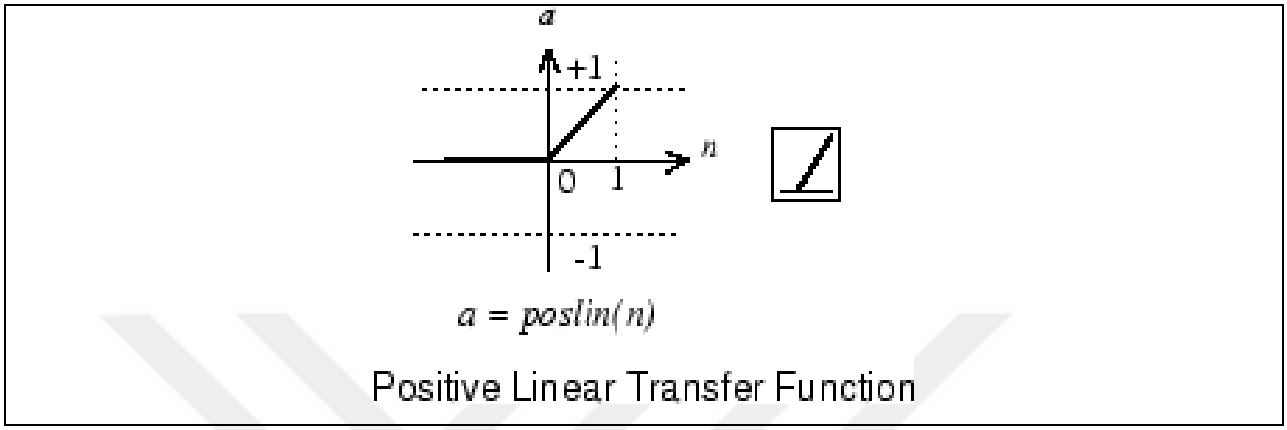

Figure 4.9: Test results using poslin can be found in Appendix D

g. Softmax: The softmax activation functions generate a vector representing the probability distributions of a list of possible outcomes. It is recommended in cases where the output of the classification network is binary, that is, it has only two classes of output, which fits into the house of this network, where the outputs are class 0 (benign) and class 1 (malignant). The performance of the softmax function, when compared to sigmoid, presented a very similar performance, and with a mathematically simpler logistic function.

Algorithm:

a = softmax(n) = exp(n)/sum(exp(n)) (4.7)

Graphic:

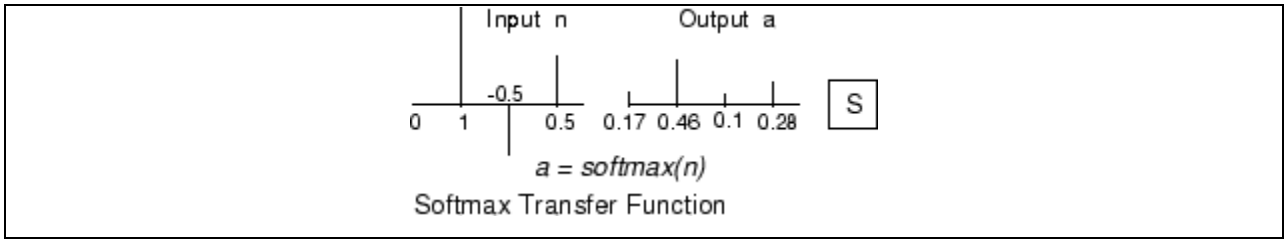

Figure 4.10: The divideint function of matlab was used to perform the separation of the data base in the training, validation and test sets

After carrying out tests with activation functions and cell distributions in a single hidden layer, the best configurations and combinations for the two layers were selected in order to obtain a more accurate classifier. To this end, two cutoffs were determined:

h. Average: greater than or equal to 97.4%

i. Standard deviation: less than or equal to 1.6%

These parameters were designed to ensure greater accuracy with less oscillation between the data.

4.3 RESULTS AND CONCLUSIONS

In total, 330 network run tests were performed to calibrate the parameters (all these tests can be found in appendices B, C, D and E). Based on these tests, we determined that the best configuration found to solve the classification problem addressed was a neural network with the following configurations: two hidden layers, the first with a 2x4 configuration and the other with a 2x2x2 configuration, both having the activation function satlins. This configuration presented an average accuracy of 97.73%, with a standard deviation of 1.29%. Tests found in Appendix E (see table E1).

Figure 7 presents in detail the confusion matrices (Training, Validation, Test and Sum Total) with the best configuration of the network, in which the final tests were performed. Below is a brief description of the arrays:

a. Training (97.3% Hits / 2.7% Mistakes):

- Class 0 hits: 300 (97.4%)
- Class 1 hits: 166 (97.1%)

b. Validation (95.1% Hits / 4.9% Errors):

- Class 0 hits: 67 (97.7%)
- Class 1 hits: 30 (90.9%)

c. Test (Hit 98.0% / Error 2.0%):

- Class 0 hits: 67 (100%)
- Class 1 hits: 33 (94.3%)

d. Final Result (Hit 97.1% / Error 2.9%):

- Class 0 hits: 434 (97.7%)
- Class 1 hits: 229 (95.8%)

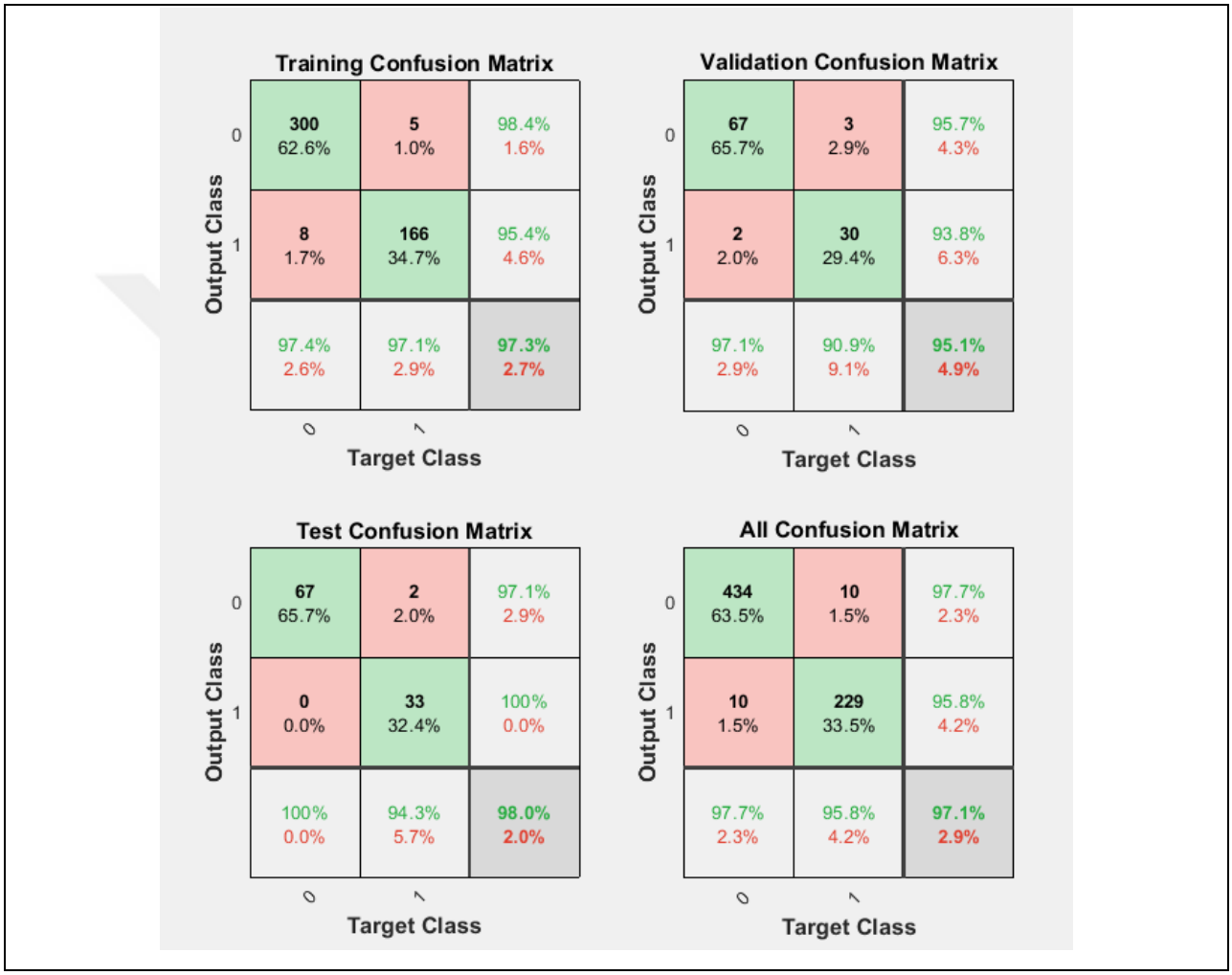

Figure 4.11: Implemented network confusion matrix

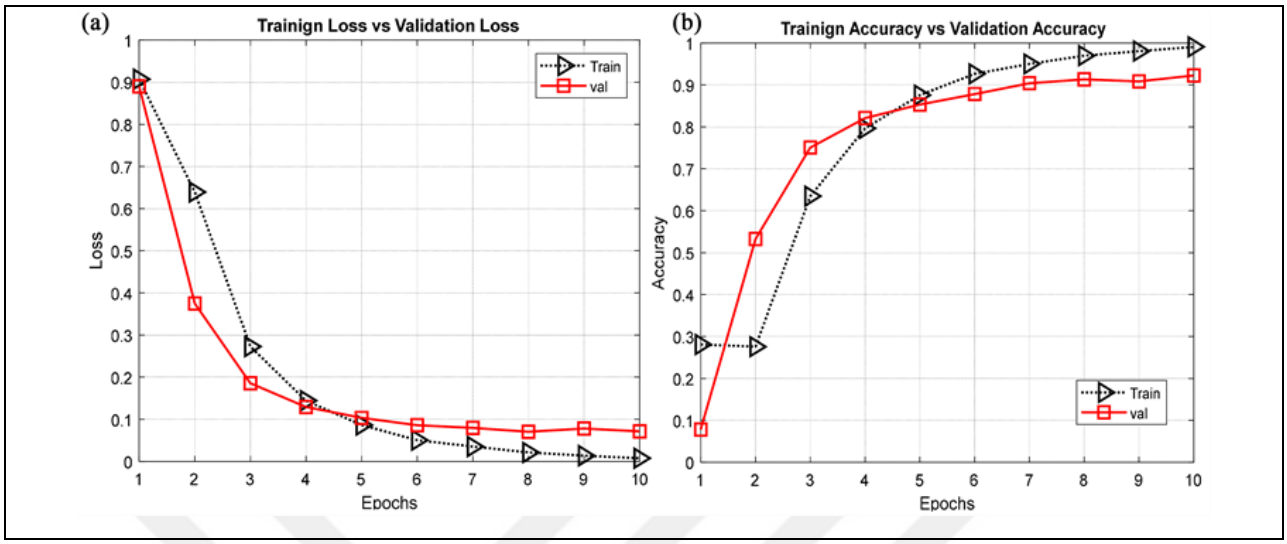

Figure 4.12: Comparing training and testing accuracy in CNN

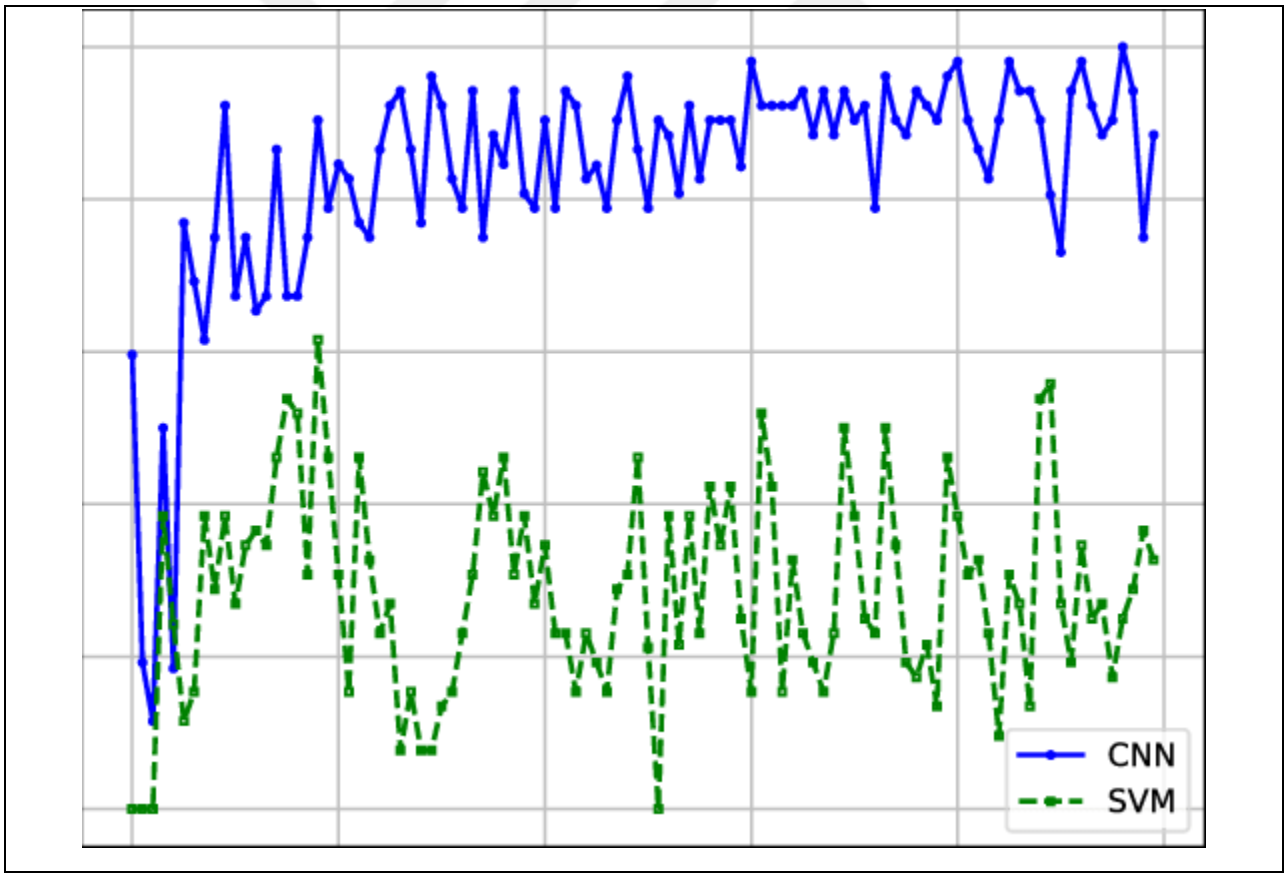

Figure 4.13: CNN vs SVM acracy and loss

5. CONCLUSION AND RECOMMENDATION

5.1 CONCLUSION

The project was initiated by a systematic review, through which it was possible to obtain a better understanding of the research area on CAD systems for breast cancer. Through the answers to the research questions carried out, it was possible to identify the most frequently used types of exams and Artificial Intelligence techniques. The main difficulties encountered during the systematic review were the omission by some researchers of important details for the understanding of their research and the lack of consistency in the use of some terminology in the area. Based on the results obtained and on the set of similar works selected, it was possible to delimit the scope of this project and formulate the solution that would be implemented, a thresholding technique was used to remove the fundus from the mammography, with the segmentation of the breast region being performed using a region growth algorithm.

The developed algorithm was applied to images extracted from the DDSM and MIAS medical databases, and their results were compared with those obtained by similar works through the metrics identified during the research. The results obtained - 0.157 false negatives per image, 0.292 false positives per image, 97.7% sensitivity, 92.3% specificity, 94.4% accuracy and an area under the ROC curve of 0.75 - were similar to obtained by softmax (but less accurate) before the selection of features, and, therefore, were below the results presented by other similar works, but within the expected given the simplifications carried out during the development of the methodology. At the beginning of the work, the following research question was formulated: What is the behavior of a Support Vector Machines algorithm when being evaluated with other cases or with a different databas?

The technique used to eliminate the image background in original CNN was not able to provide good results in several images. It is understood that the threshold value used could not have been fixed, as the gray level present in the background of the images varies from case to case. Future work would need to use a more flexible technique to accomplish this task. The technique used to segment the pectoral muscle achieved good enough results to allow 97% of the images to be approved, but in many cases, the segmentation of the muscle was not accurate. The muscle often appears with a slight curve in the image, and the algorithm is not able to handle this curve. It is

recommended the study of other techniques, capable of greater precision even in these cases, or, alternatively, the study of ROI segmentation techniques that are not influenced by the presence of the pectoral muscle. convolutional Neural Networks provided good results for high-gloss masses, but segmented much more tissue than necessary in many cases.

Regardless of the resolution of this problem, it could be seen during the analysis of the results that the dependence of the CNN on the brightness of the lesions makes it ineffective for the segmentation of masses with obscured or ill-defined edges. It would also not be effective for detecting microcalcifications, as they are too small, and would not obtain good results for detecting architectural distortions, since they do not have solid and clear cores like masses.

5.2 FUTURE WORK

Finally, the following future works are suggested: A user interface for the DDSM image database that is capable of downloading and converting mammograms, displaying the lesion markings graphically, and, most importantly, passing on auxiliary data present in the database in a simpler format that can be introduced in other systems and analyzed more easily, in particular by researchers working for the Iraqi Ministry of Health; this is particularly important for researchers working for the Iraqi Ministry of Health.

Researching methods of region-of-interest (ROI) segmentation that do not require the removal of the pectoral muscle from the image (since this step is responsible for eliminating some lesions that are located in this region); Investigating methods of segmenting the pectoral muscle that are mordant. modifying the value of Z that is set as the default in a software for cellular neural networks so that it can respond in a dynamic manner to the characteristics of each image (for example, breast density and image histogram); modifying the value of Z that is set as default in a program that uses cellular neural networks so that it adapts to the characteristics of each image.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," CA Cancer J Clin, vol. 66, pp.7-30, 2016.

[2] Poorolajal, J., et al., Breast cancer screening (BCS) chart: a basic and preliminary model for making screening mammography more productive and efficient. J Public Health (Oxf), 2017: p. 1-8.

[3] Z. Jiao, X. GAO, Y. Wang, and J. Li, "A deep feature based frameworkfor breast masses classification," Journal of Neurocomputing, vol. 197, pp. 221-231, 2016.

[4] J. Arevalo, F. A. González, R. Ramos-Pollán, J. Oliveira, and M. A.Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," Comput. Methods Program Biomed, vol. 127, pp. 248-257, 2016.

[5] Deng, L., Yu, D.: Deep Learning: Methods and Applications. Now publishers, Boston (2014).

[6] Krizhevsky, A., I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems. 2012.

[7] Zhang, Y., et al. DeepSplice: Deep classification of novel splice junctions revealed by RNA-seq. in Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference on. 2016. IEEE

[8] Hua, K.-L., et al., Computer-aided classification of lung nodules on computed tomography images via deep learning technique. OncoTargets and therapy, 2015.

[9] Pravin S. Mane, Indra Gupta, M. K. Vasantha, "Implementation of RISC Processor on FPGA", Electrical Engineering Department, Indian Institute of Technology Roorkee, 2006 IEEE.

[10] Kui YI, Yue-Hua DING, “32-bit RISC CPU Based on MIPS”, International Joint Conference on Artificial Intelligence 2019.

[11] Mrs. Rupali S. Balpande, Mrs. Rashmi S. Keote, “Design of FPGA based Instruction Fetch and Decode Module of 32-bit RISC (MIPS) Processor”, International Conference on Communication Systems and Network echnologies, 2011.

[12] Mr. S. P. Ritpurkar, Prof. M. N. Thakare, Prof. G. D. Korde, “Synthesis and Simulation of a 32Bit MIPS RISC Processor using VHDL”, IEEE International Conference on Advances in Engineering & Technology Research, 2014.

[13] David A. Patterson, John L. Hennessy, “Computer Organization and Design”, the Hardware and Software Interface, Third Edition.

[14] J. C. Tobias Charistian Cahoon, Melanie A.Sutton, “Three-class mam mogram classification based on descriptive cnn features,” 2020.

[15] S. Sharma, M. Kharbanda, and G. Kaushal, “Brain tumor and breast cancer detection using medical images,” International Journal of Engi-neering Technology Science and Research, vol. 2, 2015.

[16] P. Hankare, K. Shah, D. Nair, and D. Nair, “Breast cancer detection using thermography,” Int. Res. J. Eng. Technol, vol. 4, 2016.

[17] S. Hwang and H.-E. Kim, “Self-transfer learning for fully weakly supervised object localization,” arXiv preprint arXiv: 1602.01625, 2016.

[18] Z. Mohammadzadeh, R. Safdari, M. Ghazisaeidi, S. Davoodi, and Z. Azadmanjir, “Advances in optimal detection of cancer by image processing; experience with lung and breast cancers,” Asian Pacific journal of cancer prevention: APJCP, vol. 16, no. 14, 2015.

[19] Ding, J.; Cheng, H.-D.; Huang, J.; Liu, J.; Zhang, Y. Breast ultrasound image classification based on multiple-instance learning. J. Digit. Imaging 2012, 25, 620–627.

[20] Bing, L.; Wang, W. Sparse representation based multi-instance learning for breast ultrasound image classification. Comput. Math. Methods Med. 2017, 2017, 7894705.

[21] Prabhakar, T.; Poonguzhali, S. Automatic detection and classification of benign and malignant lesions in breast ultrasound images using texture morphological and fractal features. In Proceedings of the 2017 10th Biomedical Engineering International Conference (BMEiCON), Hokkaido, Japan, 31 August–2 September 2017; IEEE: New York, NY, USA, 2017; pp. 1–5.

[22] Zhang, Q.; Suo, J.; Chang, W.; Shi, J.; Chen, M. Dual-modal computer-assisted evaluation of axillary lymph node metastasis in breast cancer patients on both real-time elastography and B-mode ultrasound. Eur. J. Radiol. 2017, 95, 66–74.

[23] GAO, Y.; Geras, K.J.; Lewin, A.A.; Moy, L. New frontiers: An update on computer-aided diagnosis for breast imaging in the age of artificial intelligence. Am. J. Roentgenol. 2019, 212, 300–307.

[24] Geras, K.J.; Mann, R.M.; Moy, L. Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives. Radiology 2019, 293, 246–259.

[25] Fujioka, T.; Mori, M.; Kubota, K.; Oyama, J.; Yamaga, E.; Yashima, Y.; Katsuta, L.; Nomura, K.; Nara, M.; Oda, G. The utility of deep learning in breast ultrasonic imaging: A review. Diagnostics 2020, 10, 1055.

[26] Zahoor, S.; Lali, I.U.; Javed, K.; Mehmood, W. Breast cancer detection and classification using traditional computer vision techniques: A comprehensive review. Curr. Med. Imaging 2020, 16, 1187–1200.

[27] Kadry, S.; Rajinikanth, V.; Taniar, D.; Damaševiˇcius, R.; Valencia, X.P.B. Automated segmentation of leukocyte from hematological images—A study using various CNN schemes. J. Supercomput. 2021, 1–21.

[28] Abayomi-Alli, O.O.; Damaševi˘cius, R.; Misra, S.; Maskeliunas, R.; Abayomi-Alli, A. Malignant skin melanoma detection using ¯ image augmentation by oversampling in nonlinear lower-dimensional embedding manifold. Turk. J. Electr. Eng. Comput. Sci. 2021, 29, 2600–2614.

[29] Maqsood, S.; Damaševi˘cius, R.; Maskeliunas, R. Hemorrhage detection based on 3d cnn deep learning framework and feature ¯ fusion for evaluating retinal abnormality in diabetic patients. Sensors 2021, 21, 3865.

[30] Hussain, N.; Kadry, S.; Tariq, U.; Mostafa, R.R.; Choi, J.-I.; Nam, Y. Intelligent Deep Learning and Improved Whale Optimization Algorithm Based Framework for Object Recognition. Hum. Cent. Comput. Inf. Sci. 2021, 11, 34.

[31] . Kadry, S.; Parwekar, P.; Damaševi˘cius, R.; Mehmood, A.; Khan, J.A.; Naqvi, S.R.; Khan, M.A. Human gait analysis for osteoarthritis prediction: A framework of deep learning and kernel extreme learning machine. Complex Intell. Syst. 2021, 1–19.

[32] Dhungel, N.; Carneiro, G.; Bradley, A.P. The automated learning of deep features for breast mass classification from mammograms. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016; Springer: Cham, Switzerland, 2016; pp. 106–114.

[33] Alhaisoni, M.; Tariq, U.; Hussain, N.; Majid, A.; Damaševi˘cius, R.; Maskeliunas, R. COVID-19 Case Recognition from Chest CT ¯ Images by Deep Learning, Entropy-Controlled Firefly Optimization, and Parallel Feature Fusion. Sensors 2021, 21, 7286.

[34] Odusami, M.; Maskeliunas, R.; Damaševi˘cius, R.; Krilavi˘cius, T. Analysis of features of alzheimer's disease: Detection of early ¯ stage from functional brain changes in magnetic resonance images using a finetuned resnet18 network. Diagnostics 2021, 11, 1071.

[35] Nawaz, M.; Nazir, T.; Masood, M.; Mehmood, A.; Mahum, R.; Kadry, S.; Thinnukool, O. Analysis of Brain MRI Images Using Improved CornerNet Approach. Diagnostics 2021, 11, 1856.

[36] Farzaneh, N.; Williamson, C.A.; Jiang, C.; Srinivasan, A.; Bapuraj, J.R.; Gryak, J.; Najarian, K.; Soroushmehr, S. Automated segmentation and severity analysis of subdural hematoma for patients with traumatic brain injuries. Diagnostics 2020, 10, 773.

[37] Meng, L.; Zhang, Q.; Bu, S. Two-Stage Liver and Tumor Segmentation Algorithm Based on Convolutional Neural Network. Diagnostics 2021, 11, 1806.

[38] Khaldi, Y.; Benzaoui, A.; Ouahabi, A.; Jacques, S.; Taleb-Ahmed, A. Ear recognition based on deep unsupervised active learning. IEEE Sens. J. 2021, 21, 20704–20713.

[39] Majid, A.; Nam, Y.; Tariq, U.; Roy, S.; Mostafa, R.R.; Sakr, R.H. COVID19 classification using CT images via ensembles of deep learning models. Comput. Mater. Contin. 2021, 69, 319–337.

[40] Sharif, M.I.; Alhussein, M.; Aurangzeb, K.; Raza, M. A decision support system for multimodal brain tumor classification using deep learning. Complex Intell. Syst. 2021, 1–14.

[41] Liu, D.; Liu, Y.; Li, S.; Li, W.; Wang, L. Fusion of handcrafted and deep features for medical image classification. J. Phys. Conf. Ser. 2019, 1345, 022052.

[42] Alinsaif, S.; Lang, J.; Alzheimer's Disease Neuroimaging Initiative. 3D shearlet-based descriptors combined with deep features for the classification of Alzheimer's disease based on MRI data. Comput. Biol. Med. 2021, 138, 104879.

[43] Khan, M.A.; Muhammad, K.; Sharif, M.; Akram, T.; de Albuquerque, V.H.C. Multi-Class Skin Lesion Detection and Classification via Teledermatology. IEEE J. Biomed. Health Inform. 2021, 25, 4267–4275.

[44] Masud, M.; Rashed, A.E.E.; Hossain, M.S. Convolutional neural network-based models for diagnosis of breast cancer. Neural Comput. Appl. 2020, 1–12.

[45] Jiménez-Gaona, Y.; Rodríguez-Álvarez, M.J.; Lakshminarayanan, V. Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. Appl. Sci. 2020, 10, 8298.

[46] Zeebaree, D.Q. A Review on Region of Interest Segmentation Based on Clustering Techniques for Breast Cancer Ultrasound Images. J. Appl. Sci. Technol. Trends 2020, 1, 78–91.

[47] Huang, K.; Zhang, Y.; Cheng, H.; Xing, P. Shape-adaptive convolutional operator for breast ultrasound image segmentation. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.

[48] Sadad, T.; Hussain, A.; Munir, A.; Habib, M.; Ali Khan, S.; Hussain, S.; Yang, S.; Alawairdhi, M. Identification of breast malignancy by marker-controlled watershed transformation and hybrid feature set for healthcare. Appl. Sci. 2020, 10, 1900.

[49] Mishra, A.K.; Roy, P.; Bandyopadhyay, S.; Das, S.K. Breast ultrasound tumour classification: A Machine Learning—Radiomics based approach. Expert Syst. 2021, 38, e12713. [CrossRef]

[50] Hussain, S.; Xi, X.; Ullah, I.; Wu, Y.; Ren, C.; Lianzheng, Z.; Tian, C.; Yin, Y. Contextual level-set method for breast tumor segmentation. IEEE Access 2020, 8, 189343–189353.

[51] Xiangmin, H.; Jun, W.; Weijun, Z.; Cai, C.; Shihui, Y.; Jun, S. Deep Doubly Supervised Transfer Network for Diagnosis of Breast Cancer with Imbalanced Ultrasound Imaging Modalities. arXiv 2020, arXiv:2007.066342020.

[52] Moon, W.K.; Lee, Y.-W.; Ke, H.-H.; Lee, S.H.; Huang, C.-S.; Chang, R.-F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Comput. Methods Programs Biomed. 2020, 190, 105361.

[53] Byra, M.; Jarosik, P.; Szubert, A.; Galperin, M.; Ojeda-Fournier, H.; Olson, L.; O'Boyle, M.; Comstock, C.; Andre, M. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. Biomed. Signal Process. Control 2020, 61, 102027.

[54] Kadry, S.; Damaševiˇcius, R.; Taniar, D.; Rajinikanth, V.; Lawal, I.A. Extraction of tumour in breast MRI using joint thresholding and segmentation–A study. In Proceedings of the 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 25–27 March 2021; IEEE: New York, NY, USA, 2021; pp. 1–5.

[55] Lahoura, V.; Singh, H.; Aggarwal, A.; Sharma, B.; Mohammed, M.; Damaševiˇcius, R.; Kadry, S.; Cengiz, K. Cloud computing-based framework for breast cancer diagnosis using extreme learning machine. Diagnostics 2021, 11, 241.

[56] Maqsood, S.; Damasevicius, R.; Shah, F.M. An Efficient Approach for the Detection of Brain Tumor Using Fuzzy Logic and U-NET CNN Classification, International Conference on Computational Science and Its Applications; Springer: Cham, Switzerland, 2021; pp. 105–118.

[57] Rajinikanth, V.; Kadry, S.; Taniar, D.; Damasevicius, R.; Rauf, H.T. Breast-cancer detection using thermal images with marinepredators-algorithm selected features. In Proceedings of the 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Noida, India, 26–27 August 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.

[58] Khaldi, Y.; Benzaoui, A.; Ouahabi, A.; Jacques, S.; Taleb-Ahmed, A. Ear recognition based on deep unsupervised active learning. IEEE Sens. J. 2021, 21, 20704–20713.

[59] Majid, A.; Nam, Y.; Tariq, U.; Roy, S.; Mostafa, R.R.; Sakr, R.H. COVID19 classification using CT images via ensembles of deep learning models. Comput. Mater. Contin. 2021, 69, 319–337.

[60] Sharif, M.I.; Alhussein, M.; Aurangzeb, K.; Raza, M. A decision support system for multimodal brain tumor classification using deep learning. Complex Intell. Syst. 2021, 1–14.

[61] Liu, D.; Liu, Y.; Li, S.; Li, W.; Wang, L. Fusion of handcrafted and deep features for medical image classification. J. Phys. Conf. Ser. 2019, 1345, 022052.

[62] Alinsaif, S.; Lang, J.; Alzheimer's Disease Neuroimaging Initiative. 3D shearlet-based descriptors combined with deep features for the classification of Alzheimer's disease based on MRI data. Comput. Biol. Med. 2021, 138, 104879.

[63] Khan, M.A.; Muhammad, K.; Sharif, M.; Akram, T.; de Albuquerque, V.H.C. Multi-Class Skin Lesion Detection and Classification via Teledermatology. IEEE J. Biomed. Health Inform. 2021, 25, 4267–4275.

[64] Masud, M.; Rashed, A.E.E.; Hossain, M.S. Convolutional neural network-based models for diagnosis of breast cancer. Neural Comput. Appl. 2020, 1–12.

[65] Jiménez-Gaona, Y.; Rodríguez-Álvarez, M.J.; Lakshminarayanan, V. Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. Appl. Sci. 2020, 10, 8298.

[66] Zeebaree, D.Q. A Review on Region of Interest Segmentation Based on Clustering Techniques for Breast Cancer Ultrasound Images. J. Appl. Sci. Technol. Trends 2020, 1, 78–91.

[67] Huang, K.; Zhang, Y.; Cheng, H.; Xing, P. Shape-adaptive convolutional operator for breast ultrasound image segmentation. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.

[68] Sadad, T.; Hussain, A.; Munir, A.; Habib, M.; Ali Khan, S.; Hussain, S.; Yang, S.; Alawairdhi, M. Identification of breast malignancy by marker-controlled watershed transformation and hybrid feature set for healthcare. Appl. Sci. 2020, 10, 1900.