Mustafa BIÇAKCI

A Ph.D. Thesis

AGU 2021

NON-SMALL CELL LUNG CANCER TUMOR CHARACTERISATION USING DEEP LEARNING

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Mustafa BIÇAKCI
March 2021

NON-SMALL CELL LUNG CANCER TUMOR CHARACTERISATION USING DEEP LEARNING

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Mustafa BIÇAKCI
March 2021

SCIENTIFIC ETHICS COMPLIANCE

I hereby declare that all information in this document has been obtained in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name-Surname: Mustafa BIÇAKCI
Signature :

REGULATORY COMPLIANCE

Ph.D. thesis titled Non-Small Cell Lung Cancer Tumor Characterisation Using Deep Learning has been prepared in accordance with the Thesis Writing Guidelines of the Abdullah Gül University, Graduate School of Engineering and Science.

Prepared By	Advisor
Mustafa BIÇAKCI	Prof. Bülent YILMAZ

Head of the Electrical and Computer Engineering Program
Assoc. Prof. Kutay İÇÖZ

ACCEPTANCE AND APPROVAL

Ph.D. thesis titled Non-Small Cell Lung Cancer Tumor Characterisation Using Deep Learning prepared by Mustafa BIÇAKCI has been accepted by the jury in the Electrical and Computer Engineering Graduate Program at Abdullah Gül University, Graduate School of Engineering and Science.

/ / 2021

JURY:

Advisor: Prof. Bülent YILMAZ

Member: Prof. Alper BAŞTÜRK

Member: Prof. Murat EKİNCİ

Member: Assoc. Prof. Zafer AYDIN

Member: Assist. Prof. Kasım TAŞDEMİR

APPROVAL:

The acceptance of this Ph.D. thesis has been approved by the decision of the Abdullah Gül University, Graduate School of Engineering and Science, Executive Board dated ….. /….. / …….. and numbered ……..……. .

…….. /…….. / ……..

Graduate School Dean

Prof. Hakan USTA

ABSTRACT
NON-SMALL CELL LUNG CANCER TUMOR CHARACTERISATION USING DEEP LEARNING

Mustafa BIÇAKCI
Ph.D. in Electrical and Computer Engineering
Advisor: Prof. Bülent YILMAZ
March 2021

Non-small cell lung cancer (NSCLC) constitutes the vast majority of lung cancers and has two major subtypes, adenocarcinoma (ADC) and squamous cell carcinoma (SqCC). Generally, these two subtypes are distinguished from each other by considering microscopic morphological criteria. However, poor morphology makes this quite difficult. Such studies are important for subspecialty treatment methods. In this thesis, deep learning (DL) methods on the subtype classification of NSCLC were investigated.
In the first study, 73% success rate was achieved by using artificial neural networks (ANN), which form the basis of DL methods. In the second study, several DL models were investigated on subtype classification using segmented tumor slices from PET images. As a result, VGG16 and VGG19 emerged as the most successful models with a 95% F-score. Later, slice based studies were abandoned and patient based studies were initiated. In the third study, the use of three-dimensional (3D) data created by combining slices from each patient was not successful. In the fourth study, three different experiments were conducted in which PET images were directly used, cropped to include peritumoral areas, and segmented only tumor parts. This study demonstrated the positive effect of peritumoral areas and VGG19 reached an F-score of 74%. In the fifth study, transfer learning and fine tuning works did not yield successful results. The latest study involving CNN-based and ResNet-based shallow networks was promising with an F-score of 71%.

Keywords: Lung cancer, subtype classification, PET imaging, deep learning, convolutional neural network

ÖZET

DERİN ÖĞRENME YAKLAŞIMLARIYLA KÜÇÜK HÜCRELİ DIŞI AKCİĞER KANSERİNDE TÜMÖR KARAKTERİZASYONU

Mustafa BIÇAKCI
Elektrik ve Bilgisayar Mühendisliği Anabilim Dalı Doktora
Tez Yöneticisi**:** Prof. Dr. Bülent YILMAZ
Mart-2021

Küçük Hücreli Dışı Akciğer Kanseri (KHDAK) akciğer kanserlerinin büyük çoğunluğunu oluşturur ve adenokarsinom (ADC) ve skuamöz hücreli karsinom (SqCC) olmak üzere iki önemli alt tipi vardır. Genel olarak, bu iki alt tip mikroskobik olarak belirlenen morfolojik kriterler dikkate alınarak birbirinden ayrılır. Ancak, kötü morfoloji bunu oldukça zorlaştırır. Alt tipe özel tedavi yöntemleri için bu tür çalışmalar önemlidir. Bu tezde, pozitron emisyon tomografi (PET) görüntüleri kullanılarak KHDAK'nin alt tiplerinin sınıflandırılması üzerinde derin öğrenme (DÖ) yöntemleri incelenmiştir. İlk çalışmada, DÖ yöntemlerinin temelini oluşturan yapay sinir ağları (YSA) kullanılarak %73 doğru sınıflandırma başarısı elde edilmiştir. İkinci çalışmada, PET görüntülerinden alınan bölütlenmiş tümör kesitleri kullanılarak birkaç DÖ modeli incelenmiştir. Sonuçta, %95 F skoru ile VGG16 ve VGG19 en başarılı modeller olmuştur. Bu çalışmanın sonunda kesit bazlı çalışmalar bırakılarak hasta bazlı çalışmalara geçilmiştir. Üçüncü çalışmada, hasta bazlı dilimlerin birleştirilmesiyle oluşturulan üç boyutlu (3B) verilerin kullanımı yeterli başarıyı sağlamamıştır. Dördüncü çalışmada, PET görüntülerinin doğrudan kullanıldığı, tümör kısımlarının kırpılarak kullanıldığı ve bölütlenmiş tümör parçalarının kullanıldığı üç farklı deney yapılmıştır. Bu çalışma, peritümoral alanların sınıflandırmada olumlu etkisini ortaya koymuş ve VGG19 %74 F skoru değerine ulaşmıştır. Beşinci çalışmada, transfer öğrenme ve hassas ayar çalışmaları başarısızdı. CNN ve ResNet tabanlı sığ ağları içeren son çalışma %71 F skoru ile umut verici olmuştur.

Anahtar kelimeler: Akciğer kanseri, alt tip sınıflandırma, PET görüntüleme, derin öğrenme, evrişimli sinir ağı,

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Bülent YILMAZ very much. I am grateful for his contribution to my thesis as well as all his advice and guidance for me to be a good researcher.

I would like to thank my dear wife İlknur BIÇAKCI, who supported me in every aspect of my life. I am very grateful to her for always providing me with spiritual strength during my doctoral studies.

I would also like to thank Prof. Ulaş BAĞCI, who provided some key advice during my doctoral studies and never withheld his support.

Finally, I would like to thank my precious extended family, especially my mother, father and sister, who are always by my side during my doctoral studies as in all areas of my life.

Mainly, all genuine gratitude and thanks to my Lord.
"All the praise offered by all beings from pre-eternity to post-eternity is due to Him alone!" -The Letters

Table of Contents

1. INTRODUCTION 1

1.1 GENERAL 1
1.2 OBJECTIVES AND SCOPE 2

2. BACKGROUND 4

2.1 ANATOMY OF THE RESPIRATORY SYSTEM 4
2.2 LUNG CANCERS 4
2.3 IMAGING TECHNIQUES IN LUNG CANCER 6
2.3.1 Chest Radiography 6
2.3.2 Computed Tomography 7
2.3.3 Positron Emission Tomography 8
2.3.4 PET / CT Imaging 9
2.4 SUBTYPE CLASSIFICATION OF NON-SMALL CELL LUNG CANCER IN POSITRON EMISSION TOMOGRAPHY IMAGING 10
2.4.1 Feature Extraction 12
2.4.2 Image Classification 20
2.4.3 Deep Learning Models 21
2.4.4 Hyperparameter Optimization 25

3. STUDY-1 28

3.1 CLASSIFICATION OF NON-SMALL CELL LUNG CANCER SUBTYPES WITH ARTIFICIAL NEURAL NETWORKS 28
3.2 MATERIALS AND METHODS 29
3.3 RESULTS AND DISCUSSION 34

4. STUDY-2 35

4.1 DEEP LEARNING METHODS IN SLICE-BASED SUBTYPE CLASSIFICATION OF NON-SMALL CELL LUNG CANCER 35
4.2 MATERIALS AND METHODS 35
4.3 RESULTS AND DISCUSSION 39

5. STUDY-3 41

5.1 STUDIES ON 3D CNN MODELS USING 3D PATIENT DATA 41
5.2 MATERIALS AND METHODS 41
5.3 RESULTS AND DISCUSSION 43

6. STUDY-4 44

6.1 METABOLIC IMAGING BASED SUB-CLASSIFICATION OF LUNG CANCER 44
6.2 MATERIALS AND METHODS 45
6.3 RESULTS AND DISCUSSION 49

7. STUDY-5 53

7.1 TRANSFER LEARNING STUDIES ON SUBTYPE CLASSIFICATION OF NSCLC 53

7.2 MATERIALS AND METHODS 54
7.3 RESULTS AND DISCUSSION.....57

8. STUDY-6.....59

8.1 SHALLOW NETWORK STUDIES ON SUBTYPE CLASSIFICATION OF NSCLC.....59
8.2 MATERIALS AND METHODS 59
8.3 RESULTS AND DISCUSSION.....60

9. CONCLUSIONS AND FUTURE PROSPECTS.....62

9.1 CONCLUSIONS 62
9.2 SOCIETAL IMPACT AND CONTRIBUTION TO GLOBAL SUSTAINABILITY 65
9.3 FUTURE PROSPECTS.....65

BIBLIOGRAPHY 67

List of Figures

Figure 2.1 Anatomy of the respiratory system 5
Figure 2.2 Chest radiography procedure 6
Figure 2.3 CT scan procedure 7
Figure 2.4 PET scanning 8
Figure 2.5 CT and PET images of a cancerous lung 9
Figure 2.1 CT (left), FDG-PET (middle) and FDG-PET / CT (right) images of a patient diagnosed with NSCLC 9
Figure 2.2 FDG-PET / CT of a patient with NSCLC 10
Figure 2.3 FDG-PET / CT of a patient with NSCLC (supraclavicular lymph node metastasis) 10
Figure 2.4 Various segmentation techniques 12
Figure 2.5 Schematic diagram of a convolutional neural network 15
Figure 2.6 Representation of input image and feature detector in the convolutional operation 15
Figure 2.7 First step of the convolution operation to generate a feature map 16
Figure 2.8 Generated feature map as a result of the convolution operations 16
Figure 2.9 Obtaining the convolutional layer 17
Figure 2.10 The rectified linear unit (ReLU) 17
Figure 2.11 Max pooling operation 18
Figure 2.12 Flattening operation 19
Figure 2.13 A representation of the fully-connected layer 19
Figure 2.14 Summary of operations in a CNN model 20
Figure 2.15 A representation of the MLP architecture 22
Figure 2.16 LeNet architecture 22
Figure 2.22 Schematic diagram of (a) VGG16 and (b) VGG19 models 23
Figure 2.23 Organization of convolution filters in the fire module of SqueezeNet 24
Figure 2.24 View of SqueezeNet architecture 24
Figure 3.1 Summary of the work from which the features were obtained 29
Figure 3.2 Block diagram representation of nervous system **Hata! Yer işareti tanımlanmamış.**
Figure 3.3 (a) Biological neuron and (b) artificial neuron 32
Figure 4.1 Segmented slices with a tumor obtained from one patient 37

Figure 4.2 Segmented slices obtained from different patients with ADC.....37
Figure 4.3 Segmented slices obtained from different patients with SqCC.....38
Figure 5.1 2D SqueezeNet architecture.....42
Figure 5.2 3D SqueezeNet architecture.....42
Figure 6.1 Images from different patients with (A) ADC and (B) SqCC.....45
Figure 6.2 Schematic representation of three experiments studied in this work.....46
Figure 6.3 The average accuracy and loss of training and validation in the 10-fold cross-validation for VGG19 model in the second experiment.....51
Figure 7.1 A representation of CNN model.....55
Figure 7.2 Fine tuning the last convolutional block of the VGG16 network.....56
Figure 8.1 CNN-S architecture.....60
Figure 8.2 RN-S architecture.....60

List of Tables

Table 3.1 Feature extraction methods and features extracted using these methods.....30
Table 3.2 Train and test success of training algorithms 34
Table 4.1 Hyperparameters.....39
Table 4.2 Classification success of the models.....40
Table 5.1 Resuls of 3D Models.....43
Table 6.1 Hyperparameters.....47
Table 6.2 Tuned CNN Architectures.....48
Table 6.3 Results from three experiments showing classification performance of the models 49
Table 6.4 Hyperparameter values yielding performances given in Table 3 51
Table 7.1 Results of transfer learning works 58
Table 8.1 Results of CNN-S model.....61
Table 8.2 Results of RN-S model.....61

List of Abbreviations

18F-FDG	[^{18}F] 2-Fluoro-2-Deoxy-D-Glucose
Adam	Adaptive Moment Estimation
ADC	Adenocarcinoma
AUC	Area Under Curve
ANN	Artificial Neural Networks
COPD	Chronic Obstructive Pulmonary Disease
CT	Computed Tomography
CNN	Convolutional Neural Networks
DL	Deep Learning
FANN	Fast Artificial Neural Network
GLCM	Gray-Level Co-Occurrence Matrix
GLSZM	Gray-Level Size Zone Matrix
GLRLM	Gray-Level Run-Length Matrix
KRTH	Kayseri Research and Training Hospital
LCC	Large Cell Carcinoma
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MLP	Multi-Layer Perceptron
NGTDM	Neighborhood Gray Tone Difference Matrix
NSCLC	Non-Small Cell Lung Cancer
PVC	Partial Volume Correction
PET	Positron Emission Tomography
ReLU	Rectified Linear Unit
RMSprop	Root Mean Square Propagation
SCLC	Small Cell Lung Cancer
SqCC	Squamous Cell Carcinoma
SGD	Stochastic Gradient Descent

To my lovely wife and daughter

Chapter 1

Introduction

1.1 General

The lungs are the main organ of the respiratory system. The human body needs oxygen to stay alive and healthy. In addition, the body needs to get rid of carbon dioxide, which is a cellular waste product. The lungs are specially designed to change these gases by breathing in and out. Therefore, the health of the whole body is dependent on the health of the lungs. There are many lung diseases such as asthma, chronic obstructive pulmonary disease (COPD), bronchitis, pneumonia, but the most dangerous is lung cancer. Smoking and air pollution are among the main causes of lung cancer, but lung cancer can also develop in those who are not exposed to these factors. According to global cancer research and statistics, lung cancer, the most common cause of cancer-related deaths worldwide, is a public health problem that is increasing in severity day by day. Worldwide in 2020, the estimated number of new cases of lung cancer was 2,206,771 and it is estimated that 1,796,144 deaths were due to lung cancer. According to the global estimates for the year 2020, among all types of cancer, the incidence of lung cancer is 11.4%, while the mortality rate comes first with 18% [1].

Histological type and tissue of origin are decisive in classifying lung cancer. This classification is important in defining the optimal treatment and determining the prognosis. There are two main types of lung cancer, non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC are classified into squamous cell carcinoma (SqCC), adenocarcinoma (ADC), and large cell carcinoma [2]. Common imaging modalities used in patients with NSCLC include chest radiography, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and fused PET / CT [3]. [^{18}F] 2-Fluoro-2-Deoxy-D-Glucose (18F-FDG) PET / CT is a non-invasive imaging modality which combines metabolic and anatomic information in the evaluation of cancer. PET / CT imaging method is useful in staging, treatment

response evaluation and radiotherapy planning of lung cancer. PET scanners have multiple detectors arranged in a ring around the patient, allowing simultaneous gamma ray detection. Patient motion, which is one of many factors that can affect image quality, can cause potential spatial incompatibility between CT and PET. Respiratory movement that can occur, especially when viewing the chest or abdomen, can cause errors. PET data is acquired in minutes, covering many respiratory cycles, and may cause blurring of the resulting images [2]. In this thesis, subtype classification of NSCLC was made using PET images from PET / CT scans. While doing this, deep learning methods were investigated on the differentiation of ADC and SqCC, which are the main subtypes of NSCLC.

1.2 Objectives and Scope

PET / CT is a combination of positron emission tomography and computed tomography. PET gives a picture of the physiological processes in the body. By combining PET device with CT, anatomical and physiological information can be obtained in a single session. In this thesis, only the PET images obtained as a result of PET / CT scanning were considered. Although there are many studies on PET images in the literature, it is striking that there are no deep learning studies on automatic classification of NSCLC subtypes using this image type. In this thesis, on the automatic classification of NSCLC subtypes, various deep learning approaches, which are modern machine learning methods, were investigated. In addition, this study includes some preprocessing studies on PET slices obtained from PET / CT images provided by Nuclear Medicine Department of Acıbadem Kayseri Hospital. These PET slices were images containing the whole lung taken from different angles of each patient, and there were different numbers of slices for each patient. During the classification studies, it was aimed to compare the success of deep learning methods as well as to investigate the effect of peritumoral areas on classification. In addition, the effects of various regularization and optimization techniques used with these models were also revealed.

This thesis covers 6 different approaches on subtype classification. The first examines the classification of 39 features by artificial neural networks (ANN). These features were obtained from PET images in a previous study [4] conducted to investigate some classical machine learning methods. Secondly, a slice-based study was conducted using tumor images segmented from PET slices by random walk method. Although quite successful results were obtained, at the end of this study, the data set was split on a patient

basis and the study was continued. The purpose in doing this was to provide more appropriate progress to clinical studies. In the third study, patient-based stacked slices were used as the dataset. Here, three-dimensional (3D) data were obtained by combining all slices belonging to each patient. After that, the 3D versions of the models were created and classification studies were carried out. The fourth study is a comprehensive study [5] that includes three different experiments and contributes to the literature by revealing the positive effect of peritumoral areas. As the fifth study, transfer learning and fine-tuning studies were carried out. In this study, pre-trained models for different tasks such as object detection were examined on the medical dataset we have. In the last study, two shallow networks were created and successful results were tried to be obtained with the little data we have. All these studies include also many additional studies including various regularization, optimization and augmentation techniques.

Chapter 2

Background

2.1 Anatomy of The Respiratory System

The lungs are a pair of respiratory organs located in the chest. While the lungs carry oxygen to the body during breathing in, they release carbon dioxide from the body during breathing out. The right and left lungs each have lobes. While the left lung has two lobes, the slightly larger right lung has three lobes. The bronchi diverge from the trachea (windpipe) into the right and left lungs. The bronchi are sometimes involved in lung cancer. Inside the lungs are air sacs called alveoli and tubes called bronchioles. Figure 2.1 shows the anatomy of the respiratory system. This figure shows the lungs, trachea, airways, lymph nodes, and diaphragm. Oxygen taken into the lungs while breathing passes through the thin membranes of the alveoli into the bloodstream. The pleura, a thin membrane, covers the outside of the lungs and the inner wall of the chest cavity, forming a sac. This sac, called the pleural space, contains a small amount of fluid. This fluid helps the lungs move smoothly while breathing [6].

2.2 Lung Cancers

There are many lung diseases such as asthma, chronic obstructive pulmonary disease (COPD), bronchitis, pneumonia, but the most dangerous is lung cancer. Cancer is a disease known as the spread of some cells in the body to surrounding tissues by dividing uncontrollably and without stopping. As a result of this uncontrolled spread, tissue masses called tumors may occur. Cancer can start anywhere in the human body, and if that location is the lung, it is called lung cancer.

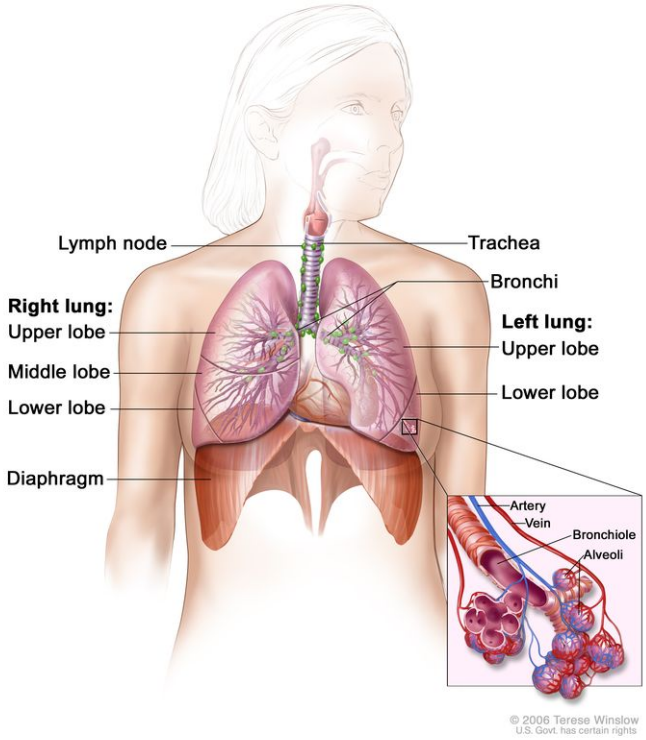

Figure 2.17 Anatomy of the respiratory system [6]

There are two main types of lung cancer, non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC is more common than SCLC. NSCLC are classified into squamous cell carcinoma (SqCC), adenocarcinoma (ADC), and large cell carcinoma [2]. Within the scope of this thesis, ADC and SqCC, which are the main two subtypes, have been dealt with and the automatic classification of these two subtypes has been studied. The most important risk factor for lung cancer is smoking, pipe or cigar smoking. Apart from this, many factors can be counted, such as radiation therapy, air pollution, and exposure to substances such as asbestos and arsenic. When smoking and these factors combine, the risk increases even more [6]. After some tests and procedures performed for the detection and diagnosis of NSCLC, a biopsy is performed if lung cancer is suspected. The process used to find out if the cancer has spread is called staging. It is important to determine the stage so that the treatment can be planned correctly. The main imaging modalities used for diagnosing and staging NSCLC are MRI, CT and PET.

2.3 Imaging Techniques in Lung Cancer

Conventional chest radiography, computed tomography (CT), magnetic resonance (MR) and positron emission tomography (PET) are the main imaging techniques used in the detection, characterization, staging and follow-up of lung cancer. CT is successful in matters such as tumor size, localization, detection of lymph nodes and metastatic disease. PET / CT, which means combining anatomical information from CT and metabolic information from PET, is one of the best imaging techniques for staging lung cancer. MR imaging is considered promising in the evaluation of lung cancer patients [7]. This thesis covers studies using only PET images obtained from PET / CT scans.

2.3.1 Chest Radiography

The first imaging technique in lung cancer screening is chest radiography. Chest radiography, known as the most commonly used diagnostic imaging test, has the advantages of being simple, low cost and low radiation [8]. This method can also be used in defining lung lesions as benign or malignant. In addition, this method can be used to see some features about staging, but it cannot detect invasion and nodal involvement of the chest wall, diaphragm, and mediastinum [9]. The purpose of staging is to determine the degree of the disease and to select patients who will benefit from the surgery. Figure 2.2 shows the procedure for chest radiography.

Figure 2.18 Chest radiography procedure [10]

2.3.2 Computed Tomography

Computed tomography (CT) is an imaging technique that creates cross-sectional images of the body with more detailed information than conventional X-rays by rotating X-rays rapidly around the body. These tomographic cross-sectional images, called slices, can be collected and stacked consecutively to obtain a three-dimensional (3D) image of the patient. This is to make it easier to identify and locate abnormalities or tumor tissues. In CT scanners, the X-ray tube is not fixed, they use a rotating X-ray source with a motor. During the CT scan, the patient lies in a slowly moving bed. Meanwhile, the X-ray tube rotating around the patient shoots X-rays through the body. X-rays are taken by special digital detectors and transmitted to the computer while leaving the patient. As the X-ray source completes a full rotation, a 2D image slice of the patient is created. Then, the bed is gradually shifted to get a new slice with a new scan. This process continues until the desired number of slices is obtained. The image slices can be kept separately or combined to obtain 3D images. In this way, in addition to certain structures in the body, clear visualization and localization of abnormalities can be achieved. CT scans can be used to view disease, tumors, bleeding, clots, and abnormalities in different parts of the body. CT scans, which are more detailed than conventional x-rays in imaging bone, joint, cartilage or tendons, are frequently used to reveal the size, shape and location of the lung tumor and to find disseminated cancer tissues [11] [12]. Figure 2.3 shows the CT scan procedure.

Figure 2.19 CT scan procedure [13]

2.3.3 Positron Emission Tomography

Although conventional imaging methods such as chest radiography and computed tomography (CT) are useful in the diagnosis and staging of lung cancer, they are not accurate enough. Positron emission tomography (PET) is an imaging modality that involves the procedure of injecting a small amount of radioactive glucose into a vein. The purpose of using this radioactive glucose, called [^{18}F] 2-Fluoro-2-Deoxy-D-Glucose (18F-FDG), is to reveal the differences between glucose metabolism of tissues. Because malignant cancer cells are more active, they take up more glucose than normal cells. Making a picture of the glucose utilization of the tissues, by rotating around the body, the PET scanner more brightly displays the parts that consume glucose more, namely malignant tissues. PET scanning, which enables successful detection of metabolic and physiological abnormalities in tumor imaging, gives more accurate results than CT and magnetic resonance imaging (MRI) in staging lung cancer and evaluating therapeutic response [6][14-17]. Figure 2.4 shows the procedure for PET scanning. Figure 2.5 shows CT and PET images of a cancerous lung [18].

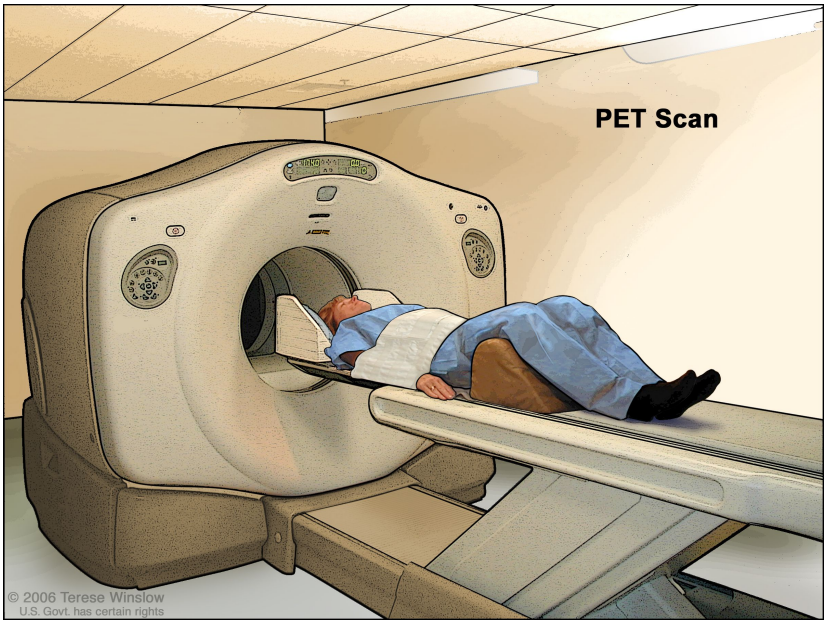

Figure 2.20 PET scanning [6]

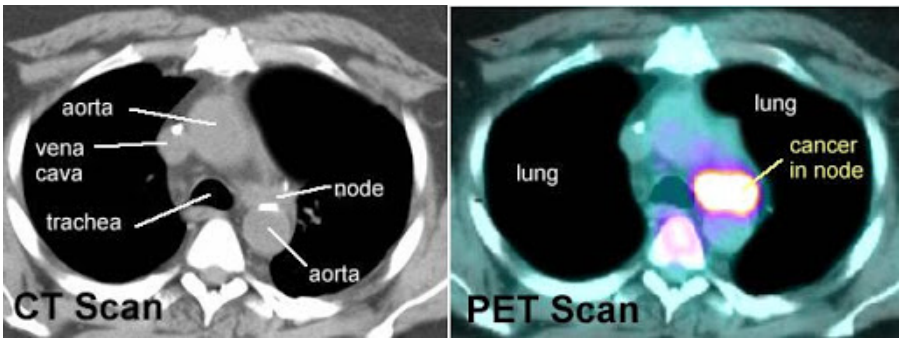

Figure 2.21 CT and PET images of a cancerous lung [18]

2.3.4 PET / CT Imaging

PET / CT imaging, which refers to the combination of CT and PET techniques in order to see anatomical and metabolic details clearly, provides sharper and better definition and localization of the lesions. CT scanning provides anatomical details with 2D or 3D images. However, despite its improved image quality, it may not provide enough accurate information. PET / CT has a higher sensitivity than PET alone. In addition, tumor evaluation can be difficult because PET images are limited in terms of anatomical information. Combining CT and PET, provides easier and faster diagnosis and detection since it includes metabolic information. It is also a very successful imaging technique for lung cancer staging [7]. Figure 2.6 shows CT (left), FDG-PET (middle) and FDG-PET / CT images of an individual diagnosed with NSCLC. As indicated by the arrow in the figure, FDG-PET / CT aids in the diagnosis of a vital infracarinal metastasis. Figure 2.7 shows FDG-PET / CT of a patient with NSCLC. FDG-PET / CT provides the most accurate differentiation of tumor tissue. Figure 2.8 shows FDG-PET / CT of another patient with NSCLC. Here, a 0.5 cm diameter supraclavicular lymph node metastasis is seen on PET / CT, which is not visible on CT alone and is uncertain on FDG-PET [19].

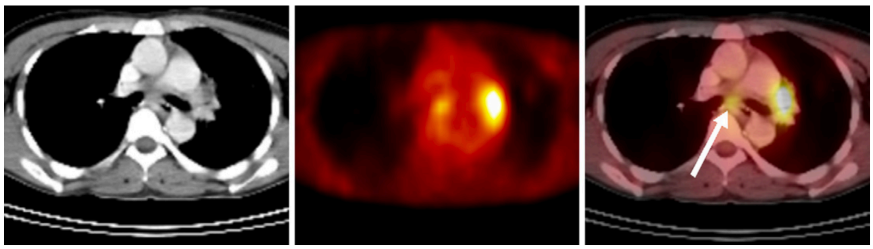

Figure 2.22 CT (left), FDG-PET (middle) and FDG-PET / CT (right) images of a patient diagnosed with NSCLC [19]

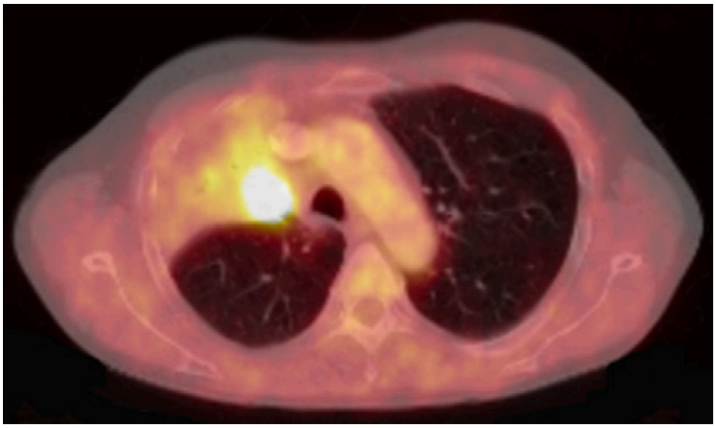

Figure 2.23 FDG-PET / CT of a patient with NSCLC [19]

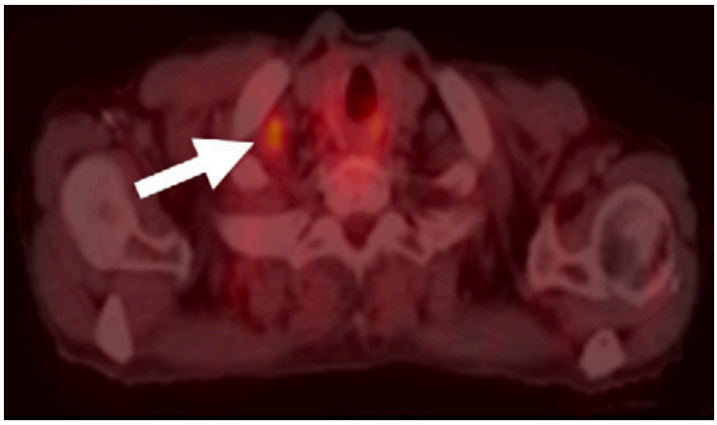

Figure 2.24 FDG-PET / CT of a patient with NSCLC (supraclavicular lymph node metastasis) [19]

2.4 Subtype Classification of Non-Small Cell Lung Cancer in Positron Emission Tomography Imaging

84% of lung cancers are non-small cell lung cancer (NSCLC) [20]. Adenocarcinoma (ADC) and squamous cell carcinoma (SqCC) are the two major subtypes of NSCLC. While 40% of lung cancers are ADC, 25-30% are of SqCC [21]. Understanding the effects of cytotoxic and biological agents suggests that subtype-specific treatment methods may be developed in the future. From this point of view, it shows the importance of the differentiation of ADC and SqCC [22]. Small lung biopsies (bronchoscopic, needle or core biopsies) and cytology specimens are used for lung cancer diagnosis. In general, standard morphological criteria by routine microscopy are essential to differentiate these two subtypes. However, bad morphology, especially in small

specimens, can sometimes lead to difficult differentiation of tumors. On the other hand, sampling error that may occur during the biopsy procedure would be problematic in representing intratumoral heterogeneity.

Positron emission tomography (PET) is a functional imaging approach that has been widely used in medicine, and provides significant diagnostic benefits. It is also a very effective and efficient method for staging and therapy of tumors. According to radiomics, medical images carry more information than obtained by visual examination [23]. With high-resolution PET images, it has become possible to obtain inferences with image processing methods. In this regard, it has become possible to recognize tumor characteristics using PET images.

Machine learning examines the algorithms that recognize complex patterns, and make predictions from the available data (such as the medical images in our case) to come up with intelligent decisions. In oncology, machine learning methods are used in different applications such as cancer prognosis and prediction [24], survival analysis [25], drug response [26], gene expression [27] and subtype differentiation [4]. As a machine learning methodology, artificial neural networks (ANN) based approaches involve learning and prediction algorithms that mimic the human brain in terms of recognition and decision-making. ANN methods are used in oncology in different applications such as tumor detection [28] and diagnosis [29]. Deep learning (DL) is an advanced neural network type with more layers to allow higher levels of abstraction. Deep convolutional neural networks (CNN) have brought breakthroughs in image-based studies [30]. They are highly successful in solving difficult problems such as recognizing objects in real world images [31], [32]. In recent years, there have been increasing number of cancer related studies using deep learning such as cancer detection and gene identification [33], skin cancer classification [34], histopathological diagnosis [35]. In addition, there are several studies [36], [37] showing that deep learning approaches are more effective and successful when compared to other machine learning methods in cancer-related classification problems. Diagnosis and classification studies using CT images [38], [39] are also available for lung cancer. However, CT images do not reflect the metabolic and heterogeneity information about the tumor which is highly critical in tumor subtype determination and is accessible when PET imaging is used. We should note that there are limited number of DL-based studies using PET images aiming lung cancer diagnosis [40]. Especially, subtype classification in NSCLC has not been explored using PET images and DL.

2.4.1 Feature Extraction

A computer diagnostic analyzer is used to provide information that will help radiologists make decisions. However, such information is difficult to interpret due to difficulties such as poor morphology. For this reason, various techniques are used in order for the radiologist to interpret and classify the images correctly. The features obtained from the image are used to classify them correctly. Here, it is necessary to mention feature extraction methods. But before this, segmentation methods are used to segment tumor-like structures in medical images and to serve feature extraction methods afterwards. Since manual segmentation is difficult and takes long time, computer-aided segmentation techniques are generally investigated. Various segmentation methods developed over the years have emerged. Figure 2.9 shows the diagram in which these methods are grouped under some headings [41].

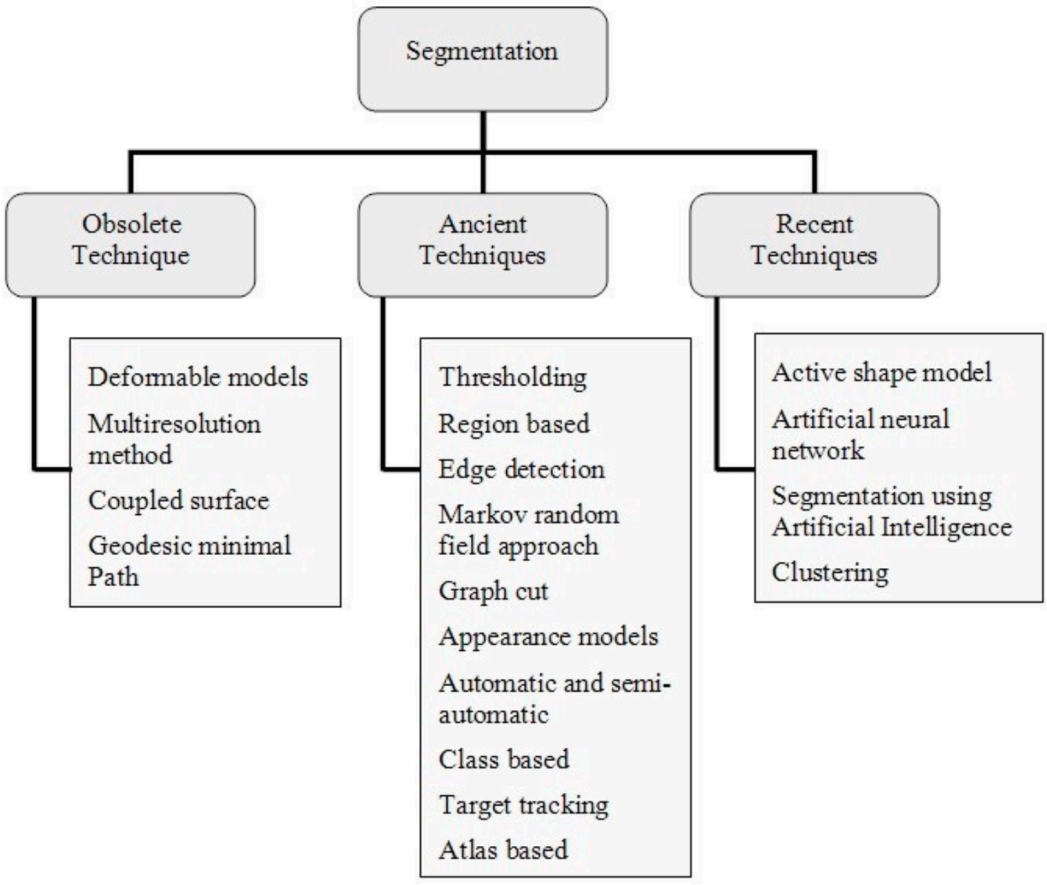

Figure 2.25 Various segmentation techniques [41]

The purpose of segmentation is basically to segment the image into regions based on attributes. By applying segmentation, the region of interest is defined in order to explain the anatomical structures on the medical image. Obsolete segmentation techniques, developed in the late 1990s, are mostly no longer used. Ancient techniques are also old, but still used in many areas. Recent segmentation techniques, which are relatively new in medical imaging, include some artificial intelligence-based methods to address some of the uncertainties of medical images.

The definition of radiomics was made by Kumar et al. as: "The extraction and analysis of large amounts of advanced high throughput of imaging features with high throughput from medical images obtained with computed tomography, positron emission tomography or magnetic resonance imaging. Importantly, these data are designed to be extracted from standard-of-care images, leading to a considerable potential subject pool" [42]. It is possible to classify radiomic features according to feature extraction techniques. Tissue and shape features are widely used. For example, intensity levels obtained from an area of tumor tissue in an image can contain valuable information. The gray intensity level relationships among the volumetric elements (voxels) are explained by the textural feature of the radiological image. On the other hand, it is necessary to focus on the shape-based features for morphological analysis [43]. Feature extraction methods commonly used in the analysis of PET images are briefly explained below.

- Statistical distributions of density combinations of certain positions in an image are used to statistically calculate texture features. The number of pixels identified as intensity points defines the statistical level. These are first-order, second-order, or higher-order statistics. A technique for extracting second-order statistical tissue features is the gray-level co-occurrence matrix (GLCM) [44]. Defining the relationship between neighboring pixels, the GLCM indicates the repetition frequency of the image brightness at a given distance and direction. Generally, the distance between pixels is considered to be one pixel when constructing the co-formation matrix. Homogeneity, autocorrelation, contrast correlation, and energy are some of the features extracted by GLCM [45].
- Gray-level run-length matrix (GLRLM) is used to extract the regional features in order to reveal the coarseness of image textures in specific directions for a certain segmented area. The length of consecutive voxels that share the same gray level intensity in a specific linear direction is defined as a run. Fine textures have shorter

runs with similar gray level intensities, while coarse textures have longer runs with different gray level intensities [46]. With the application of this technique, which is the essence of the GLRLM method, the features of these fine and coarse image tissues are extracted.

- Gray-level size zone matrix (GLSZM) deals with the size of each area formed by pixels with the same intensity level in an image. Because a homogeneous texture means large areas of the same density rather than small groups of pixels in a certain direction. The homogeneity of the texture and the width and flatness of the matrix are directly proportional. Also, this matrix does not require calculations in various directions with respect to different areas and is advantageous in this sense. However, the number of gray levels should be specified, which makes robust the calculations against noise [47].
- The textural features calculated by the neighborhood gray tone difference matrix (NGTDM) are related to the differences between a centre voxel and its neighboring voxels within a given window size. It has some parameters such as coarseness, busyness and contrast. Coarseness, one of the most basic texture features, is related to the granularity information of the image. Busyness is related to the rate of change in intensity in an image. Contrast is related to the dynamic range of the intensity levels in the image [48].

The convolutional neural network (CNN) model, which is a deep learning method specialized on images, has a feature extraction part in itself. Figure 2.10 shows a representation of a convolutional network. Designed with the aim of classification, this network has 5 different layers as input, convolution, pooling, fully-connected and output layers. The input layer determines the size for images that are input to the network. Images may need to be resized depending on the situation. Here, three layers representing RGB channels are seen. The image coming into the convolution layer is convolved with multiple kernels using weights. Then, the pooling layer reduces the size of the image while trying to prevent possible information loss. These two layers (convolution and pooling) form the feature extraction part. Just before the output layer there is a fully-connected layer. The weighted features are combined in this layer. The fully-connected layer forms the classification part of the network. Finally, in the output layer there is one output neuron for each class [49]. In this way, the features of an image given as input are extracted and the classification process is performed.

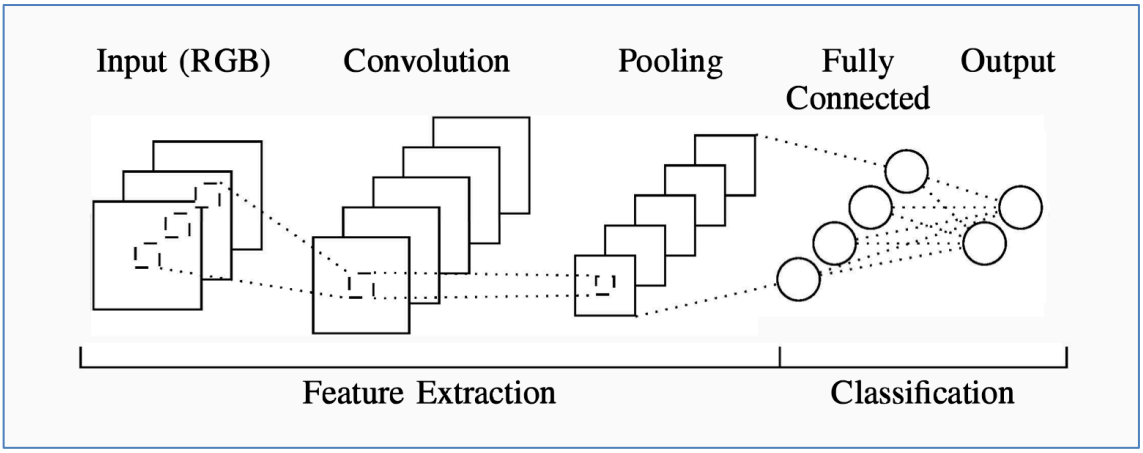

Figure 2.26 Schematic diagram of a convolutional neural network [49]

As a mathematical summary, convolution operation refers to how one of two functions changes the other. The mathematical formula of convolution operation is as follows.

$$(\boldsymbol{f} * \boldsymbol{g})(\boldsymbol{t}) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} \boldsymbol{f}(\boldsymbol{\tau})\boldsymbol{g}(\boldsymbol{t}-\boldsymbol{\tau})\, \boldsymbol{d\tau} \tag{2.1}$$

Figure 2.11 shows the input image and feature detector matrix [50]. Generally, 3x3 matrix is used as feature detector, but sometimes 5x5 or 7x7 matrices are also used. Feature detector is often referred to as kernel or filter.

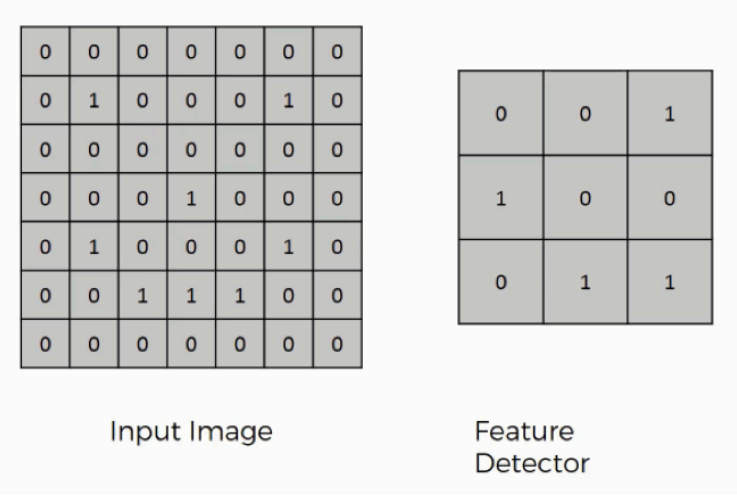

Figure 2.27 Representation of input image and feature detector in the convolutional operation [50]

The 3x3 feature detector is placed in the upper left corner as the starting point over the input image, and the number of cells that feature detector matches the input image is counted. The matching cell number is then put to the upper left cell of the feature map. Figure 2.12 shows this process [50].

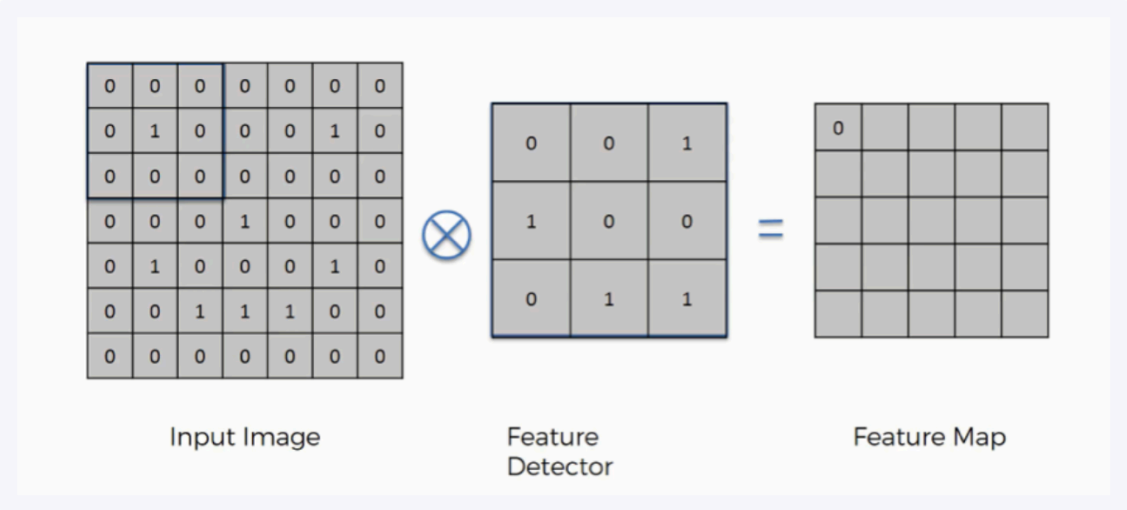

Figure 2.28 First step of the convolution operation to generate a feature map [50]

As a second step, the same operation is done by shifting feature detector one cell to the right. These steps are performed for all rows. The feature map created as a result of all operations is seen in Figure 2.13 [50].

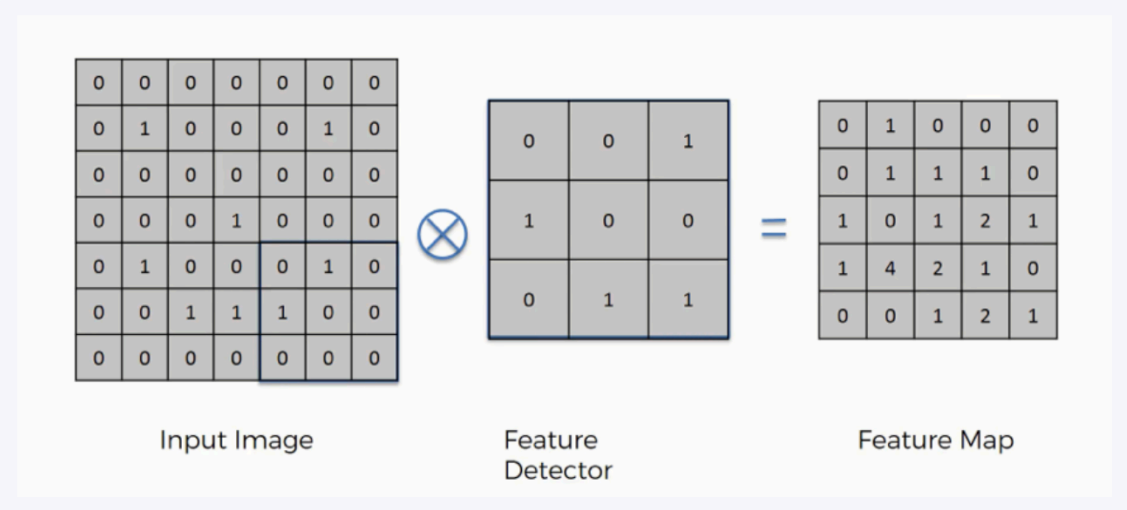

Figure 2.29 Generated feature map as a result of the convolution operations [50]

Operations are not that simple in CNN models. More than one feature maps are generated using multiple feature detectors. Then, these feature maps form the convolutional layers as shown in Figure 2.14 [50].

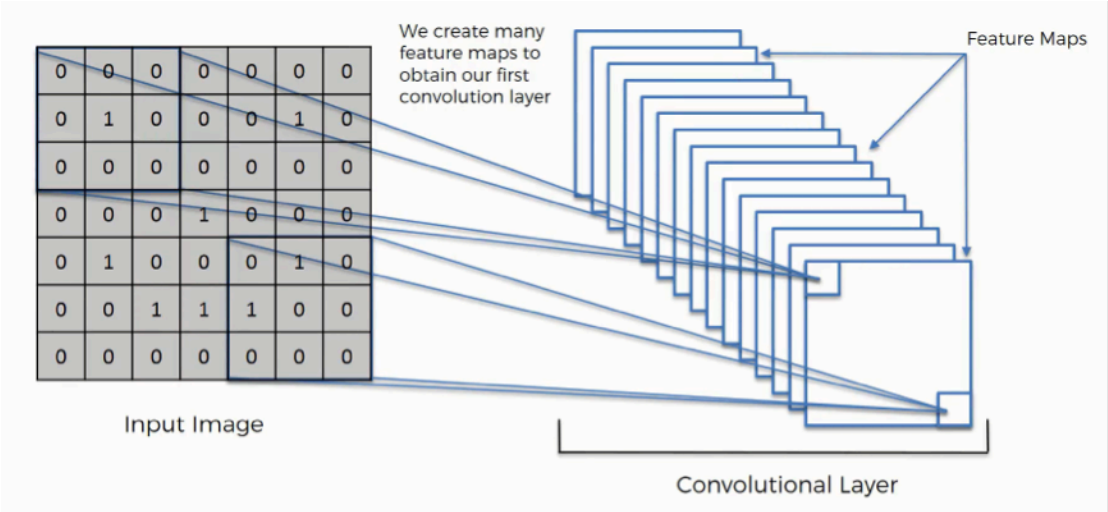

Figure 2.30 Obtaining the convolutional layer [50]

The rectified linear unit (ReLU) is a supplementary step (Figure 2.15) to the convolution operation. This process is applied to increase the non-linearity in the image. Because images are not actually linear anyway. The image can evolve towards linearity with the convolution operation and for this reason this is tried to compensate by using ReLU [51].

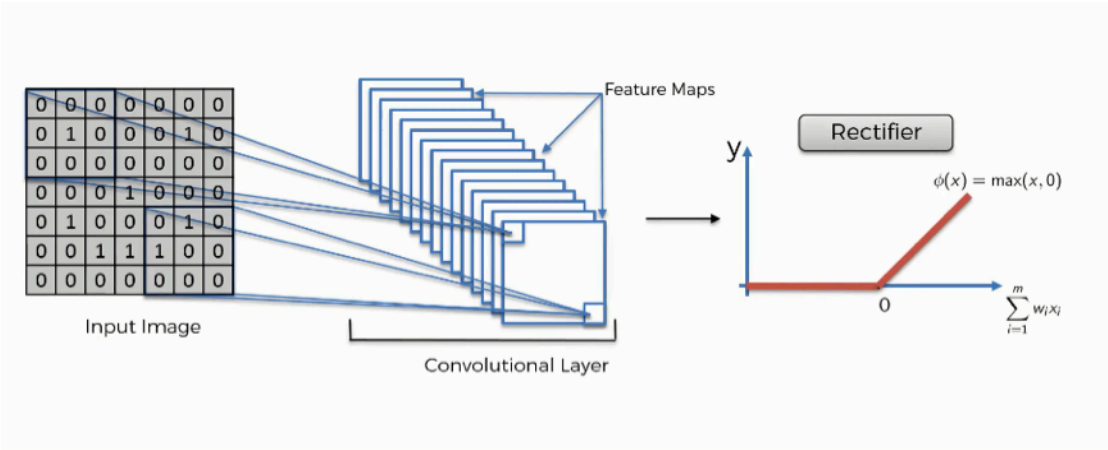

Figure 2.31 The rectified linear unit (ReLU) [51]

In max pooling, a 2x2 box is placed in the upper left corner. The largest of the values in the 4 cells on which this box stands is taken and placed in the pooled feature map. Figure

2.16 shows a representative example of this process. Here, as common, 2-pixel steps are used when shifting the box. The resulting 3x3 pooled feature map is seen. Even if a part of the box is empty while shifting the box, the operations are continued in the same way. The purpose of this process is to get rid of unnecessary information and some possible distortions. In addition, this process also serves to prevent overfitting that may occur in complex networks as it will minimize the image size and number of parameters [52].

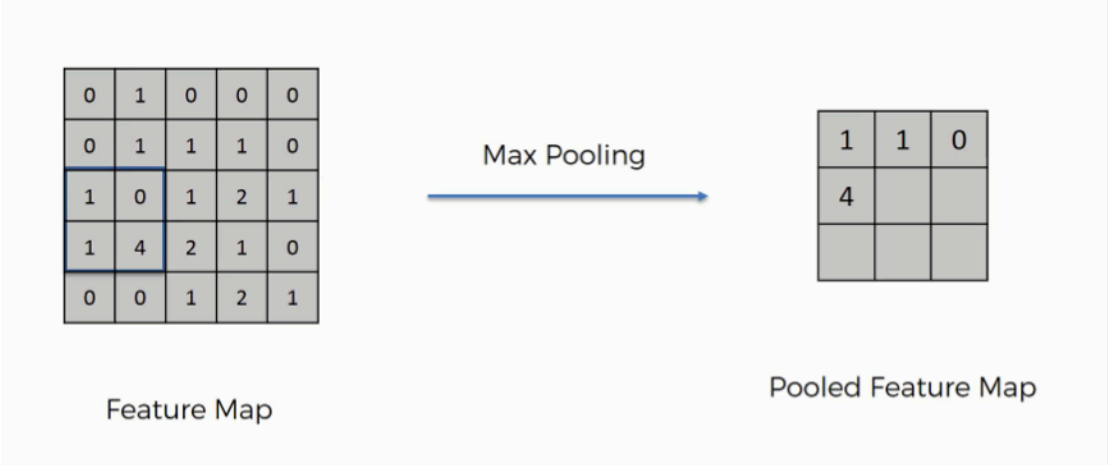

Figure 2.32 Max pooling operation [52]

The pooled feature map obtained after max pooling is subjected to flattening operation. With the application of this operation, the pooled feature map will turn into a column as shown in Figure 2.17. The purpose of the flattening operation is to prepare the features for the fully-connected layer that will come in the next step, namely the artificial neural network [53]. The feature vector obtained after flattening is passed through the artificial neural network. As a result, an estimate is obtained. If the network's task is to classify cats and dogs, for example, then a cat or dog prediction will come as output. If an incorrect estimation is made, the error is calculated using the loss function. As a result, information about how well the network works is obtained. Then, in line with this information, weights are updated again to increase prediction success, and in this way, the network continues to be optimized until the loss value is minimized. Figure 2.18 shows a fully-connected layer between input and output layers. However, it is not uncommon to use more than one fully-connected layer at the end of CNN models [54].

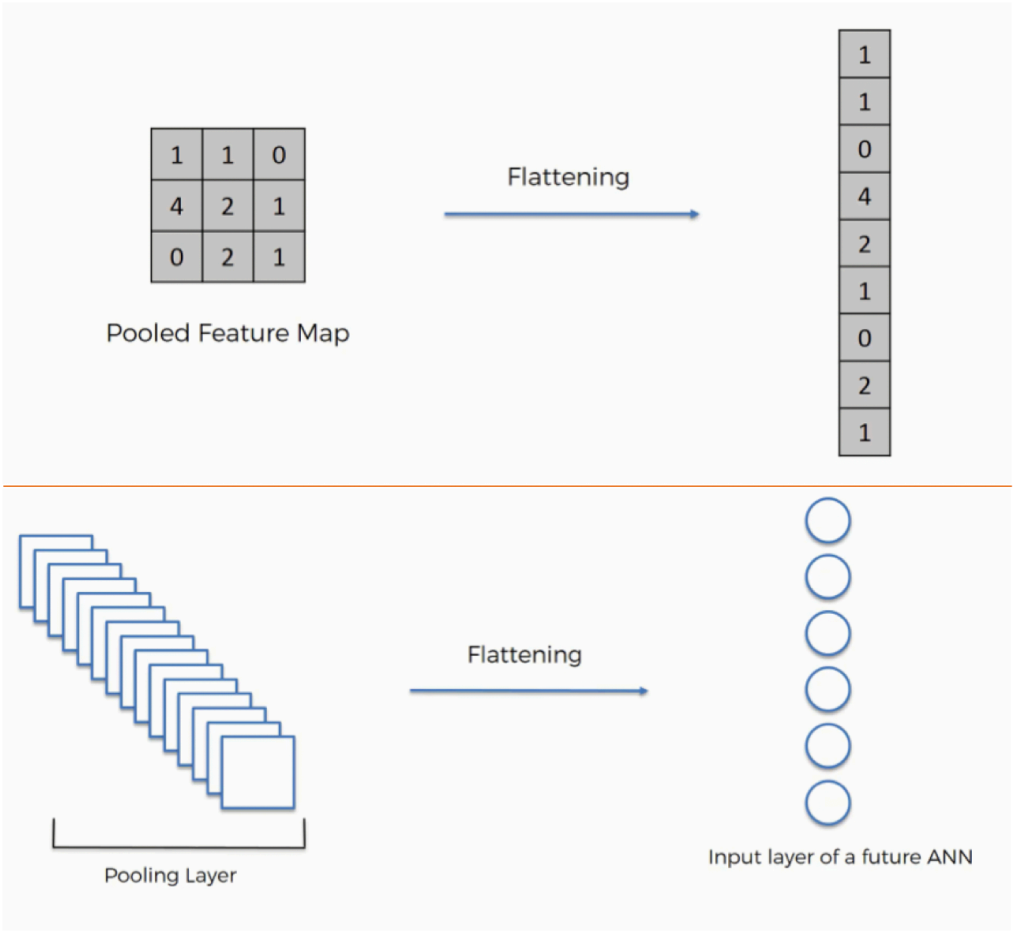

Figure 2.33 Flattening operation [53]

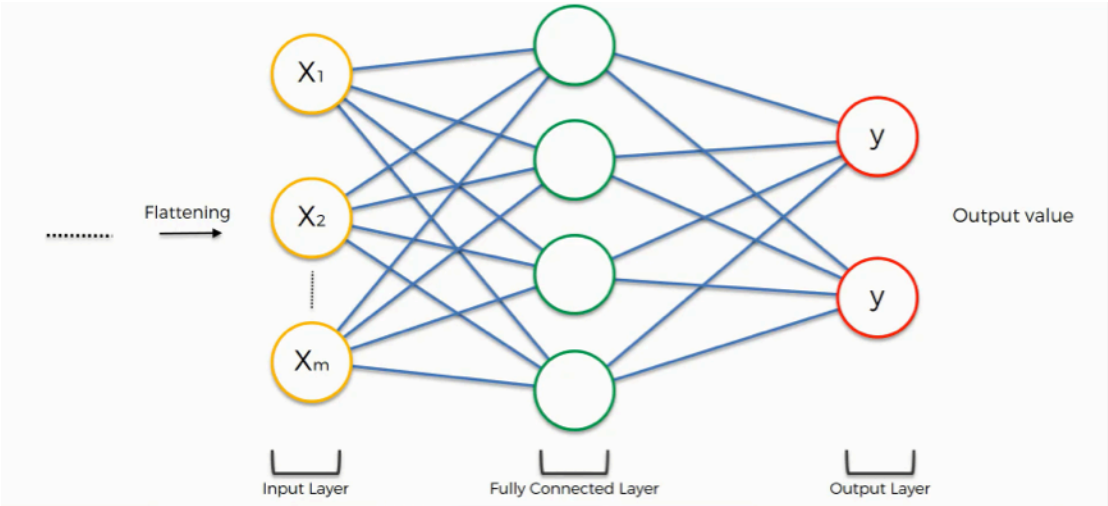

Figure 2.34 A representation of the fully-connected layer [54]

Figure 2.19 represents a summary of all essential operations (input, convolution, ReLU, pooling, flattening, fully connection and prediction) in a CNN model as a whole picture [55].

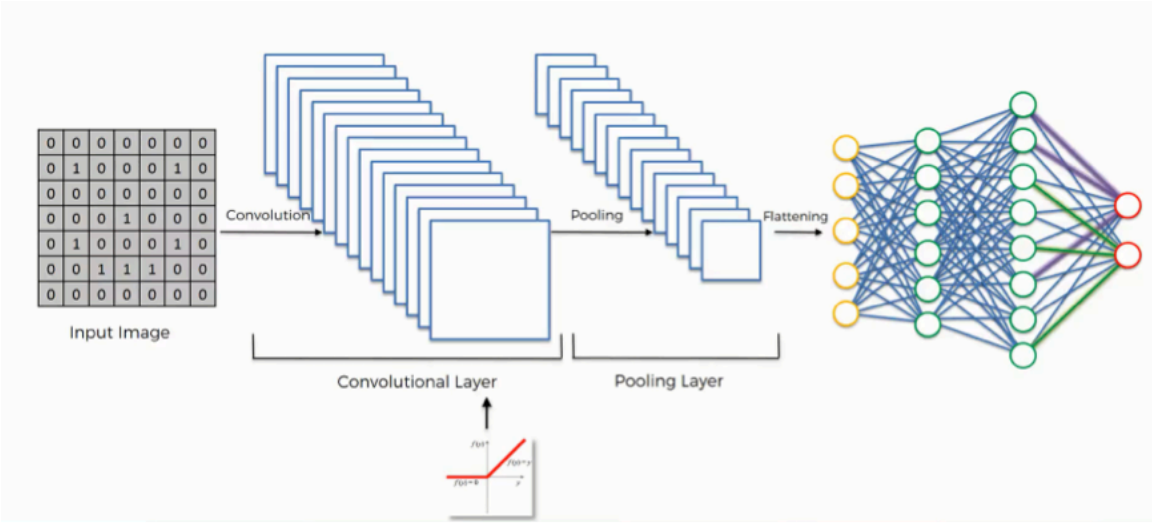

Figure 2.35 Summary of operations in a CNN model [55]

2.4.2 Image Classification

Image classification is the process of predicting which class the relevant image belongs to by passing the features taken from the images through certain processes and this is a supervised learning problem. Based on this, it can be said that the obtained feature values have a direct effect on the classification success. And, if there is a segmentation stage before feature extraction, the success of segmentation is directly proportional to the success of feature extraction and hence classification. Especially when it comes to medical images, this becomes more important. Because medical images are generally scarce and also contain more valuable information. The classification process includes a number of processes such as data splitting, training, testing and optimization. First of all, the data is split into two parts, for example 80% training and 20% test sets. Here, while the network is trained with the training set, it is tested how well the network is trained with the test set. Also, k-fold cross-validation method can be applied. For example, in a 10-fold cross-validation application, all data is split into 10 equal parts and one of these parts is used as a validation set for each training stage, while the remaining data is used as a training set. In addition, the stratify method, which ensures balanced splitting of the data, is important and widely used. After data splitting, training process is started. The

classification error (loss) is calculated by testing after the network trained in the training process. The loss value reveals how well the network is performing. Weights on the network are then updated and the training process is repeated in order to optimize the network for more successful results. One forward pass and one backward pass of all the training samples through the network during the training process is called one epoch. Generally, these samples are given in batches (partially). Therefore, dividing the total number of samples by the number of samples in each batch (batch size) gives the total number of iterations in an epoch. Epoch number is indicated before training begins. As the number of epochs increases, success may not always increase, and too much training may cause overfitting. In other words, after a certain number of epochs, the success rate may remain stable or decay. One of the ideal ways to solve this problem is to apply early stopping. The early stopping method stops the training process if the success does not increase or the loss value does not decrease after a certain epoch and guarantees the best success or loss values achieved in the training process.

2.4.3 Deep Learning Models

Deep learning is a popular machine learning approach. Deep learning networks have successive layers that use the output from the previous layer as input. While the early layers carry more basic features, later layers carry more complex features. For example, information, such as edges and corners, is carried in the early layers, while the later layers carry structures such as parts and organs formed by these edges and corners. The difference and advantage of deep learning from classical machine learning methods is that it does not require handcrafted feature extractions from the image. In classification problems, images are given to the deep learning model and the classification process is started by transferring the required features to the fully-connected layers at the end of the network by applying a series of processes. Fully-connected layers, which are in essence an artificial neural network, perform the classification process. In other words, if we take a CNN model, fully-connected layers are responsible for the classification process, while the previous parts that have operations such as convolution and pooling are responsible for feature extraction and transfer. Since deep learning models are generally many layered and complex networks, they require powerful GPUs for image-based studies.

Multi-layer perceptron (MLP) models, which are made deeper by adding many hidden layers, have a structure consisting of fully-connected layers without operations

such as convolution and pooling. Figure 2.20 shows a schematic representation of the MLP network [56]. In this figure, there is one input, one output and 2 hidden layers. However, it is possible to form deeper MLP models by using more hidden layers.

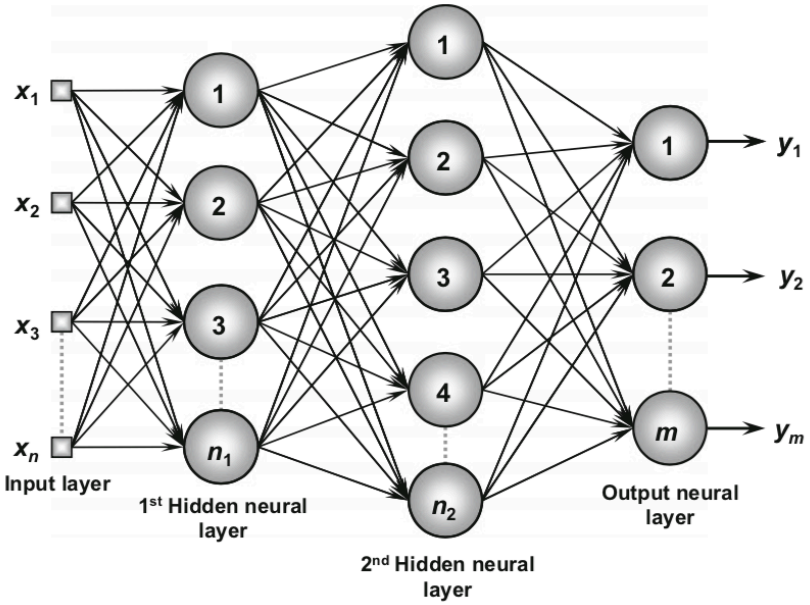

Figure 2.36 A representation of the MLP architecture [56]

Almost all modern deep learning methods include convolutional and pooling layers. In addition, models are being introduced day by day, which are more complex, contain more parameters and therefore require more calculations. LeNet, the first network has convolution-based operations was used by Yann LeCunn in 1998 to recognize handwritten numbers (MNIST dataset). Figure 2.21 shows LeNet architecture [57]. This model has 2 convolutional and 2 fully connected layers.

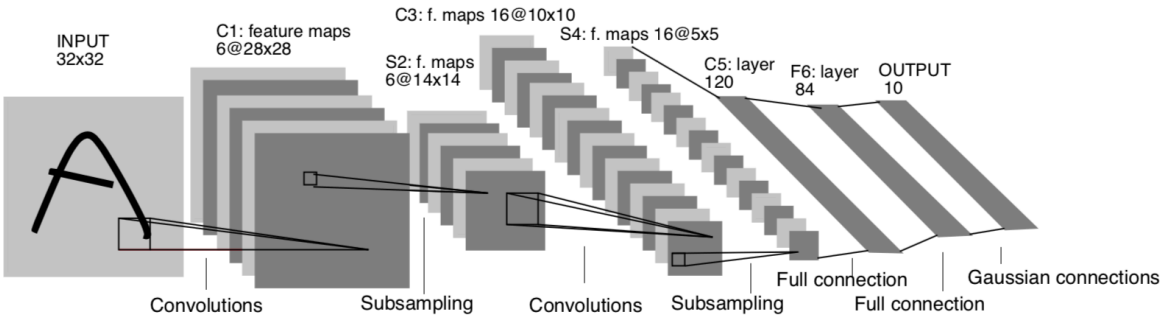

Figure 2.37 LeNet architecture [57]

With the increase in the number of images and the development of powerful GPUs, networks that require more complex and high computing power have become experimentable (e.g., AlexNet [58]). Competitions on ImageNet data helped to reveal deeper CNN models. There are different versions of the VGG model, which performed very well in the field of classification and localization in the ImageNet competition held in 2014. The VGG16 model, which has 13 convolutional and 3 fully-connected layers in its architecture, has 138 million parameters, while VGG19 has 144 million parameters with 16 convolutional layers and 3 fully-connected layers [59]. The architectures of the VGG16 and VGG19 models are shown in Figure 2.23.

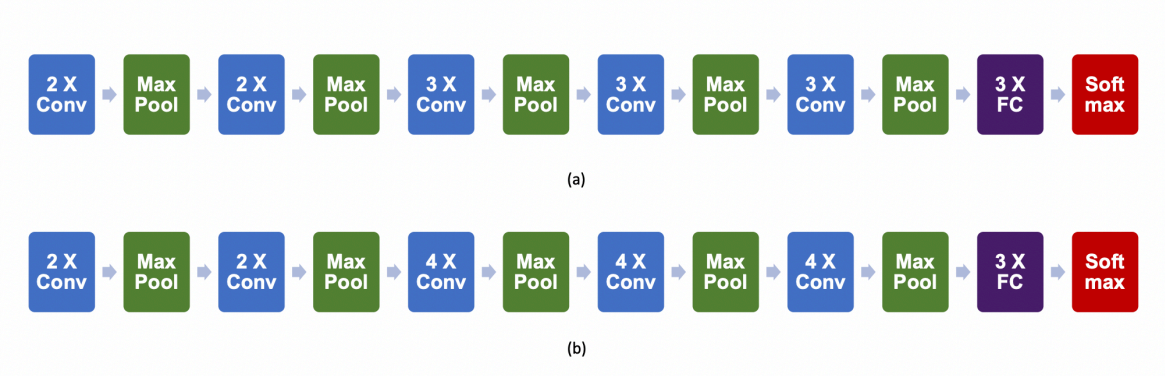

Figure 2.38 Schematic diagram of (a) VGG16 and (b) VGG19 models

While deeper networks continued to be developed, the search for success with simpler networks also continued. One of these studies, SqueezeNet, achieved the success of AlexNet with 50 times less parameters. Some structural improvements underlie its success. SqueezeNet is used to reduce the model size, and the number of parameters while maintaining competitive accuracy. To achieve this, there are three main strategies used in SqueezeNet architecture [60].

- Strategy-1: Smaller network by replacing 3×3 filters with 1 × 1 filters.
- Strategy-2: Limited number of input channels (3 × 3 filters).
- Strategy-3: Delayed downsampling for higher classification accuracy of large activation maps.

The squeeze and expand structures inside the fire modules developed to realize these strategies are shown in the figure. These structures contain 1x1 and 3x3 convolutional filters. Figures 2.24 and 2.25 show SqueezeNet architecture. This model, which contains eight fire layers between two convolutional layers, ends with a classifier, softmax.

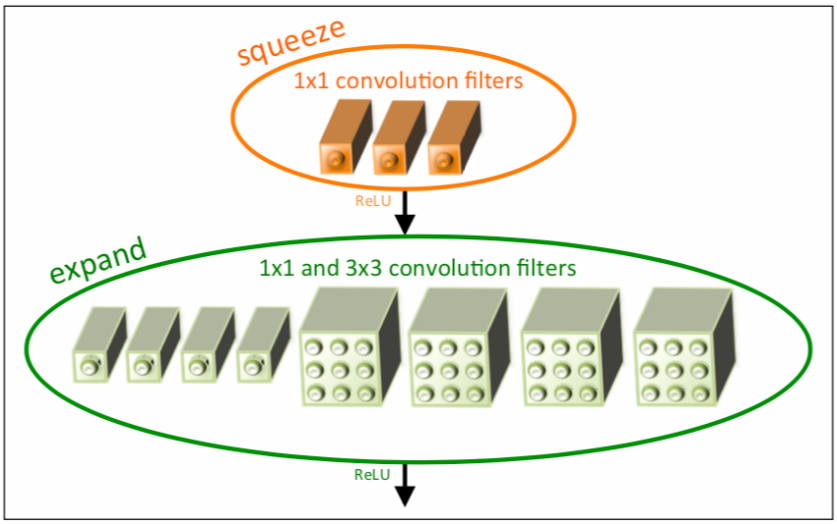

Figure 2.23 Organization of convolution filters in the fire module of SqueezeNet [60]

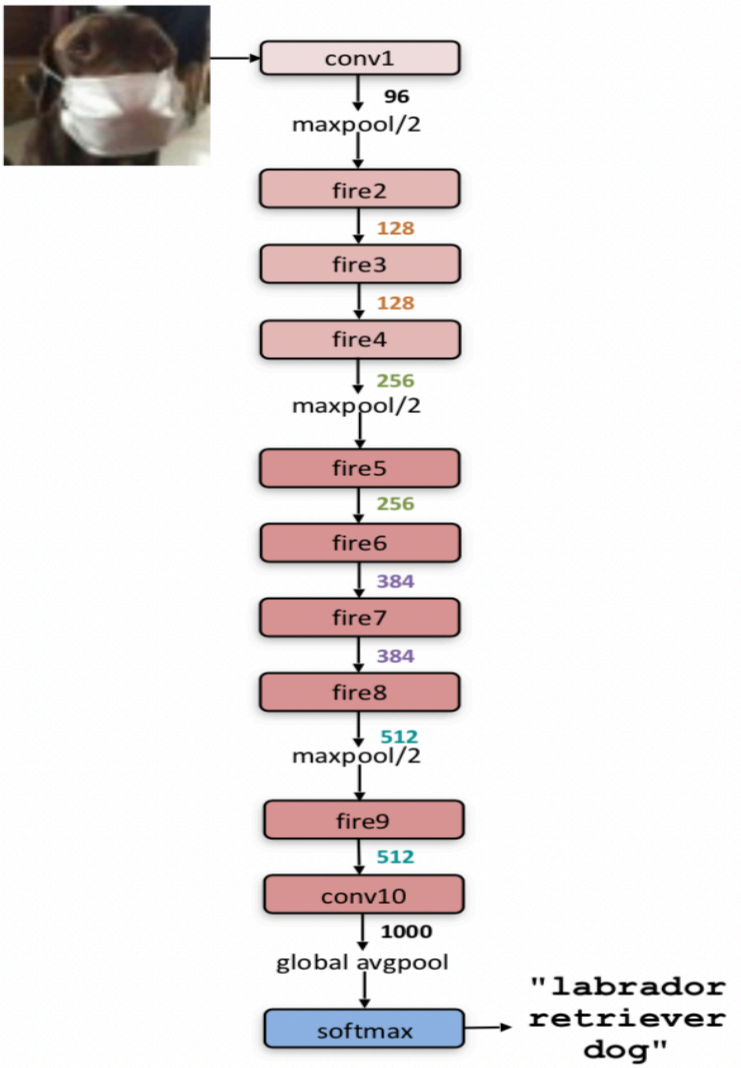

Figure 2.24 View of SqueezeNet architecture [60]

2.4.4 Hyperparameter Optimization

Hyperparameter is a parameter used by the designer of the model to control the learning process and can be optimized with different values. Although very successful results obtained especially in image-based studies, there are still many difficulties in achieving success. Computational costs are also increasing as more and more deep models are developed to achieve more successful results, and the image quality increases. For this reason, faster GPU requirements are constantly emerging. It is not enough to use only standard models to achieve successful results. It is possible to increase the standard performance of the model to higher levels if the correct hyperparameters and the correct values corresponding to them are used. In addition, overfitting is very likely in very deep and complex networks. With some hyperparameters such as optimizers and regularizers, it is possible to increase the success of the model by avoiding overfitting problems.

Updating of parameters in deep learning is done by backpropagation process. In the backpropagation process, firstly, the difference is found by taking a backward derivative called “chain rule”. Then, the difference value found is multiplied with the learning rate parameter. Then the result is subtracted from the weight values and the new weight value is calculated. The learning rate parameter used during the training process can be determined in four ways:

1. It can be determined as a fixed value.
2. It can be determined as a step incrementing value.
3. It can be determined depending on the momentum value.
4. It can be learned during learning by adaptive algorithms.

In deep learning applications, training by processing all the data in the data set at the same time is a costly task in terms of time and memory. Because, in every iteration of learning, gradient descent calculation is performed backward on the network with the backpropagation process and the weight values are updated in this way. To solve this problem; the data set is divided into small groups and the learning process is carried out on these small groups. Processing more than one input in parts in this way is called mini-batch. The value determined as mini-batch parameter while designing the model; it means how many data the model will process at the same time. The specified batch value must fit into the GPU memory. For this reason, batch size should be determined in multiples

of 2. For example, 2, 4, 8, 16, 32,…512 etc. If it is not specified in this way, there may be sudden drops in performance.

The most important feature that distinguishes the deep learning method from other artificial neural networks, especially in complex problems, is the number of layers. The concept of depth comes from here. There is no common method for deciding the number of layers. However, one can start with a single-layer or two-layer network and concentrate on finding suitable hyperparameters. The number of layers can then be increased gradually. Increasing the number of layers does not have much effect after a certain point. After this point, the performance of the model should be tried to be increased with other hyperparameters. However, layer limit should be investigated in order to observe the limits of the model. The number of neurons indicates the number of information kept in memory. The high number of neurons increases the memory need and calculation time. In the case of a non-GPU environment, these criteria should be considered. However, low number of neurons can causes underfitting. Another feature is the different distribution of the number of neurons between layers. In other words, while using more neurons in the first layers, continuing it by decreasing it in the advancing layers produces a regularization effect. This is the reason why the dropout method works [61].

It is known that the learning process in deep learning applications is basically an optimization problem. Optimization methods are used to find the optimum value in the solution of nonlinear problems. Deep learning applications commonly use optimization algorithms such as stochastic gradient descent (SGD), RMSprop, Adadelta and Adam. These algorithms differ in performance and speed.

- Gradient Descent with Momentum: It is aimed to reach the best result in the fastest way in gradient descent. It is inconvenient that the steps taken to reach the global minima are too small or too large. Because if very small steps are taken, it may be possible to get stuck in local minima. Very big steps, on the other hand, become harder to reach the target due to excessive deviating attempts. In both cases, the global minima is reached too late or not at all. Gradient descent with momentum has been proposed to eliminate this problem [62]. With this method, dynamic averages of past learning steps guide future steps. This way, the target is approached faster with fewer deviations.
- RMSprop (Root Mean Square Propagation): The RMSprop method was proposed by Geoffrey Hinton [63]. This method implements learning steps using a different learning rate for each parameter. These learning rates can be updated according

to the results obtained. Since this method does not consider distant points, it can accelerate the convergence. In the Adadelta optimization method [64], the parameters have a gradually decreasing learning speed and the learning process cannot take place after a certain point. RMSprop has been proposed to solve this problem.

- Adam (Adaptive Moment Estimation): Adam, an adaptive optimization algorithm, works more efficiently than stochastic gradient descent method [65]. It keeps the momentum changes as well as dynamically updating the learning rate for each parameter. That is, Adam combines the momentum and RMSprop methods.

The most commonly used regularization methods to prevent overfitting in deep learning methods are lasso [66] and ridge regression [67]. Lasso and ridge are also known as L1 and L2, respectively. To avoid overfitting, L1 deals with the estimation of the median of the data while L2 deals with the estimation of the mean of the data. The sum of the absolute values of the weights in the L1 method and the sum of the squares of the weights in the L2 method are calculated and added to the total error function after multiplying the values obtained by a regularization coefficient. If the regularization coefficient is determined too small here, it may cause overfitting. If this coefficient is chosen too large, it will cause underfitting. The aim of dropout method, which is one of the most frequently used techniques, is to drop some neurons randomly from the hidden layer according to a determined ratio [68]. These neurons, which are dropped to prevent overfitting, are not included in the process during backpropagation also.

Chapter 3

Study-1

3.1 Classification of Non-Small Cell Lung Cancer Subtypes with Artificial Neural Networks

Artificial Neural Networks (ANN) based approaches are artificial intelligence methods inspired by the ability of the human brain to process data with the neurons. Neurons in the system process the incoming inputs with some techniques and convert them into decision outputs. In the literature, ANN have been used to solve different problems in the field of medicine as in many other fields. Used in computer aided diagnosis, ANN have an important place in terms of detecting and recognizing certain structures (region of interest) in medical images [69-78]. The variety of problems and imaging methods used in these studies show that the ANN method can be applied to a wide range of areas. Edge detection and segmentation studies are common in analyzing medical images, especially in region-based investigations. Within the scope of these studies, the ANN method is frequently used [79-85]. In addition, cancer research is among the medical studies using ANN. Studies such as classification of brain cancer [86], early detection of prostate cancer [87], lung cancer detection [88], [89], survival analysis on breast cancer [90], prediction of breast cancer malignancy [91] can be cited as examples.

Lung cancer is one of the most dangerous diseases. This cancer type has two main varieties. First one is non-small cell lung cancer (NSCLC). And the other one is small cell lung cancer. Adenocarcinoma (ADC) and squamous cell carcinoma (SqCC) are two subtypes of NSCLC. Ayyildiz et al. conducted a study using classical machine learning methods to differentiate these two subtypes [4]. First, they performed tumor segmentation from PET images using the random walk method. Later, they extracted 39 features from the segmented images to be used in classical machine learning methods. In our study, these 39 features were used to produce a solution to the subtype classification problem

with the ANN approach. For this reason, brief information about the feature extraction methods used is given in the coming section.

3.2 Materials and Methods

This study was performed using 18F-FDG PET / CT images of 93 patients with NSCLC. Images were obtained from patients with the help of PET / CT device (Siemens Biograph 6, HiRez). in Nuclear Medicine Department of Acıbadem Kayseri Hospital in Kayseri between March 2010 and April 2014. The study was approved by the Research Ethics Committee of the Kayseri Research and Training Hospital (KRTH) for 93 patients, 5 of whom were female and 88 were male and whose mean age was around 63 (Age interval: 39-84). As tumor subtype, 39 patients from 93 patients were diagnosed with ADC and the rest were diagnosed with SqCC. The specimens obtained with fine needle or excisional biopsy was evaluated in the pathology section of Kayseri Research and Training Hospital in terms of tumor subtype.

Ayyildiz et al. performed the study shown in Figure 3.1 using classical machine learning methods to differentiate two subtypes of NSCLC (ADC and SqCC) [4]. The 39 features obtained within the scope of this study were used as inputs to ANN in our study.

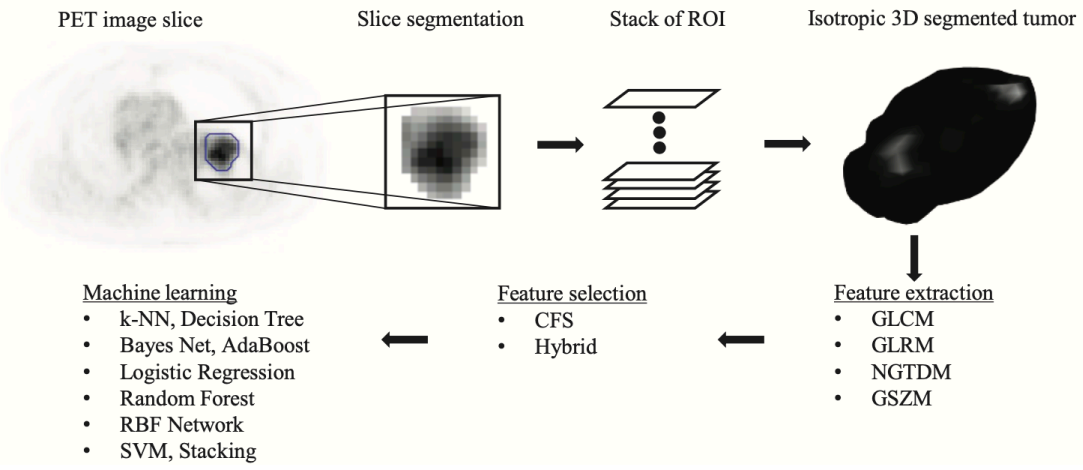

Figure 3.4 Summary of the work from which the features were obtained [4]

From 39 features, 8 using gray level cooccurrence matrix (GLCM) [44], 13 using gray level run length matrix (GLRLM) [46], 13 using gray level size zone matrix (GSZM) [47], and 5 using neighborhood gray tone difference matrix (NGTDM) [48] were

obtained. The details of these approaches can be found elsewhere [92]. In addition, these methods, which are widely used in the literature, have been investigated in recent studies [93] [94]. The features we used in our classification study as the inputs to the ANN to be trained are shown in Table 3.1.

Table 3.2 Feature extraction methods and features extracted using these methods

Feature Extraction Methods	Features
GLCM	Energy
	Contrast
	Entropy
	Homogeneity
	Correlation
	Sum Average
	Variance
	Dissimilarity
GLRLM	Short Run Emphasis (SRE)
	Long Run Emphasis (LRE)
	Gray Level Non-Uniformity (GLN)
	Run Length Non-Uniformity (RLN)
	Run Percentage (RP)
	Low Gray Level Run Emphasis (LGLRE)
	High Gray Level Run Emphasis (HGLRE)
	Short Run Low Gray Level Emphasis (SRLGLE)
	Short Run High Gray Level Emphasis (SRHGLE)
	Long Run Low Gray Level Emphasis (LRLGLE)
	Long Run High Gray Level Emphasis (LRHGLE)
	Gray Level Variance (GLV)
	Run-Length Variance (RLV)
GLSZM	Small Zone Emphasis (SZE)
	Large Zone Emphasis (LZE)
	Gray Level Non-Uniformity (GLN)
	Zone Size Non-Uniformity (ZSN)
	Zone Percentage (ZP)
	Low Gray Level Zone Emphasis (LGZE)
	High Gray Level Zone Emphasis (HGZE)
	Small Zone Low Gray Level Emphasis (SZLGE)
	Small Zone High Gray Level Emphasis (SZHGE)

	Large Zone Low Gray Level Emphasis (LZLGE)
	Large Zone High Gray Level Emphasis (LZHGE)
	Gray Level Variance (GLV)
	Zone Size Variance (ZSV)
NGTDM	Coarseness
	Contrast
	Busyness
	Complexity
	Strength

According to the definition made by Engelbrecht [95], the human brain is a complex system that functions like a parallel computer that has the potential to perform tasks extremely quickly. ANN, an artificial learning approach developed inspired by the learning ability of the human brain, imitate the transmission of information by biological nerve cells with the help of signals. Figure 3.2 shows the human nervous system consisting of a three-stage system [96]. At the center of this system is the brain. In the figure, arrows from left to right indicate transmission of information, while arrows from right to left represent feedback. Receptors convert external impulses into electrical signals and transmit them to the neural net, i.e. the brain. Effectors convert the electrical impulses sent by the neural net into interpreted outputs.

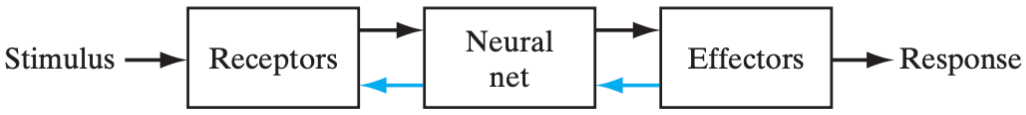

Figure 3.5 Block diagram representation of nervous system [97]

Figure 3.3 shows a biological neuron and an artificial neuron inspired by it. The dendrites in the biological neuron correspond to the input in the artificial neuron, while the axons represent the output. Cell body corresponds to transfer functions. Weights are used to represent the function of synapses. The basic components of the artificial neural network can be explained as follows.

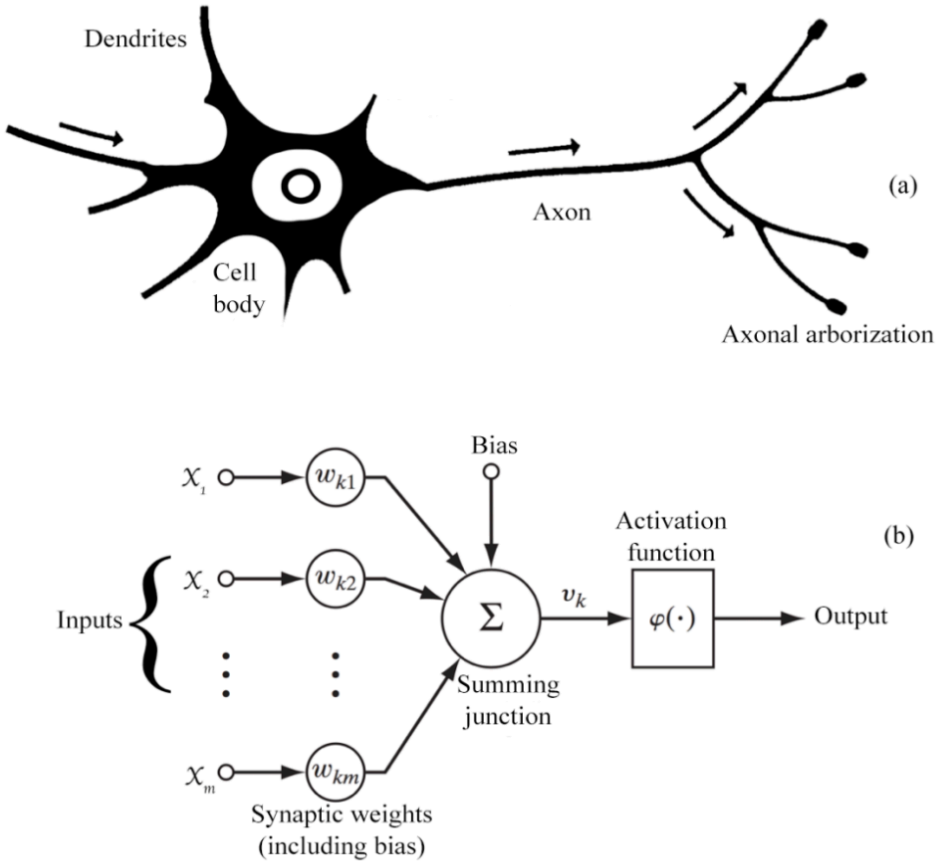

Figure 3.6 (a) Biological neuron [98] and (b) artificial neuron [97]

The task of the input layer is to send the inputs directly to the next layer without any calculation or processing. The weights contain certain coefficients to show how important the inputs are. Therefore, they are very important elements for the success of the network. Hidden layers are layers located between the input and output layers and can be multiple. A small number of hidden layers may cause insufficient learning, and too many of these layers cause the network to fail due to memorization. Therefore, the number of hidden layers is among the hyperparameters that need to be optimized. The values collected by the summation junction are sent to the activation function. The activation function is also known as the transfer function. There is an activation function for each neuron, and these functions make certain calculations to provide neuron outputs. Different types of summation and activation functions can be used. Output layer is the layer in which inputs given to the network from the input layer are transformed into final network outputs. There is only one input and one output layer in a neural network. ANN basically consist

of these components, but network models customized according to the problem can be designed [98].

In this study, fast artificial neural network library (FANN) was used for ANN trainings [99]. Applying multi-layer feed forward networks, FANN is designed to be fast and easy to use. In the use of this library, FANN Tool was used as a GUI [100]. There are some training algorithms in the FANN library. The algorithm to be used for the training process is chosen with the fann_set_training_algorithm function. Here, FANN_TRAIN_RPROP algorithm is used by default. We can briefly explain these algorithms used in our study as follows:

- FANN_TRAIN_INCREMENTAL: It is the standard backpropagation algorithm. Weights are updated after each training. In this method, weights are updated many times during each epoch and it is not suitable enough for complex problems while working fast for some problems.
- FANN_TRAIN_BATCH: In this method using the standard backpropagation algorithm, the weights are updated only once during an epoch according to the mean square error calculation. In this sense, it is different from the incremental training algorithm. Batch training algorithm works slower, but due to more accurate error calculation, it gives more successful results in some problems than incremental training algorithm.
- FANN_TRAIN_RPROP: This method, which is a more advanced version of batch training algorithm, does not use learning_rate and works adaptively. It gives good results for many problems. The RPROP algorithm is described by the work done by Riedmiller and Braun [101]. However, actually used here is the iRPROP algorithm, a type of RPROP algorithm introduced by Igel and Husken [102].
- FANN_TRAIN_QUICKPROP: Quickprop algorithm [103], which emerged as a more advanced version of batch training algorithm, has gained the potential to give better results by including the learning_rate parameter.

Some activation functions were used during the training processes. These activation functions, which are standard in the FANN library and used with the training algorithms mentioned above, are as follows:

- FANN_SIGMOID: Sigmoid, one of the most widely used activation functions, outputs between 0 and 1. It is especially used in binary classification problems.

- FANN_SIGMOID_STEPWISE: This method, which outputs between 0 and 1 and is defined as a stepwise linear approximation to sigmoid, works faster than sigmoid, but is less precise.
- FANN_SIGMOID_SYMMETRIC: This function, known as tanh and one of the most used activation functions, outputs between -1 and 1.
- FANN_SIGMOID_SYMMETRIC_STEPWISE: This method, which outputs between -1 and 1 and is defined as a stepwise linear approximation to symmetric sigmoid, works faster than symmetric sigmoid, but is less precise.

3.3 Results and Discussion

In summary, for subtype classification of NSCLC, 39 features obtained from PET images using feature extraction methods was given to ANN. Training and tests were carried out using mentioned training algorithms and activation functions. The results of the four algorithms of FANN Library are shown in Table 3.2. These are developed and specialized algorithms based on the backpropagation algorithm. According to Table 3.3.1, Incremental, the basest training algorithm, has the lowest success. Batch and RPROP, which are more advanced algorithms, achieved 68% success rate. Quickprop, the most advanced algorithm, achieved the highest success with 73% test accuracy.

Such subtype classification studies are very important in terms of subtype-specific therapies. Therefore, the methods developed to solve the problem of differentiating ADC and SqCC from each other and the optimization studies on these methods are equally important. It can be said that ANN, which is a machine learning method, forms the basis of deep learning methods. Based on this, the results obtained provided motivation for the deep learning studies in the subtype classification of NSCLC problem carried out thereafter.

Table 3.3 Train and test success of training algorithms

Training Algorithm	Train Accuracy (%)	Test Accuracy (%)
Incremental	86	54
Batch	87	68
RPROP	88	68
Quickprop	87	**73**

Chapter 4

Study-2

4.1 Deep Learning Methods in Slice-Based Subtype Classification of Non-small Cell Lung Cancer

Computer-aided technologies have been used in the medical field for many years. With the development of computer-aided imaging technologies as well as machine learning methods, research on decision support mechanisms, automatic detection, segmentation and classification have gained momentum. The major developments in the automatic analysis of medical images have been thanks to the convolutional neural networks (CNN) method as in other image types.

In this study, two-dimensional slices containing tumor tissues segmented using the random walk method from PET images were used. Here, we examined multilayer perceptron (MLP) model and LeNet, SqueezeNet, VGG16 and VGG19 as CNN models to classify two subtypes of NSCLC using these slices. This study is a slice-based study, and while the train and test data sets were split, they were kept on a slice basis, not on a patient basis. We used some optimizers and regularization methods to increase the success of the models. We also considered optimizing certain hyperparameters of these models by performing cross-validation on each training set.

4.2 Materials and Methods

This study was performed using 18F-FDG PET / CT images of 94 patients with NSCLC. The images were obtained in the Nuclear Medicine Department of Acıbadem Kayseri Hospital in Kayseri between March 2010 and April 2014. The PET / CT imaging device was Siemens Biograph 6, HiRez. 10 to 15 mCi of 18F-FDG was injected to the patients and PET / CT acquisitions were performed approximately 60 minutes after the injection. At 8 or 9 bed positions PET scanning was performed for 2 to 3 minutes at each

position. Three-dimensional iterative reconstruction algorithm was used for the reconstruction of the images. Two nuclear medicine experts evaluated the images and 3D whole body projection using the e-Soft software platform (Siemens, USA). This study was approved by the Research Ethics Committee of the Kayseri Research and Training Hospital (KRTH) with protocol number 20.02.2013/55. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. Five of the patients were female, and 88 were male (mean age was around 63, age interval was 39-84). Until now, no studies were published regarding the tumor variability among male or female NSCLC patients. Thus, we have designed this study without considering the effect of gender on cancer subtype characteristics. Thirty-eight patients were diagnosed with ADC, and the rest were diagnosed with SqCC. For the diagnosis, the specimens were obtained with fine needle or excisional biopsy, and the subtype evaluations were performed in the pathology department of KRTH.

Even though we obtained PET and CT images, we mainly focused on the PET images, especially the slices with tumors. Here, tumor tissues segmented from PET images by random walk method were used. As a result, we had a total of 1457 segmented PET slices that consisted of 516 ADC and 941 SqCC subtypes. The images and their labels were stored in order to get prepared for the next stage. While Figure 4.1 shows four segmented slices with a tumor obtained from one patient, Figure 4.2 and 4.3 depict segmented slices obtained from different patients with ADC and SqCC subtypes, respectively. In order to provide the sub-images containing the tumor with a uniform size to the deep learning models, we had to specify a common and appropriate size for them. The maximum row and column values of the segmented images were taken as the limits of the bounding box with the tumor in 3D. As a result, the bounding box size was specified as 64-by-64 for the training and test phases of the classification process. In the bounding box, all the pixel intensity values were assigned as zero, except for the segmented region that included the tumor.

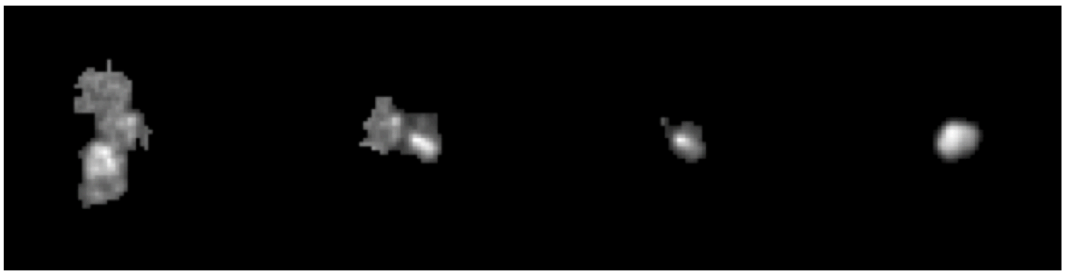

Figure 4.4 Segmented slices with a tumor obtained from one patient

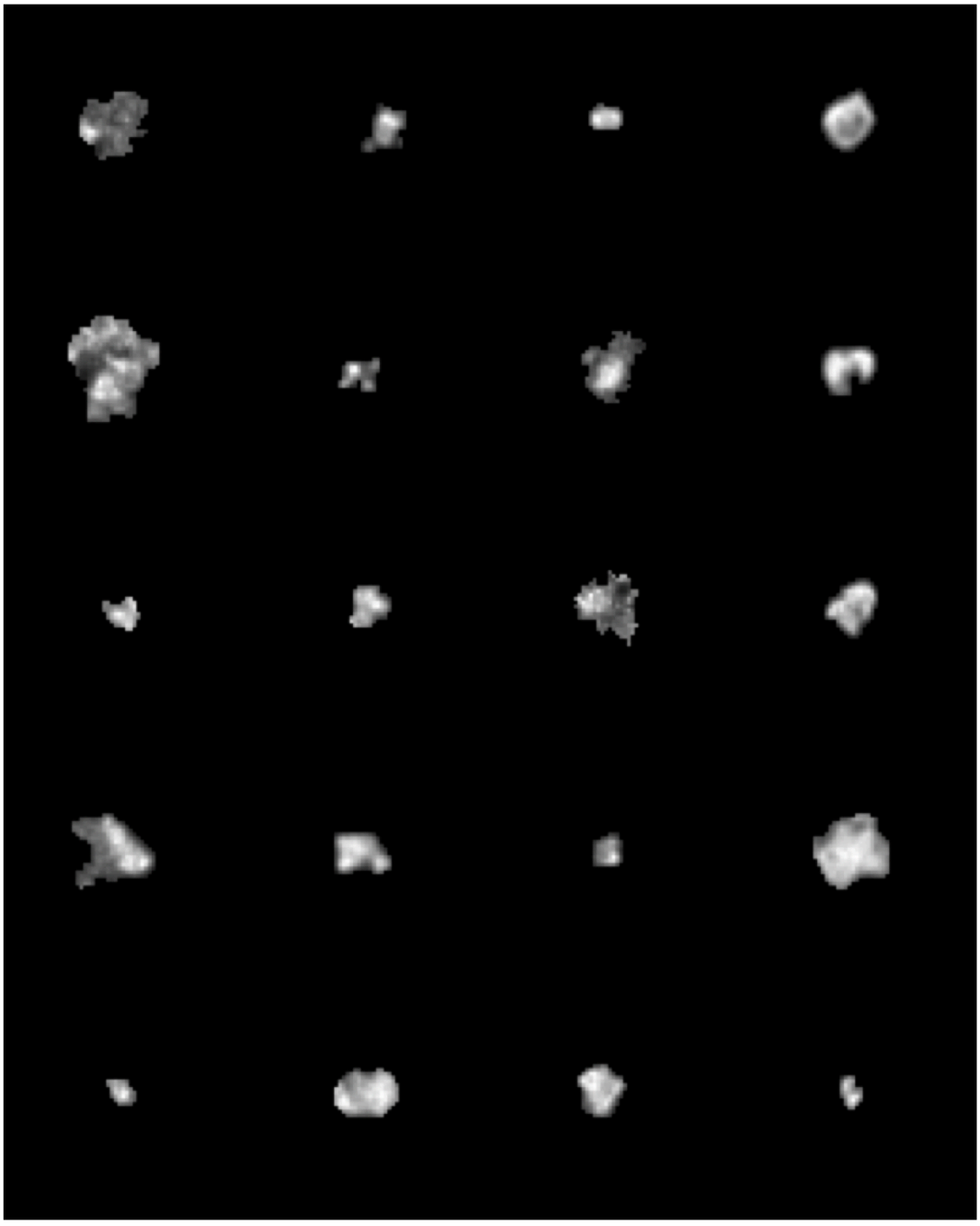

Figure 4.5 Segmented slices obtained from different patients with ADC

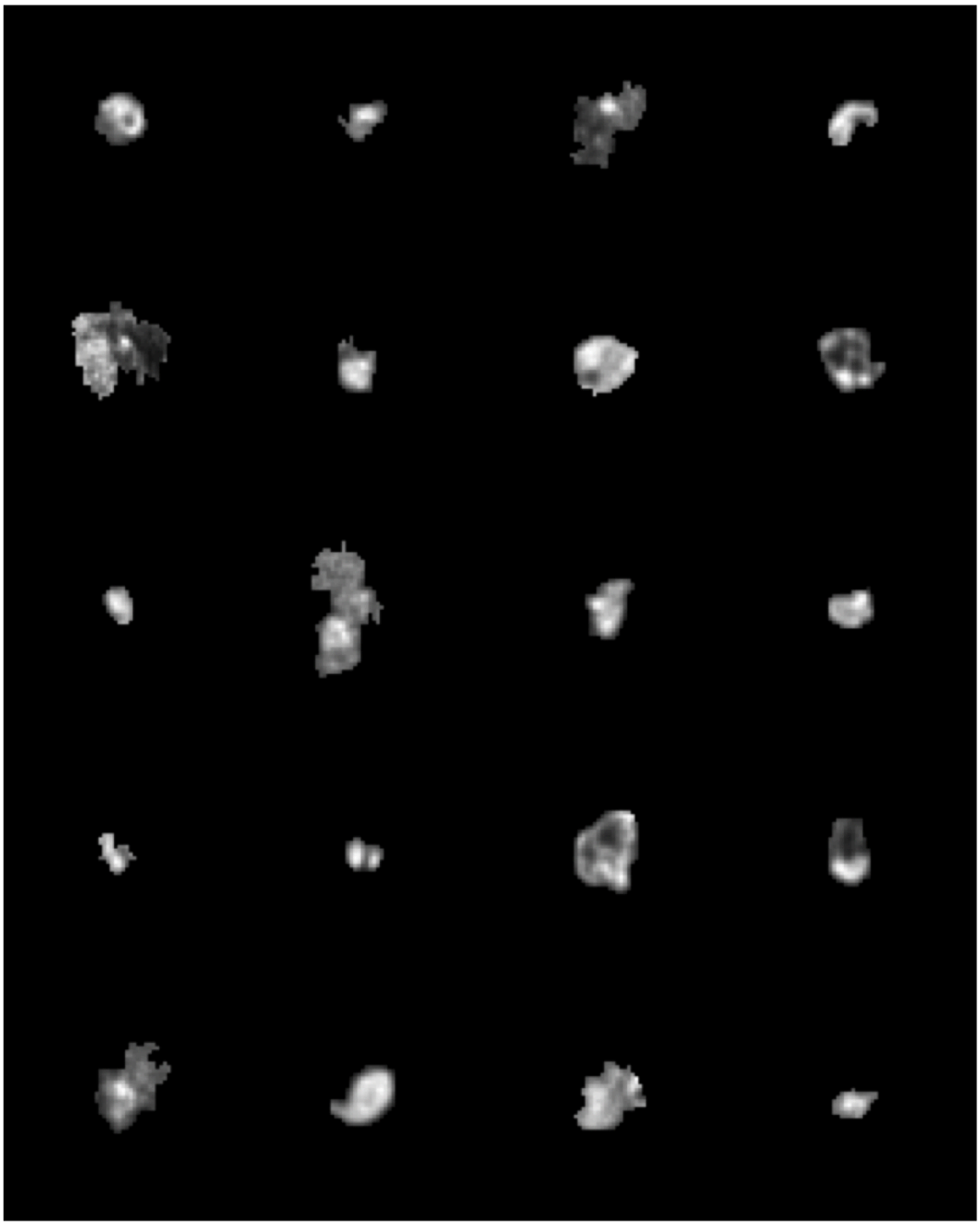

Figure 4.6 Segmented slices obtained from different patients with SqCC

Before training process, the data were divided into two groups: train dataset and test dataset. For this process, 20% of the data was randomly selected and collected as the test data. As a result, 291 slices were determined as the test data and 1166 slices were set as the train data. For the training process, 20% of train data, 234 images were randomly selected and collected for validation. As a result, 932 train, 234 validation and 291 test

data were generated for training and testing processes in CNN and MLP implementations. The models have been trained and tested using computers with Nvidia Tesla K80 GPU.

In this study, in addition to using LeNet, SqueezeNet, VGG16 and VGG19 models as CNN methods, MLP model was also used as a deep ANN approach. During the training of all these models, hyperparameter optimization was made. Table 4.1 shows the hyperparameters used and their corresponding values.

Table 4.2.1 Hyperparameters

Hyperparameters	Values
Optimizers	Adam / RMSprop / Adadelta / Momentum
Regularizations	L1 / L2 / Elastic / None
Drop Out	0.5 / None
Batch Size	16 / 32
Learning Rate	1 / 0.1 / 0.01 / 0.001 / 0.0001
Hidden Layer (for only MLP)	3 / 7 / 11 / 15
Activation Function	ReLu / Tanh (for only MLP)
Augmentation	Rotation range: 90, Horizontal Flip, Vertical Flip / None

As can be seen in Table 4.1, there is "None" value for regularization and dropout. It means that no regularization or dropout has been used on the model. Dropout was used for fully connected layers at the end of the models and the dropout value was set to 0.5. For MLP, 3, 7, 11 and 15 hidden layer models were examined. In addition, the success of the models was evaluated by using ReLu and Tanh functions to be used as the activation functions of the hidden layers. For augmentation, the rotation range value was 90 degrees. In addition, horizontal flip and vertical flip were used. As a result, 5592 train data were obtained after augmentation. In other words, the number of train data was increased by 6 times. However, the validation and test data were not augmented. Here, the value "None" means that no augmentation is performed and only the original data were used.

4.3 Results and Discussion

In this study, some CNN and MLP models were examined to classify two subtypes (ADC and SqCC) of NSCLC. LeNet, SqueezeNet, VGG16, VGG19 models were used as CNN models. In addition, some optimization and regularization techniques were applied to prevent overfitting and increase the success and hyperparameter optimization was

done. Table 4.2 shows the results. According to this ranking, the most successful models appeared to be VGG19 and VGG16 with 95% success with test loss around 0.17. At the same time, VGG models had the highest run time.

Table 4.3 Classification success of the models

Model	Run Time	Val. Loss	Val. Acc.	Test Loss	Test Acc.	Recall	Precision	F-score (%)
VGG19	316	0.197	0.944	0.169	0.955	0.948	0.954	**95.1**
VGG16	244	0.222	0.949	0.178	0.955	0.951	0.952	**95.1**
LeNet	38	0.181	0.936	0.233	0.935	0.920	0.938	92.8
MLP	27	0.235	0.927	0.224	0.931	0.930	0.923	92.6
SqueezeNet	55	0.355	0.842	0.262	0.890	0.889	0.878	88.3

LeNet and MLP models yielded almost the same performance levels and they were the second most successful models with 92% success. When run times were compared, it can be said that LeNet worked about 8 times faster than VGG19, and MLP worked almost 1.5 times faster than LeNet. The inference that can be made from these is that in order to increase the success by 1% or 2%, it may be necessary to use a model several times slower.

The importance of this study is to perform subtype classification from PET images with NSCLC without the need for a pathological procedure. In doing so, a popular method, deep learning has been used. Moreover, no deep learning-based study on subtype classification of NSCLC was found from PET images. In this sense, this study is quite original in terms of its place in the literature. All these efforts are expected to promote the identification of cancer subtypes, which are very important for future targeted therapies. Although this study has yielded highly successful results, in order to be able to progress more appropriately to clinical studies, the data in future studies were split on patient basis and train and test stages were carried out.

Chapter 5

Study-3

5.1 Studies on 3D CNN Models using 3D Patient Data

Many architectures based on convolutional neural networks (CNN) method, which is an image-based deep learning method, have been derived due to the efforts to increase success or adapt to various types of images. And in this area new developments are being made every day. Dimension-based studies are also included in these development activities. 3D CNN architectures are used for 3D image studies. These architectures have 3D convolution and 3D pooling operations. Although these types of studies are generally used in cases where volumetric context is important, they can also be used to give the interrelated data to the network as a stack. For example, in the problem of action recognition in a video, there is a series of 2D images that are interrelated and sequential, and there is a need for 3D conversion to analyze them together. Similarly, 3D structures gain importance in analyzing patient images obtained by medical imaging methods such as PET, CT and MRI. Here, for example, this method can be used for volumetric analysis of a tumor or for different patient-based evaluations.

In the literature, there are many studies in different fields made with 3D CNN models. In addition to natural image recognition problems such as object [104] [105], human action [106], hand gesture [107] and facial expression [108], there are also many medical studies such as Alzheimer's disease prediction [109], skull stripping [110], sclerosis lesion segmentation [111] and lung cancer detection and classification [112].

In this study, the purpose of using 3D images was to combine our 2D slices on a patient basis. Then, these 3D data were trained and tested using 3D CNN models.

5.2 Materials and Methods

In this study, PET scans of the same patients mentioned in the second study were used. In order to keep the data patient-based, 2D slices belonging to each patient was

combined to form a 3D data. Here, because the slice amount of each patient was different, the empty sections were completely filled with black slice. As a result, a single data of 64x64x64 size was generated for each patient.

In this study, 3D versions of LeNet, SqueezeNet, VGG16 and VGG19 models were used to work with 3D data. In order to create 3D models, convolution and pooling layers were transformed from 2 dimensions to 3 dimensions. As an example, in Figure 5.1 and Figure 5.2, 2D and 3D versions of the SqueezeNet model are shown respectively.

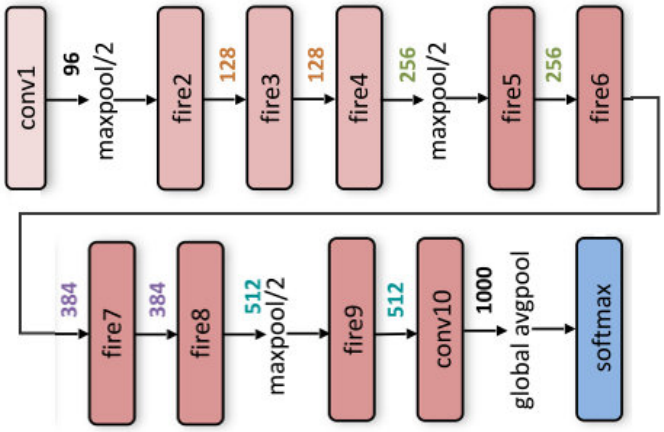

Figure 5.3 2D SqueezeNet architecture [113]

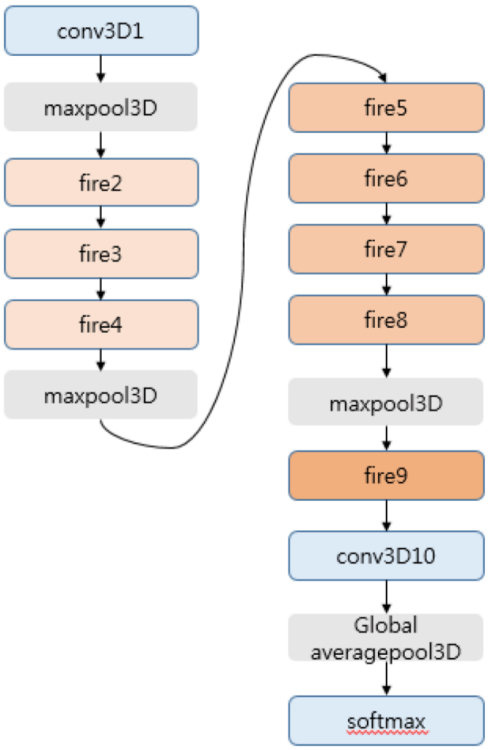

Figure 5.4 3D SqueezeNet architecture [114]

5.3 Results and Discussion

The previous study was a slice-based study, and that study had good success using 2D images and 2D models. With this study, we turned to patient-based studies in terms of a more appropriate progression to clinical studies, and one way to achieve this was to switch from 2 to 3 dimensions to keep slices patient-based. However, unfortunately, this study did not yield a good success and as seen in the Table 5.1, 60% validation accuracy was obtained in all models.

Table 5.31 Results of 3D Models

CNN Models (3D versions)	Validation Accuracy (%)
LeNet	60
SqueezeNet	60
VGG16	60
VGG19	60

As a result of combining 2D images of each patient as 3D data, a single data belonging to each patient was created, and for this reason, 94 3D data were studied instead of 1457 slices. Considering that the success of deep learning methods is often directly proportional to the amount of data, such a failure may not come as surprise. In addition, when converting 64x64 sized 2D images to 64x64x64 sized 3D images, a very large part of the 3D patient data does not contain any information, since zeros are assigned to areas other than tumor tissue. In other words, there is a tiny tumor tissue inside a large and pitch-black cube. Therefore, it has been evaluated that this may be one of the factors that make learning difficult. Apart from these, it has also been discussed that there may be a special case for the data we have.

Chapter 6

Study-4

6.1 Metabolic Imaging Based Sub-Classification of Lung Cancer

Lung cancer is one of the deadliest cancer type whose 84% is non-small cell lung cancer (NSCLC). In this study, deep learning-based classification methods were investigated comprehensively to differentiate two subtypes of NSCLC, namely adenocarcinoma (ADC) and squamous cell carcinoma (SqCC). The study used 1457 18F-FDG PET images/slices with tumor from 94 patients (88 men), 38 of which were ADC and the rest were SqCC. Three experiments were carried out to examine the contribution of peritumoral areas in PET images on subtype classification of tumors. We assessed multilayer perceptron (MLP) and three convolutional neural network (CNN) models such as SqueezeNet, VGG16 and VGG19 using three kinds of images in these experiments: 1) Whole slices without cropping or segmentation, 2) cropped image portions (square subimages) that include the tumor and 3) segmented image portions corresponding to tumors using random walk method. Several optimizers and regularization methods were used to optimize each model for the diagnostic classification. The classification models were trained and evaluated by performing stratified 10-fold cross validation, and F-score and area-under-curve (AUC) metrics were used to quantify the performance. According to our results, it is possible to say that inclusion of peritumoral regions/tissues both contributes to the success of models and makes segmentation effort unnecessary. To the best of our knowledge, deep learning-based models have not been applied to the subtype classification of NSCLC in PET imaging, therefore, this study is a significant cornerstone providing thorough comparisons and evaluations of several deep learning models on metabolic imaging for lung cancer. Even simpler deep learning models are found

promising in this domain, indicating that any improvement in deep learning models in machine learning community can be reflected well in this domain as well.

6.2 Materials and Methods

In this study, PET scans of the same patients mentioned in the second study were used. With this data, three different datasets were prepared to be used in three experiments. In the first experiment, FDG-PET images containing 168x168 pixels obtained from the scans were determined as the first dataset without any processing. For the second experiment, each FDG-PET slice was manually cropped to include tumor and peritumoral tissue, and these ROIs formed our second dataset. Here, instead of specifying a standard bounding box size, the boxes were manually cropped to the extent that the tumor could fit inside. Since tumors did not have a standard shape, the amount of peritumoral areas varied in different slices. For the last experiment, the tumors were segmented on each slice using a standard "random walk" algorithm comprising our last dataset. It was used to distinguish tumor from background semi-automatically. In an unpublished preliminary study, we have tested the performances of segmentation approaches suggested for PET images such as Otsu's, active contour and random walk methods. We found random walk approach to perform the best among these approaches when compared to manual drawing of a nuclear medicine expert. Finally, we worked with a total of 1457 images that consisted of 516 ADC and 941 SqCC subtypes for all experiments. Figure 6.1 shows sample images from different patients with ADC and SqCC.

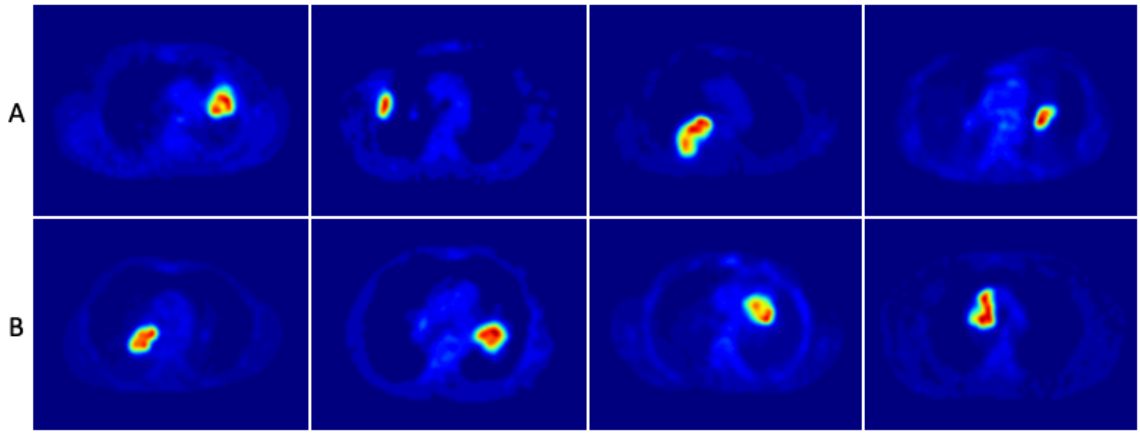

Figure 6.4 Images from different patients with (A) ADC and (B) SqCC

In order to feed the images with a uniform size to deep learning models as the input, we had to specify a common and appropriate image size. In the first experiment, all images had already a fixed size, as there was no need for further processing. The datasets prepared for the second and third experiments, we considered the largest subimage size as 64x64 pixels that all the tumors in all slices could fit inside. In the bounding box, zero padding was performed by setting all pixel values to zero except ROIs for the second experiment and except segmented regions for the last experiment. At the end of all these processes, the datasets were fed to deep learning models and the training and test procedures were performed. Figure 6.2 depicts the schematic representation of abovementioned experiments.

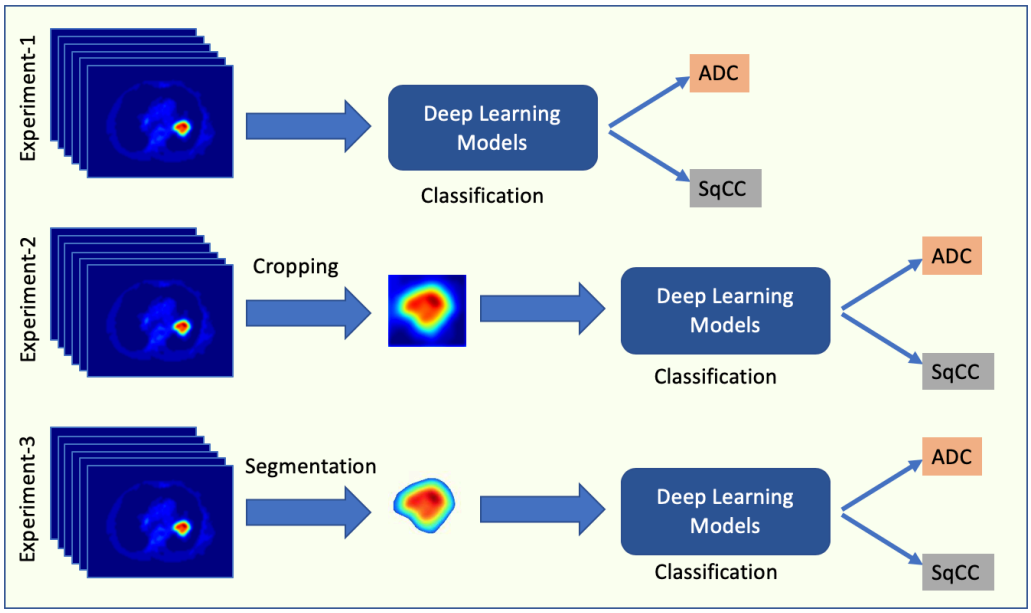

Figure 6.5 Schematic representation of three experiments studied in this work

In the literature, 10-fold cross validation method has been commonly used. However, there is no rule on this issue yet. In this study, the dataset was randomly split for stratified 10-fold cross validation which was used in many deep and conventional machine learning based bioimage and biosignal studies [29]. CNN and MLP models were trained from scratch starting with random weights using NVIDIA DGX-1 with NVIDIA Tesla V100 GPUs at Abdullah Gül University, High Performance Computing (HPC) Laboratory. The number of epochs for each training process was 100, and early stopping approach was used to prevent overfitting.

For each model mentioned above several regularization and optimization techniques were used during the training process. The parameters were optimized to increase the classification success, and are listed in Table 6.1.

Table 6.2 Hyperparameters

Hyperparameters	Values
Optimizers	Momentum / RMSprop / Adadelta / Adam
Regularizations	None / L1 / L2 / Elastic
Dropout	None / 0.5
Batch Size	16 / 32
Learning Rate	1 / 0.1 / 0.01 / 0.001 / 0.0001
Hidden Layer (only MLP)	3 / 7 / 11 / 15

In the training process, stochastic gradient descent (SGD) approach helped us to find the optimum direction for minimizing the cost. The aim of momentum, which is based on an advanced SGD logic, was to accelerate the progression in cases where the gradient did not change direction, and to slow the progression for situations where it changed direction. In addition, some popular optimization methods, such as RMSprop, Adam, and Adadelta, which are similar to momentum, were used to estimate the optimum direction and speed for cost to move towards global minima.

The regularization methods we used were lasso, ridge regression and elastic net approaches. Lasso and ridge regression are also known as L1 and L2, respectively. As the penalty term, L1 dealt with the sum of the absolute values of the model parameters, and L2 dealt with the sum of their squares. Elastic net was a convex combination of lasso and ridge regression.

The main idea of the dropout regularization, which is known to prevent overfitting problems especially for deep and complex networks with large number of parameters (such as VGG models), is to randomly drop nodes on the network during training, based on a certain ratio. In this study, the dropout with the ratio of 0.5 was used for the fully connected layers at the end of the models.

Table 6.2 shows the tuned CNN architectures where 64x64 images were used as the input. As can be seen in this table, the total number of parameters was directly proportional to the depth of the network.

Table 6.2 Tuned CNN Architectures

SqueezeNet			VGG16			VGG19		
Layer	**Output**	**Param.**	**Layer**	**Output**	**Param.**	**Layer**	**Output**	**Param.**
Input	64x64x1		Input	64x64x1		Input	64x64x1	
Conv.	31x31x64	640	2XConv.	64x64x64	37568	2XConv.	64x64x64	37568
Max Pooling	15x15x64		Max Pooling	32x32x64		Max Pooling	32x32x64	
Fire	15x15x128	11408	2XConv.	32x32x128	221440	2XConv.	32x32x128	221440
Fire	15x15x128	12432	Max Pooling	16x16x128		Max Pooling	16x16x128	
Max Pooling	7x7x128		3XConv.	16x16x256	1475328	4XConv.	16x16x256	2065408
Fire	7x7x256	45344	Max Pooling	8x8x256		Max Pooling	8x8x256	
Fire	7x7x256	49440	3XConv.	8x8x512	5899776	4XConv.	8x8x512	8259584
Max Pooling	3x3x256		Max Pooling	4x4x512		Max Pooling	4x4x512	
Fire	3x3x384	104880	3XConv.	4x4x512	7079424	4XConv.	4x4x512	9439232
Fire	3x3x384	111024	Max Pooling	2x2x512		Max Pooling	2x2x512	
Fire	3x3x512	188992	Flatten	2048		Flatten	2048	
Dropout	3x3x512		FC	4096	8392704	FC	4096	8392704
Fire	3x3x512	197184	Dropout	4096		Dropout	4096	
Dropout	3x3x512		FC	4096	16781312	FC	4096	16781312
Conv.	3x3x2	1026	Dropout	4096		Dropout	4096	
Global Average Pooling	2		FC	2	8194	FC	2	8194
Total Parameters		722370	Total Parameters		39895746	Total Parameters		45205442

In this study, MLP models with different number of hidden layers with 64 neurons were examined. The classification performance of MLP model was optimized by making its structure much deeper. For this purpose, MLP versions with 3, 7, 11 and 15 hidden layers have been generated and optimization studies have been performed. In the literature there is no rule of thumb for the number of neurons used in MLP architectures. While optimizing the number of hidden layers by using different values, we tried to optimize the number of neurons by the dropout method which randomly drops nodes from the network during the training process, based on the 0.5 ratio.

6.3 Results and Discussion

The results of three experiments carried out within the scope of this study are shown in Table 6.3. The first experiment was performed on PET images of 168x168 size without any processing. The aim was to observe the success of deep learning models when the images were subject to classification without any extra effort. The purpose of the last two experiments was to examine the contribution of peritumoral regions and segmentation effort to subtype classification. Table 6.3 shows the performance and run time values of the SqueezeNet, VGG16, VGG19 and MLP models. To evaluate the performances of models, the F-score and area-under-curve (AUC) metrics were used in terms of correct detection of NSCLC subtype.

Table 6.3 Results from three experiments showing classification performance of the models

	Experiment-1			**Experiment-2**			**Experiment-3**		
Model	Run Time (s)	F-score (%)	AUC (%)	Run Time (s)	F-score (%)	AUC (%)	Run Time (s)	F-score (%)	AUC (%)
SqueezeNet	572	54	52	508	71	66	459	70	69
VGG16	2906	68	63	1165	73	68	523	73	69
VGG19	3758	68	65	1384	**74**	**69**	779	71	70
MLP	287	65	62	231	73	69	72	70	66

According to the results of the first experiment shown in Table 6.3, it requires significantly high run time due to the size of the images. Here, we may conclude that the model performances were relatively low due to extra tissues unrelated to the tumor. According to the results of all experiments, the most successful classification performances were obtained in the second experiment in which peritumoral areas were included. Here, VGG19 was the most successful model with 74% F-score and 69% AUC. In order to demonstrate the training process of this model, the graphs of average accuracy and average loss in 10-fold cross-validation are shown in Figure 6.3. Although the number of epochs were adjusted as 100, it can be seen in the figure that maximum 60 epochs were run for VGG19 due to early stopping. VGG16 and MLP have an F-score of 73%, followed by SqueezeNet with an F-score of 71%. Also, it is seen that MLP can achieve the success of complex CNN models with less run time thanks to the positive

effect of peritumoral areas. When the second and third experiments were compared, the effect of the number of pixels containing information on run time is revealed. In the second experiment, due to the inclusion of tumor tissue as well as peritumoral areas, run time values were slightly higher than those of the third experiment. When the second and third experiments were evaluated together, it is possible to say that peritumoral regions both contribute to the success of models and make segmentation effort unnecessary. Here, the positive effect of peritumoral regions on the classification performance means an important contribution in this field. In a recent prostate study using MRI, the contribution of peritumoral areas was emphasized [115]. Therefore, we think that the effect of peritumoral areas in medical image analysis will be studied more in the future.

The hyperparameter values that led to the results given in Table 6.3 are shown in Table 6.4. According to this table, Adadelta was successful as an optimizer for SqueezeNet, while Momentum was successful on all other models. While SqueezeNet used L2 regularization to achieve the best performance, all other models reached the optimum point without using any of L1, L2 or Elastic regularization approaches. While dropout method worked for all CNN models, for MLP it did not work. The optimal batch size and learning rate values were 32 and 0.1 respectively for SqueezeNet and MLP, 16 and 0.001 for VGG models. In addition, it is not surprising that the same combinations succeeded on both VGG16 and VGG19 models, which do not differ except for the number of layers.

We recognize that we have only considered the classification of subtypes as ADC and SqCC which constitute approximately 70% of lung cancer, and neglected other possible subtypes such as large cell carcinoma (LCC). However, LCC accounts for only 5–10% [116]. That is why we have limited subtypes to the first two categories.

Since the variety of models used in the study can contribute to the comparative evaluation of the results, it may be useful to examine different models such as ResNet and DenseNet in addition to the deep learning models tested here. Furthermore, it is thought that expanding the datasets using data augmentation techniques may be useful for increasing the subtype classification success.

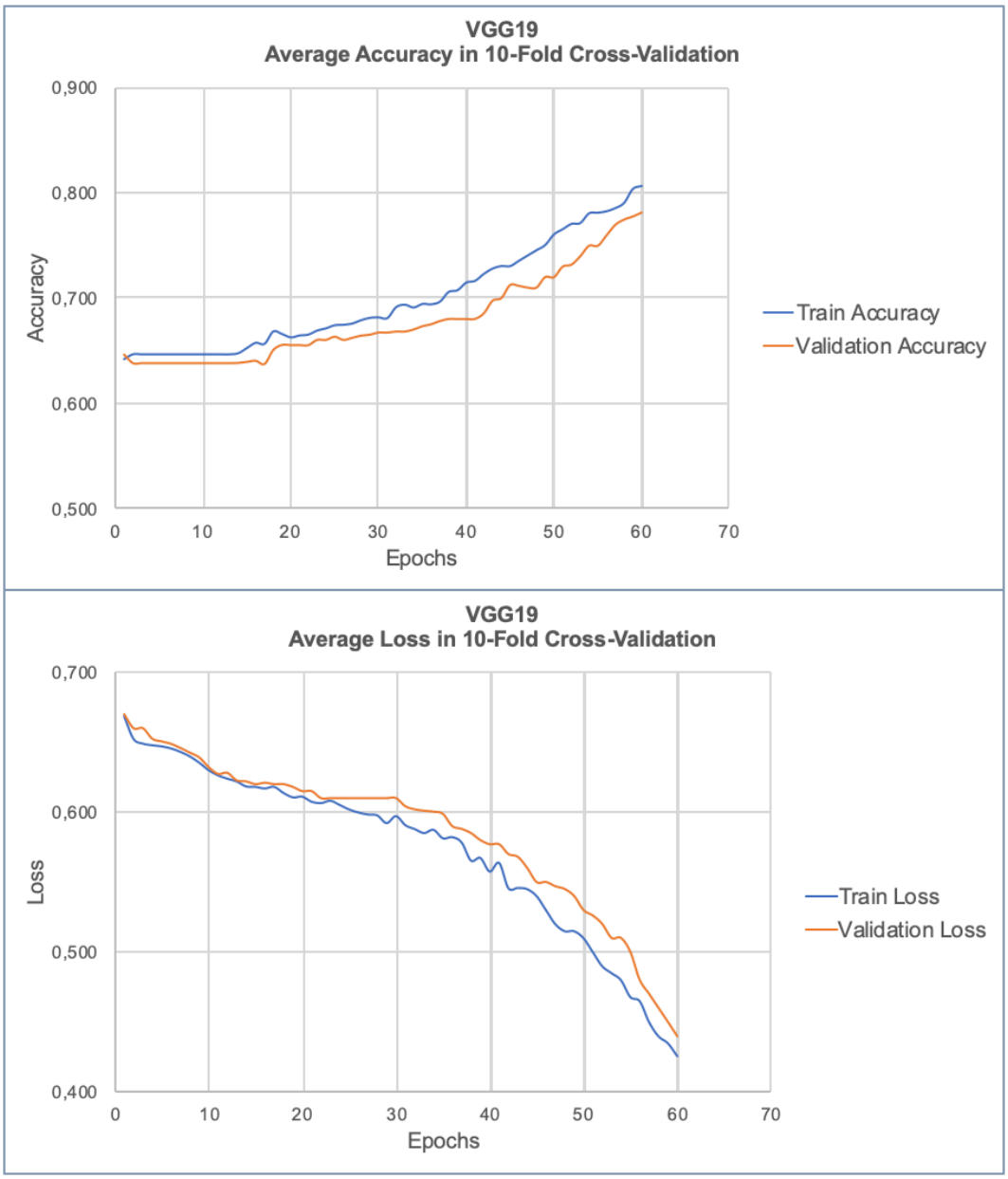

Figure 6.6 The average accuracy and loss of training and validation in the 10-fold cross-validation for VGG19 model in the second experiment

Table 6.3 Hyperparameter values yielding performances given in Table 6.3

Model	Optimizer	Regularization	Dropout	Batch Size	Learning Rate
SqueezeNet	Adadelta	L2	0.5	32	0.1
VGG16	Momentum	None	0.5	16	0.001
VGG19	Momentum	None	0.5	16	0.001
MLP	Momentum	None	None	32	0.1

A study on PET images [117] revealed that noise reduction and partial volume correction (PVC) methods improve the segmentation accuracy. Improvements in segmentation may indirectly lead to an increase in classification performance. Therefore, these methods have the potential to be examined in future classification studies with segmented images.

Moreover, the effect of different segmentation approaches may also be investigated in this context. In random walk approach the selection of seed points, one of which should be at a representative background region and the other be on the tumor region, may have an impact on the segmentation accuracy. It is clear that when the semi/automatic approach segments the tumor region incorrectly and misses critical image features related to the subtype then the classification performance would be diminished.

Currently, biopsy is still needed to characterize lung cancer subtypes. In this study, we investigated the feasibility of using PET images directly without the need for pathological studies to classify the subtypes of NSCLC which is a common effort put forward by many researchers working in this field.

This study was carried out on 1457 PET slices obtained from 94 patients. In the future, we aim to increase the success by applying data augmentation methods, and plan to make further contributions to this topic by using various deep learning models such as ResNet and DenseNet, as well as multi-task learning studies. In addition, we aim to examine the performance of deep learning models by making segmentation improvements such as noise reduction and partial volume correction. Finally, we think that better performances can be obtained by combining PET and CT images when compared to using solely PET or CT images. In our research laboratory we plan on such a study in the near future.

Chapter 7

Study-5

7.1 Transfer Learning Studies on Subtype Classification of NSCLC

The medical literature has been growing exponentially, especially in recent years [118]. Thanks to the contributions of the medical imaging studies to fields such as monitoring, diagnosis and treatment, clinical solutions of diseases have become more practical [119-121]. With the development and widespread use of various medical imaging methods, the number of medical images is increasing day by day. However, it is still very difficult to find enough labeled medical data in terms of performing artificial intelligence studies specific to a certain problem.

Convolutional neural networks (CNN) are designed to provide solutions to problems such as recognition, segmentation and classification by learning the attributes of images. ImageNet, currently a very large dataset of 14 million images, has become a reference point for different CNN studies [58], [59], [122], [123] and benchmarks. The presence of large datasets is effective in getting successful results on general object recognition by facilitating learning [32]. But for real-world problems such as medical studies, it is not always possible to find large labelled datasets and bear the cost of "training from scratch". Therefore, the use of pre-trained models is a good option, rather than "training from scratch" with little data [124], [125]. In addition, fine-tuning studies can be performed using different levels of features obtained from pretrained models as a starting point for training [126], [127].

In the literature, there are transfer learning studies [128-130] as well as "training from scratch" studies [131-135] on the classification of medical images.

In this study, on the problem of subtype classification of NSCLC, VGG16, ResNet50 and DenseNet121 models trained with ImageNet dataset were used for transfer

learning and fine tuning. In addition, data augmentation studies have been carried out because we have little data. In addition to applying momentum and RMSprop as an optimizer, overfitting was tried to be prevented by dropout method.

7.2 Materials and Methods

In this study, PET scans of the same patients mentioned in the second study were used. With the second experiment we conducted in the fourth study, we demonstrated the effect of peritumoral areas. Therefore, in this study, the dataset used in that experiment was used. Here, each FDG-PET slice was manually cropped to include tumor and peritumoral tissue, and these ROIs formed our dataset. Since it is necessary to enter 3D data into the model in transfer learning studies, firstly, the grayscale images were filled with the same gray values corresponding to the RGB channels, and thus the data reached 64x64x3. We used MATLAB for these operations.

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. In transfer learning, we first train a base network on a base dataset and task, and then we repurpose the learned features, or transfer them, to a second target network to be trained on a target dataset and task. This process will tend to work if the features are general, meaning suitable to both base and target tasks, instead of specific to the base task [136]. Here, the features obtained from the transferred model are used in a new classifier.

The feature extractor part seen in Figure 7.1 represents the part to which we transfer the features, while the classifier part at the end represents the part we just added and trained on. The feature extractor part seen in the figure represents the part to which we transfer the features, while the classifier part at the end represents the part we just added and trained on.

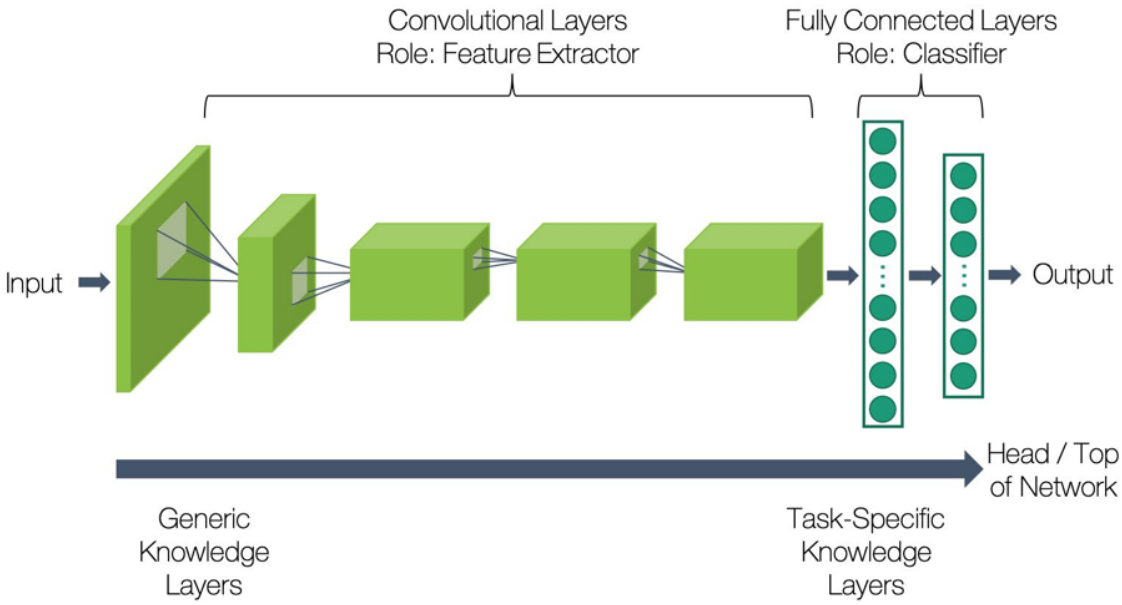

Figure 7.3 A representation of CNN model [137]

The steps for deep learning are as follows [138]:

1. Take layers from a pre-trained model.
2. Freeze these layers to prevent information loss in future trainings.
3. Add new trainable layers on top of frozen layers.
4. Train the new layers on the current dataset.

In this study, for transfer learning and fine tuning, we used VGG16 [59], ResNet50 [139] and DenseNet121 [140] models, originally trained with ImageNet data. In transfer learning works, we only took the convolutional blocks from the models. After that, we have carried out the training process by adding new fully connected layers at the end. In other words, we froze the convolutional blocks and made only newly added fully connected layers trainable. Fine-tuning consists of unfreezing the entire model obtained (or part of it), and re-training it on the new data with a very low learning rate. This is an optional last step that can potentially achieve meaningful improvements, by incrementally adapting the pretrained features to the new data. It could also potentially lead to quickly overfitting [138]. Figure 7.2 shows the fine-tuning process for the VGG16 model as an example. Here, the last convolutional block and the newly added fully connected layer were left as trainable and the rest was frozen. In fine-tuning step, we made many experiments by unfreezing convolution blocks one by one and making them trainable. In addition to using the dropout method during our transfer learning study, we also used RMSprop and momentum optimizers.

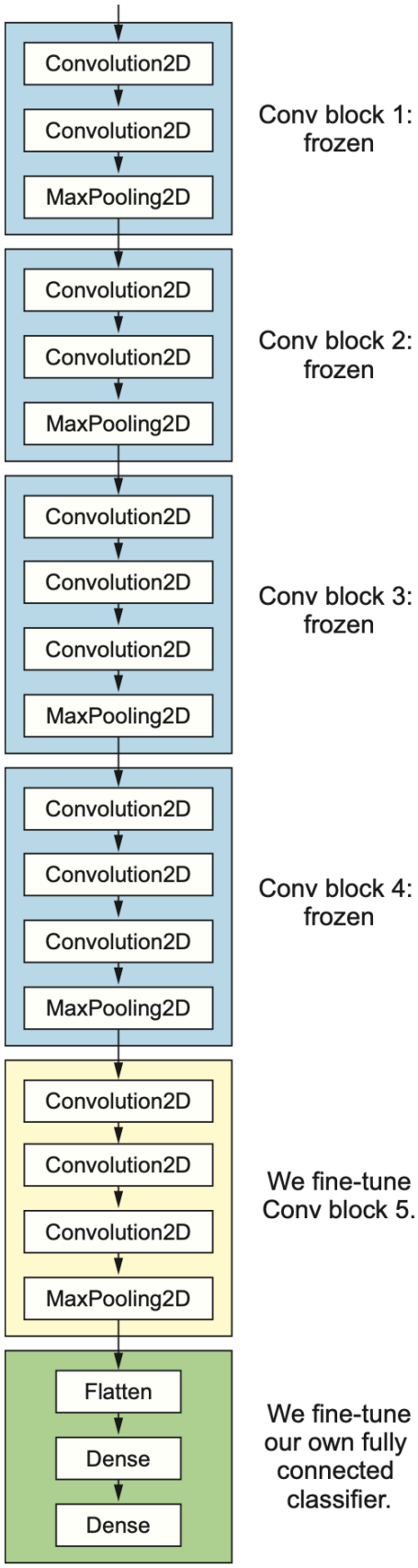

Figure 7.4 Fine tuning the last convolutional block of the VGG16 network [141]

Data augmentation is the artificial expansion of a dataset consisting of a small number of images for deep learning studies. Keras library has ImageDataGenerator class that covers different augmentation techniques [142].

In augmentation works, we have focused on some main types of data augmentation techniques for image data [143]:

- rotation_range is a value in degrees (0-180), a range within which to randomly rotate images.
- width_shift and height_shift are ranges (as a fraction of total width or height) within which to randomly translate images vertically or horizontally.
- horizontal_flip is for randomly flipping half of the images horizontally.
- vertical_flip is for randomly flipping half of the images vertically.

The values we used for these techniques are as follows:

- rotation_range=20
- width_shift_range=0.2
- height_shift_range=0.2
- horizontal_flip=True
- vertical_flip=True

For all these transfer learning and fine-tuning studies, 10-fold stratified cross-validation was applied in the training processes. The training process, which was applied as 100 epochs, was stopped at the most successful point with early stopping.

7.3 Results and Discussion

In this work, transfer learning and fine-tuning studies were performed with VGG16, ResNet50 and DenseNet121 pretrained models. Later, some augmentation methods were also included in these, and training and test processes were carried out again. In all these processes, dropout technique and some optimizers such as RMSprop and momentum were used. Table 7.1 shows the results with and without augmentation. As can be seen in this table, despite the use of augmentation, 65% validation accuracy, 52% F-score and 50% AUC values were obtained for all models.

Table 7.3 Results of transfer learning works

Models	Val. Acc. (%)	Precision (%)	Recall (%)	F-score (%)	AUC (%)
VGG16	65	60	65	52	50
ResNet50	65	60	65	52	50
DenseNet121	65	55	65	52	50
VGG16-aug	65	60	65	52	50
ResNet50-aug	65	60	65	52	50
DenseNet121-aug	65	60	65	52	50

Possible reasons for the failure of the study as a result of the evaluation are as follows:

1. The type of image in which the transferred models were used is completely different from the images we have. The features that the models were obtained from ImageNet data cannot help to solve the problem of classifying the medical images we have.
2. Considering the failure of 3D data studies in the third chapter, it is possible to say that 3D studies do not work, specific to the data we have.

Chapter 8

Study-6

8.1 Shallow Network Studies on Subtype Classification of NSCLC

Classification studies were carried out by training from scratch with SqueezeNet, VGG16, VGG19 and MLP in Study 4, and by transfer learning and fine tuning with VGG16, ResNet50 and DenseNet121 in Study 5. When the results of VGG16 for these two studies are compared, it is clear that training from scratch is more successful than transfer learning for the dataset we have.

If the dataset is small, for training from scratch, it may make sense to use shallower and smaller networks rather than using very deep networks. There are studies in the literature that have achieved successful results on medical data, especially using shallow networks [144] [145].

In this study, the CNN-based and ResNet-based shallow networks we designed have been investigated on the subtype classification problem. In other words, the success potential of shallow CNN (CNN-S) and shallow ResNet (RN-S) models instead of standard deep networks known in the literature, such as the models in Study 4, were examined.

8.2 Materials and Methods

In this study, PET scans of the same patients mentioned in the second study were used. With the second experiment we conducted in the fourth study, we demonstrated the effect of peritumoral areas. Therefore, in this study, the dataset used in that experiment was used. Here, each FDG-PET slice was manually cropped to include tumor and peritumoral tissue, and these ROIs formed our dataset.

In this study, two types of networks that are not very deep and complex were created by using the basic structures in the architectures used for image-based deep learning studies. While one of these networks, CNN-S is a CNN formed as a result of repeating the block consisting of 2 convolution and 1 max pooling layers 2 times consecutively (like first two blocks in VGG models), the other network is RN-S that consists of the first 2 stages of the classical ResNet. Later, a fully connected layer was added to these networks. Figure 8.1 and Figure 8.2 show the CNN-S and RN-S structure, respectively.

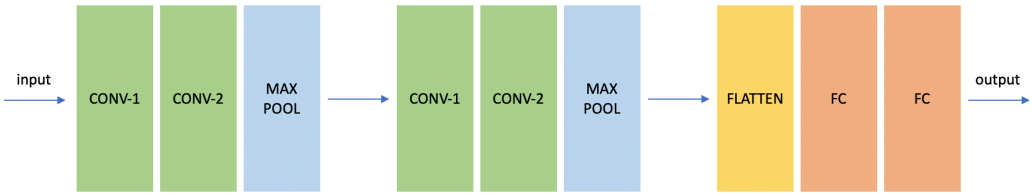

Figure 8.1 CNN-S architecture

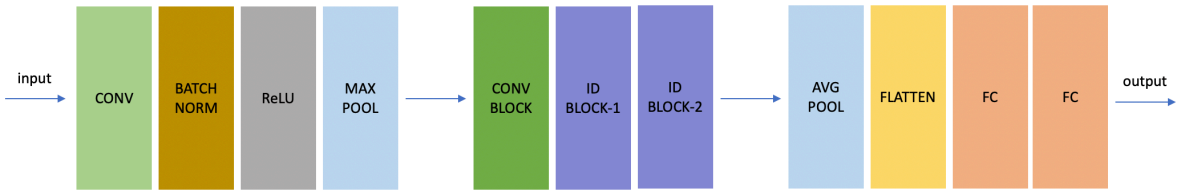

Figure 8.2 RN-S architecture

After the models were created, 100 epoch training was done by applying early stopping. During the training, some optimizers were applied to increase the performance. These optimizers were Momentum, RMSprop and Adam. Also, 10-fold stratified cross-validation was applied in the training processes.

8.3 Results and Discussion

In this study, 2 types of models are designed to train from scratch. These models are based on classical CNN and ResNet structure and have a shallow architecture. The performance metrics obtained using the CNN-S model is depicted in Table 1 and the RN-S model in Table 2. These tables also show the effect of optimizers. Here, model successes were calculated by averaging the 10-fold stratified cross validation results.

Table 8.3 Results of CNN-S model

Model	Optimizer	Learning Rate	Val_Acc (%)	Precision (%)	Recall (%)	F-Score (%)	AUC (%)
CNN-S	RMSprop	0.001	73	70	73	**71**	66
CNN-S	Momentum	0.001	72	68	72	69	63
CNN-S	Adam	0.0001	71	73	71	68	64

Table 8.3 Results of RN-S model

Model	Optimizer	Learning Rate	Val_Acc (%)	Precision (%)	Recall (%)	F-Score (%)	AUC (%)
RN-S	Momentum	0.001	73	73	73	**71**	67
RN-S	RMSprop	0.01	73	73	73	**71**	65
RN-S	Adam	0.001	72	73	72	70	66

As can be seen from the tables, the highest F-score value for both models was 71%. In the RN-S model, all optimizers achieved an F-score of 70% or more, while Momentum and Adam were below 70% in the CNN-S model. In addition, learning rate values can be seen in the tables. For the CNN-S model, RMSprop was the most successful optimizer with a learning rate of 0.001, while for the RN-S model, momentum yielded the best results with the same learning rate.

We obtained 74% F-score success with VGG19, which is a very deep network in Study 4. In that study, we achieved this success by doing intense hyperparameter optimization work with the grid search method. In this study, 71% F-score value was obtained with shallow networks by using only limited number of optimizers and learning rate values. In this respect, it is possible to say that comprehensive hyperparameter optimization works with different shallow network variations have the potential to yield good results.

Chapter 9

Conclusions and Future Prospects

9.1 Conclusions

In the literature, there are some studies on the subtype classification of NSCLC, although there is no study that exactly overlaps with our thesis study. In 2013, more than 90% success was achieved in a study using hematoxylin and eosin tissue images with some classical machine learning methods, including ANN [146]. In a study conducted in 2018 [147], ADC and SqCC differentiation was performed with a radiomic signature containing five quantitative CT image features. At the end of this study, which achieved 89% validation success rate, it was evaluated that PET could increase the success by containing functional information about the tumor. In a different study [148] conducted in 2019, subtypes of NSCLC were classified with a hybrid method using features extracted from CT images and achieved 86% test accuracy. In a study using PET / CT images, ten machine learning methods were examined in addition to the VGG16 model for the problem of differentiating ADC and SqCC [149]. The VGG16 surpassed all other machine learning methods with a 90% AUC. In addition, it was stated in this study that PET images provide advantages in reflecting metabolic heterogeneity. In a different study [150] conducted on the same problem, some machine learning methods were examined using exhaled breath data. As a result, the highest AUC value achieved with the K-nearest neighbor classifier combining borderline2-SMOTE and feature reduction methods was 63%. In another study [151] conducted in 2020, 78% test success was achieved with some classical machine learning methods using CT images. In a study [152] examining the classification ability of deep learning methods using CT images, F-score values of more than 70% were obtained. The study [153] performed using radiomic features obtained from PET images in 2021 increased over 85% and in this study the performance of stage-specific PET radiomic prediction models was emphasized.

In this thesis, classification of two subtypes of NSCLC, ADC and SqCC, has been investigated. The importance of this thesis is to perform subtype tumor classification from only PET images without the need for a pathological procedure. In order to provide a successful solution to this problem, several methods were tried and many optimization and regularization studies were carried out on these methods. These studies have been examined under 6 different main headings/chapters.

In the first study, classical ANN methods were applied on subtype classification. As input to ANN, 39 features obtained from PET images using GLCM, GLRLM, GLSZM and NGTDM techniques which are popular texture analysis based feature extraction methods, were provided. Later, training activities were carried out on ANN. While doing this, the FANN library and the FANN Tool developed for this library were used. During the training phase, incremental, batch, RPROP and quickprop algorithms, which are four training algorithms in the FANN library, were used, while the activation functions were sigmoid, sigmoid stepwise, sigmoid symmetric and sigmoid symmetric stepwise functions, which are also available in this library. As a result, quickprop training algorithm was found to be the most successful of the training algorithms. ANN is considered the foundation of deep learning. Therefore, this study is important as it encouraged us to switch to deep learning methods.

In the second study, a solution was sought for subtype classification with a slice-based approach. Here, MLP and four CNN models LeNet, SqueezeNet, VGG16 and VGG19 models were used for the training and test of a total of 1457 segmented PET slices belonging to patients. At the end of this study, highly successful results were obtained in all models. The VGG16 and VGG19 models, which are the most complex ones and have the highest run times, yielded the best results with an F-score of 95%. Therefore, it has emerged that working with more complex and costly networks should be considered in order to achieve higher success.

The third study was the first transition from slice-based studies to patient-based studies. Here, the idea of combining 2D slices as 3D images has emerged, which is considered as a solution to separate the data as patient-based. 3D versions of LeNet, SqueezeNet, VGG16 and VGG19 models were used in this study. Unfortunately, sufficient performance was not achieved as a result of this study. It has been argued that the reason for this failure may be that data, which is already scarce for deep learning methods, is further reduced as a result of 3D transformation.

The fourth study compares three different experiments on the same subtype classification problem. In the first experiment, PET images are given directly to the networks without any processing. In the second experiment, the images obtained as a result of cropping the same PET images including the peritumoral areas are used. And finally, it is an experiment in which tumor tissues segmented with random walk are used. We investigated the feasibility of several popular deep learning methodologies. To the best of our knowledge, deep learning-based models have not been applied to subtype classification of NSCLC in PET imaging domain. Therefore, this study is a significant contribution providing thorough comparisons and evaluations of several deep learning models on metabolic imaging for lung cancer. Even simpler deep learning models were found to be promising in this field, indicating that any improvement in deep learning models in machine learning community can be reflected well in this domain as well. In addition, no studies investigating the contribution of peritumoral areas to subtype classification of NSCLC using deep learning on PET images were found in the literature. From this perspective, this study includes reviews that shed light on similar studies in the future. As a result, this study was deemed worthy of publication due to its contribution to the literature [5].

In the fifth study, instead of training CNN models from scratch, pretrained models were examined by transfer learning. In addition, the contribution of pretrained features at different levels was examined by performing fine tuning. During all these studies, in addition to some optimization methods, data augmentation was applied to increase the success. As a result, despite all this effort, successful results have not been achieved. Here it was emphasized that there may be incompatibilities specific to the available data and the problem. In addition, considering that the 3D transformations in the third study do not have a positive effect on success, it is possible to say that 3D studies do not work, specific to the data we have.

In the last study, two shallow networks based on classical CNN and ResNet structures were created and successful results were tried to be obtained with the little data we have. In this study, the success of shallow networks instead of complex networks on small datasets is investigated. Because standard complex networks in the literature may not always be successful on small datasets. In this case, successful results can be obtained by obtaining shallower models by reducing and modifying the nets. Despite the lack of comprehensive hyperparameter optimization works, this study is promising considering the success achieved.

9.2 Societal Impact and Contribution to Global Sustainability

According to the cancer report published by the World Health Organization in 2020, more than 18 million people are diagnosed with cancer each year, and it is estimated that this number will increase to 29.4 million by 2040. Also, according to this report, lung cancer is the most frequently diagnosed (11.6% of all cases) and has the highest mortality rate (18.4% of all deaths) among all types of cancer. In addition, patients with lung cancer have a high symptom burden and often become debilitated after treatments. All these bad effects cause psychological and mental depressions on the family and social environment of the patients. Considering all these, the importance of research and investments in lung cancer is obvious. Early diagnosis studies, one of the most effective public health measures in cancer, cover rapid clinical and pathological diagnosis and referral to appropriate treatment. In this study, it was aimed to accelerate clinical processes in diagnosis and to support subtype-specific treatments. This study is very important for public health, as it is focused on PET imaging, which exposes a relatively lower radiation dose than CT scans and is a good option against the risk of infection due to medical procedures. In addition, PET imaging examines metabolic functions and is shown as an alternative to biopsy in this sense. Since biopsy has many disadvantages for patients and specialists, one of our goals was to reduce the workload of the experts and the rate of misdiagnosis. However, currently, a biopsy is still required to characterize lung cancer subtypes. In this thesis, the classification of ADC and SqCC, subtypes of NSCLC, is tried to be performed automatically. Combining this information obtained from PET images with information from conventional diagnostic methods will create a strong decision support mechanism. In this way, rapid and reliable diagnosis will be provided, and guidance to subtype-specific appropriate treatments will be provided.

9.3 Future Prospects

All studies in this thesis are on the classification of ADC and SqCC, which account for approximately 70% of lung cancer, and other types have been ignored. In addition, studies such as staging and prognosis were excluded from the thesis, apart from the classification problem, due to the insufficient data available. In addition, the imbalance

in the dataset was an important limitation for classification studies. Within the scope of our thesis studies, many deep learning studies have been conducted on the problem of subtype classification. In the light of these studies, the importance of various methods such as regularization, optimization and augmentation as well as preprocessing studies such as segmentation and the inclusion of peritumoral areas in the image have emerged. Since the variety of models used in the study can contribute to the comparative evaluation of the results, it may be useful to examine different models in addition to the deep learning models used here. In addition, expanding the dataset by increasing the labeled data can increase the success of deep learning. Comprehensive projects involving radiologists may be required for this. By analyzing PET and CT images together, problems such as classification, staging and prognosis can be solved automatically. In addition, real time responses will be possible to obtain thanks to a system that will be integrated into PET / CT scanning devices. Thus, decision support will be provided for rapid and reliable diagnosis as well as reducing workload for experts. Although biopsy is still required currently, the need for biopsy will decrease with the increase of such studies in the future. Therefore, complications related to biopsy such as bleeding, bruising and infection will be eliminated.

BIBLIOGRAPHY

[1] Global cancer observatory (GCO), https://gco.iarc.fr/today/home (29 December 2020).

[2] A. Agrawal, V. Rangarajan, "PET/CT in lung cancer," Springer International Publishing (2018).

[3] R. F. Munden, S. S. Swisher, C. W. Stevens and D. J. Stewart, "Imaging of the patient with non–small cell lung cancer," Radiology, 237, 803-818 (2005).

[4] O. Ayyildiz, Z. Aydin, B. Yilmaz, S. Karaçavuş, K. Şenkaya, S. İçer, A. Taşdemir and E. Kaya, "Lung cancer subtype differentiation from positron emission tomography images," Turk J Elec Eng & Comp Sci, 28, 262–274 (2020).

[5] M. Bicakci, O. Ayyildiz, Z. Aydin, A. Basturk, S. Karacavus and B. Yilmaz, "Metabolic imaging based sub-classification of lung cancer," IEEE Access, 8, 218470-218476 (2020).

[6] Non-small cell lung cancer treatment, https://www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq (29 December 2020).

[7] W. De Wever, J. Coolen and J.A. Verschakelen, "Imaging techniques in lung cancer," Breathe, 7, 338-346 (2011).

[8] C. Schaefer-Prokop and M. Prokop, "New imaging techniques in the treatment guidelines for lung cancer," European Respiratory Journal, 19, 71-83 (2002).

[9] S. L. S. MacDonald, D. M. Hansell, "Staging of non-small cell lung cancer: imaging of intrathoracic disease," European Journal of Radiology, 45, 18-30 (2003).

[10] X-ray (Radiography) - Chest, https://www.radiologyinfo.org/en/info.cfm?pg=chestrad (29 December 2020).

[11] Computed tomography (CT), https://www.nibib.nih.gov/sites/default/files/2020-06/Computed_Tomography_Fact_Sheet.pdf (29 December 2020).

[12] Early detection, diagnosis, and staging, https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html (29 December 2020).

[13] CT lung screening, https://www.cedars-sinai.org/programs/imaging-center/exams/ct-scans/lung.html (29 December 2020).

[14] J. F. Vansteenkiste and S. S. Stroobants, "PET scan in lung cancer: Current recommendations and innovation," Journal of Thoracic Oncology, 1, 71–73 (2006).

[15] Liesbet Schrevens, Natalie Lorent, Christophe Dooms, Johan Vansteenkiste, "The role of PET scan in diagnosis, staging, and management of non-small cell lung cancer," The Oncologist, 9, 633-643 (2004).

[16] J. F. Vansteenkiste, "Imaging in lung cancer: positron emission tomography scan," European Respiratory Journal, 19, 49-60 (2002).

[17] E. F. Patz Jr. and J. J. Erasmus, "Positron emission tomography imaging in lung cancer," Clinical Lung Cancer, 1, 42-48 (1999).

[18] PET scans (Positron emission tomography), http://www.aboutcancer.com/lung_nodes_pet_DH_sah_sep_2006.jpg (29 December 2020).

[19] L. S. Freudenberg, S. J. Rosenbaum, T. Beyer, A. Bockisch, G. Antoch, "PET versus PET/CT dual-modality imaging in evaluation of lung cancer," PET clinics, p. 347–352 (2006).

[20] F. C. Detterbeck, D. J. Boffa, and L. T. Tanoue, "The new lung cancer staging system," Chest, 136, 260–271 (2009).

[21] V. K. Anagnostou, A. T. Dimou, T. Botsis, E. J. Killiam, M. D. Gustavson, R. J. Homer, D. Boffa, V. Zolota, D. Dougenis, L. Tanoue, S. N. Gettinger, F. C. Detterbeck, K. N. Syrigos, G. Bepler, and D. L. Rimm, "Molecular classification of nonsmall cell lung cancer using a 4-protein quantitative assay," Cancer, 118, 1607–1618 (2012).

[22] W. D. Travis, E. Brambilla, A. G. Nicholson, Y. Yatabe, J. H. M. Austin, M. B. Beasley, L. R. Chirieac, S. Dacic, E. Duhig, D. B. Flieder, K. Geisinger, F. R. Hirsch, Y. Ishikawa, K. M. Kerr, M. Noguchi, G. Pelosi, C. A. Powell, M. S. Tsao, and I. Wistuba, "The 2015 world health organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification," J. Thorac. Oncol., 10, 1243–1260 (2015).

[23] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. W. L. Aerts, A. Dekker, D. Fenstermacher, D. B. Goldgof, L. O. Hall, P. Lambin, Y. Balagurunathan, R. A. Gatenby, R. J. Gillies, "Radiomics: The process and the challenges," Magn. Reson. Imag., 30, 1234–1248 (2012).

[24] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Comput. Struct. Biotechnol. J., 13, 8-17 (2015).

[25] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, "Machine learning methods for quantitative radiomic biomarkers," Sci. Rep., 5, 1–11 (2015).

[26] M. P. Menden, F. Iorio, M. Garnett, U. McDermott, C. H. Benes, P. J. Ballester, and J. Saez-Rodriguez, "Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties," PLoS ONE, 8, e61318 (2013).

[27] A. C. Haury, P. Gestraud, and J. P. Vert, "The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures," PLoS ONE, 6, 1-12 (2011).

[28] L. Matulewicz, J. F. A. Jansen, L. Bokacheva, H. A. Vargas, O. Akin, S. W. Fine, A. Shukla-Dave, J. A. Eastham, H. Hricak, J. A. Koutcher, and K. L. Zakian, "Anatomic segmentation improves prostate cancer detection with artificial neural networks analysis of 1H magnetic resonance spectroscopic imaging," J. Magn. Reson. Imag., 40, 1414–1421 (2014).

[29] H. G. Zadeh, J. Haddadnia, M. Hashemian, and K. Hassanpour, "Diagnosis of breast cancer using a combination of genetic algorithm and artificial neural network in medical infrared thermal imaging," Iran. J. Med. Phys., 9, 265–274 (2012).

[30] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, 521, 436-444 (2015).

[31] J. Deng, W. Dong, R. Socher, L. -J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," IEEE Conference on Computer Vision and Pattern Recognition, 248-255 (2009).

[32] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," Int. J. Comput. Vis., 115, 211–252 (2015).

[33] P. Danaee, R. Ghaeini, and D. A. Hendrix, "A deep learning approach for cancer detection and relevant gene identification," Biocomputing, 53, 219–229 (2017).

[34] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, 542, 115–118 (2017).

[35] G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen- van de Kaa, P. Bult, B. van Ginneken, and J. van der Laak, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," Sci. Rep., 6, 26286 (2016).

[36] Y. Yuan, Y. Shi, C. Li, J. Kim, W. Cai, Z. Han, and D. D. Feng, "DeepGene: An advanced cancer type classifier based on deep learning and somatic point mutations," BMC Bioinf., 17, 476 (2016).

[37] M. Halicek, G. Lu, J. V. Little, X. Wang, M. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, and B. Fei, "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging," J. Biomed. Opt., 22, 060503 (2017).

[38] F. Ciompi, K. Chung, S. J. van Riel, A. A. A. Setio, P. K. Gerke, C. Jacobs, E. T. Scholten, C. Schaefer-Prokop, M. M. W. Wille, A. Marchiano, U. Pastorino, M. Prokop, and B. van Ginneken, "Towards automatic pulmonary nodule management in lung cancer screening with deep learning," Sci. Rep., 7, 1–11, (2017).

[39] Q. Song, L. Zhao, X. Luo, and X. Dou, "Using deep learning for classification of lung nodules on computed tomography images," J. Healthc. Eng., 2017, 1-10, (2017).

[40] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Med. Image Anal., 42, 60–88 (2017).

[41] Chiranji Lal Chowdhary, D. P. Acharjya, "Segmentation and feature extraction in medical imaging: A systematic review," Procedia Computer Science, 167, 26-36 (2020).

[42] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. W. L. Aerts, A. Dekker, D. Fenstermacher, D. B. Goldgof, L. O. Hall, P. Lambin, Y. Balagurunathan, R. A. Gatenby, and R. J. Gillies, "Radiomics: the process and the challenges," Magnetic Resonance Imaging, 30, 1234-1248 (2012).

[43] B. A. Altazi, G. G. Zhang, D. C. Fernandez, M. E. Montejo, D. Hunt, J. Werner, M. C. Biagioli, E. G. Moros, "Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms," Journal of Applied Clinical Medical Physics, 18, 32-48 (2017).

[44] P. Mohanaiah, P. Sathyanarayana, L. GuruKumar, "Image texture feature extraction using GLCM approach," International Journal of Scientific and Research Publications, 3, 290-294 (2013).

[45] B. Ergen and M. Baykara, "İstatistiksel uzaysal alan metotlarının içerik tabanlı tıbbi görüntü erişimi için bir uygulama," Fırat Üniv. Mühendislik Bilimleri Dergisi, 2, 87-93 (2011).

[46] D.-H. Xu, A. S. Kurani, J. D. Furst, D. S. Raicu, "Run-length encoding for volumetric texture," Heart, 27, 452-458 (2004).

[47] G. Thibault, B. Fertil, C. Navarro, S. Pereira, P. Cau, N. Lévy, J. Sequeira and J.-L. Mari, "Texture indexes and gray level size zone matrix application to cell nuclei classification," 10th International Conference on Pattern Recognition and Information Processing, 140-145 (2009).

[48] M. Amadasun and R. King, "Textural features corresponding to textural properties," IEEE Transactions on Systems, Man, and Cybernetics, 19, 1264-1274 (1989).

[49] L. Hertel, E. Barth, T. Käster and T. Martinetz, "Deep convolutional neural networks as generic feature extractors," 2015 International Joint Conference on Neural Networks (IJCNN), 1-4 (2015).

[50] Step 1 – Convolution operation, https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-1-convolution-operation (05 January 2021).

[51] Step 1(B): The rectified linear unit (RELU), https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-1b-relu-layer/ (05 January 2021).

[52] Step 2 – Max pooling," https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-2-max-pooling/ (05 January 2021).

[53] Step 3: Flattening," https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-3-flattening/ (05 January 2021).

[54] Step 4: Full connection, https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-4-full-connection (05 January 2021).

[55] Summary, https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-summary/ (05 January 2021).

[56] I. N. da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni and S. F. dos Reis Alves, "Artificial neural network architectures and training processes," Artificial Neural Networks, Springer, Cham, 21-28 (2017).

[57] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, 86, 2278-2324, (1998).

[58] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," NIPS (2012).

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

[60] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," in arXiv:1602.07360 (2016).

[61] Derin öğrenme uygulamalarında en sık kullanılan hiper-parametreler, https://medium.com/deep-learning-turkiye/derin-ogrenme-uygulamalarinda-en-sik-kullanilan-hiper-parametreler-ece8e9125c4 (10 January 2021).

[62] N. Qian, "On the momentum term in gradient descent learning algorithms," Neural Networks, 12, 145-151 (1999).

[63] T. Tieleman and G. Hinton, "Lecture6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude," Coursera, Neural Netw. Mach. Learn., 4, 26–31 (2012).

[64] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," arXiv preprint arXiv:1212.5701 (2012).

[65] J. B. Diederik, P. Kingma, "Adam: A method for stochastic optimization," ICLR (2015).

[66] R. Tibshirani, "Regression shrinkage and selection via the lasso," J.Roy. Stat. Soc., Ser. B, Methodol., 58, 267–288 (1996).

[67] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, 12, 55-67 (1970).

[68] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research, 15, 1929–1958 (2014).

[69] J. Ge, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, J. Wei, M. A. Helvie and C. Zhou, "Computer aided detection of clusters of microcalcifications on full field digital mammograms," Medical Physics, 33, 2975-2988 (2006).

[70] S.-C. B. Lo, H.-P. Chan, J.-S. Lin, H. Li, M. T. Freedman and S. K. Mun, "Artificial convolution neural network for medical image pattern recognition," Neural Networks, 8, 1201-1214 (1995).

[71] R. H. Nagel, R. M. Nishikawa, J. Papaioannou and K. Doi, "Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms," Medical Physics, 25, 1502-1506 (1998).

[72] A. Papadopoulos, D.I. Fotiadis and A. Likas, "An automatic microcalcification detection system based on a hybrid neural network classifier," Artificial Intelligence in Medicine, 25, 149-167 (2002).

[73] I. Christoyianni, E. Dermatas and G. Kokkinakis, "Fast detection of masses in computer-aided mammography," IEEE Signal Processing Magazine, 17, 54-64 (2000).

[74] A. C. Patrocinio, H. Schiabel, R. H. Benatti, C. E. Goes and F. L. S. Nunes, "Investigation of clustered microcalcification features for an automated classifier as part of a mammography CAD scheme," Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No.00CH37143), Chicago (2000).

[75] R. Setiono, "Extracting rules from pruned neural networks for breast cancer diagnosis," Artificial Intelligence in Medicine, 8, 37-51 (1996).

[76] X. W. Xu, K. Doi, T. Kobayashi, H. MacMahon, M. L. Giger, "Development of an improved CAD scheme for automated detection of lung nodules in digital chest images," Medical Physics, 24, 1395-1403 (1997).

[77] Z.-H. Zhou, Y. Jiang, Y.-B. Yang and S.-F. Chen, "Lung cancer cell identification based on artificial neural network ensembles," Artificial Intelligence in Medicine, 24, 25-36 (2002).

[78] B. Keserci and H. Yoshida, "Computerized detection of pulmonary nodules in chest radiographs based on morphological features and wavelet snake model," Medical Image Analysis, 6, 431-447 (2002).

[79] S. Kobashi, N. Kamiura, Y. Hata and F. Miyawaki, "Volume-quantization-based neural network approach to 3D MR angiography image segmentation," Image and Vision Computing, 19, 185-193 (2001).

[80] I. Middleton and R. I. Damper, "Segmentation of magnetic resonance images using a combination of neural networks and active contour models," Medical Engineering & Physics, 26, 71-86 (2004).

[81] J. Lin, "Segmentation of medical images through a penalized fuzzy Hopfield network with moments preservation," Journal of the Chinese Institute of Engineers, 23, 633-643 (2000).

[82] K. C.-R. Lin, M.-S. Yang, H.-C. Liu, J.-F. Lirng and P.-N. Wang, "Generalized Kohonen's competitive learning algorithms for ophthalmological MR image segmentation," Magnetic Resonance Imaging, 21, 863-870 (2003).

[83] Z. Dokur and T. Ölmez, "Segmentation of MR and CT images by using a quantiser neural network," Neural Computing Applications, 11, 168–177 (2003).

[84] G. Stalidis, N. Maglaveras, S. N. Efstratiadis, A. S. Dimitriadis and C. Pappas, "Model-based processing scheme for quantitative 4-D cardiac MRI analysis," IEEE Transactions on Information Technology in Biomedicine, 6, 59-72 (2002).

[85] C. Y. Chang and P. C. Chung, "Two-layer competitive based Hopfield neural network for medical image edge detection," Optical Engineering, 39, 695-703 (2000).

[86] D. M. Joshi, N. K. Rana and V. M. Misra, "Classification of brain cancer using artificial neural network," 2nd International Conference on Electronic Computer Technology, Kuala Lumpur (2010).

[87] B. Djavan, M. Remzi , A. Zlotta , C. Seitz , P. Snow and M. Marberger, "Novel artificial neural network for early detection of prostate cancer," Journal of Clinical Oncology, 20, 921-929 (2002).

[88] Z. H. Zhou, Y. Jiang, Y. B. Yang and S. F. Chen, "Lung cancer cell identification based on artificial neural network ensembles," Artificial Intelligence in Medicine, 24, 25-36 (2002).

[89] I. M. Nasser and S. S. Abu-Naser, "Lung cancer detection using artificial neural network," International Journal of EngiEngineering and Information Systems (IJEAIS), 3, 17-23 (2019).

[90] C. -L. Chi, W. N. Street and W. H. Wolberg, "Application of artificial neural network-based survival analysis on two breast cancer datasets," AMIA Annu Symp Proc., 11, 130–134 (2007).

[91] J. C. E. Floyd Jr., J. Y. Lo, A. J. Yun, D. C. Sullivan and P. J. Kornguth, "Prediction of breast cancer malignancy using an artificial neural network," Cancer, 74, 2944-2948 (1994).

[92] M. Vallières, C. R. Freeman, S. R. Skamene and I. El-Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung

metastases in soft-tissue sarcomas of the extremities," Phys. Med. Biol., 60, 5471–5496 (2015).

[93] M. E. Mayerhoefer, A. Materka, G. Langs, I. Haggstrom, P. Szczypinski, P. Gibbs and G. Cook, "Introduction to radiomics," J Nucl Med, 61, 488–495 (2020).

[94] R. N. Mahon, G. D. Hugo and E. Weiss, "Repeatability of texture features derived from magnetic resonance and computed tomography imaging and use in predictive models for non-small cell lung cancer outcome," Phys. Med. Biol., 64, 145007 (2019).

[95] A. P. Engelbrecht, "Computational intelligence: An introduction," John Wiley & Sons, Limited (2007).

[96] M. A. Arbib, "Brains, machines, and mathematics," Springer-Verlag (1987).

[97] S. Haykin, "Neural networks and learning machines," Prentice Hall (2009).

[98] M. A. Arbib, "The handbook of brain theory and neural networks," MIT press (2003).

[99] E. Akgün and M. Demir, "Modeling course achievements of elementary education teacher candidates with artificial neural networks," International Journal of Assessment Tools in Education, 5, 491–509 (2018).

[100] S. Nissen, "Implementation of a fast artificial neural network library (FANN)," Dept. Comput. Sci., Univ. (2003).

[101] Fann Tool, https://code.google.com/archive/p/fanntool/ (10 January 2021).

[102] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," in IEEE International Conference on Neural Networks, 1, 586-591 (1993).

[103] C. Igel and M. Hüsken, "Improving the Rprop learning algorithm," in Proceedings of the Second ICSC International Symposium on Neural Computation (NC 2000), Canada/Switzerland, 115-121 (2000).

[104] S. E. Fahlman, "Faster-learning variations on back-propagation: An empirical study," Proceedings of the 1988 Connectionist Models Summer School, 38-51 (1989).

[105] D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, 922-928 (2015).

[106] A. Garcia-Garcia, F. Gomez-Donoso, J. Garcia-Rodriguez, S. Orts-Escolano, M. Cazorla and J. Azorin-Lopez, "PointNet: A 3D convolutional neural network for real-time object class recognition," 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, 1578-1584 (2016).

[107] S. Ji, W. Xu, M. Yang and K. Yu, "3D convolutional neural networks for human action recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 35, 221-231 (2013).

[108] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 1-7 (2015).

[109] Y. -H. Byeon and K. -C. Kwak, "Facial expression recognition using 3d convolutional neural network," International Journal of Advanced Computer Science and Applications (IJACSA), 5(12) (2014).

[110] A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks," Proc. Int. Conf. Pattern Recog. Appl. Methods, arxiv:1502.02506, (2015).

[111] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus and A. Biller, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," NeuroImage, 129, 460-469 (2016).

[112] S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, A. Oliver and X. Lladó, "Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach," NeuroImage, 155, 159-168 (2017).

[113] W. Alakwaa, M. Nassef and A. Badr, "Lung cancer detection and classification with 3D convolutional neural network (3D-CNN)," International Journal of Advanced Computer Science and Applications, 8, 409-417 (2017).

[114] T. H. B. Nguyen, E. Park, X. Cui, V. H. Nguyen and H. Kim, "fPADnet: Small and efficient convolutional neural network for presentation attack detection," Sensors, 18, 2532 (2018).

[115] B. B. Ahn, "The compact 3D convolutional neural network for medical images," Stanford University (2017).

[116] A. Algohary, R. Shiradkar, S. Pahwa, A. Purysko, S. Verma, D. Moses, R. Shnier, A.-M. Haynes, W. Delprado, J. Thompson, S. Tirumani, A. Mahran, AR Rastinehad, L. Ponsky, PD Stricker, and A. Madabhushi, "Combination of peri-tumoral and intra-tumoral radiomic features on bi-parametric MRI accurately stratifies prostate cancer risk: A multi-site study," Cancers, 12, 2200 (2020).

[117] C. Zappa and S. A. Mousa, "Non-small cell lung cancer: Current treatment and future advances," Transl. Lung Cancer Res., 5, 288–300 (2016).

[118] Z. Xu, M. Gao, G. Z. Papadakis, B. Luna, S. Jain, D. J. Mollura, and U. Bagci, "Joint solution for PET image segmentation, denoising, and partial volume correction," Med. Image Anal., 46, 229–243 (2018).

[119] Z. Lu, "PubMed and beyond: a survey of web tools for searching biomedical literature," Database, 2011 (2011).

[120] F. F. Khan, A. Saeed, S. Haider, K. Ahmed and A. Ahmed, "Application of medical images for diagnosis of diseases-review article," World J. Microbiol. Biotechnol., 2, 135–138 (2017).

[121] J. Shi, X. Zheng, Y. Li, Q. Zhang and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," IEEE Journal of Biomedical and Health Informatics, 22, 173-183 (2018).

[122] J. Shi, J. Wu, Y. Li, Q. Zhang and S. Ying, "Histopathological image Classification with color pattern random binary hashing-based PCANet and matrix-form classifier," EEE Journal of Biomedical and Health Informatics, 21, 1327-1337 (2017).

[123] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, 1-9 (2015).

[124] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 770-778 (2016).

[125] H. Ravishankar, P. Sudhakar, R. Venkataramani, S. Thiruvenkadam, P. Annangi, N. Babu and V. Vaidya, "Understanding the mechanisms of deep transfer learning for medical images," Deep Learning and Data Labeling for Medical Applications, 10008, Springer, Cham (2016).

[126] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," IEEE Transactions on Medical Imaging, 35, 1285-1298 (2016).

[127] A. Kumar, J. Kim, D. Lyndon, M. Fulham and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," IEEE Journal of Biomedical and Health Informatics, 21, 31-40 (2017).

[128] J. Zhang, Y. Xia, Q. Wu and Y. Xie, "Classification of medical images and illustrations in the biomedical literature using synergic deep learning," Comput. Sci., arXiv:1706.09092 (2017).

[129] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," Proc. IEEE 12th Int. Symp. Biomed. Imag., 294-297 (2015).

[130] B. van Ginneken, A. Setio, C. Jacobs and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York, 286-289 (2015).

[131] F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. T. Scholten, M. Oudkerk, P. A. de Jong, M. Prokop and B. van Ginneken, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," Medical Image Analysis, 26, 195-202 (2015).

[132] B. Menze, M. Reyes and K. Van Leemput, "The multimodal brain tumor image segmentation benchmark (BRATS)," IEEE Trans. Med. Imag., 34, 1993-2024 (2015).

[133] Y. Pan, W. Huang, Z. Lin, W. Zhu, J. Zhou, J. Wong and Z. Ding, "Brain tumor grading based on neural networks and convolutional neural networks," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 699-702 (2015).

[134] W. Shen, M. Zhou, F. Yang, C. Yang and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," IPMI, 588-599 (2015).

[135] G. Carneiro, J. Nascimento and A. P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models," Proc. MICCAI, 652-660 (2015).

[136] J. M. Wolterink, T. Leiner, M. A. Viergever and I. Išgum, "Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks," Proc. MICCAI, 589-596 (2015).

[137] A Gentle Introduction to Transfer Learning for Deep Learning, https://machinelearningmastery.com/transfer-learning-for-deep-learning/ (29 December 2020).

[138] A. Koul, S. Ganju and M. Kasam, "Practical deep learning for cloud, mobile, and edge: Real-world AI & computer-vision projects using python, keras & TensorFlow," O'Reilly Media (2019).

[139] Transfer learning and fine-tuning, https://keras.io/guides/transfer_learning/ (29 December 2020).

[140] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," arXiv:1512.03385 (2015).

[141] G. Huang, Z. Liu, L. van der Maaten and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4700-4708 (2017).

[142] F. Chollet, "Deep learning with python," Manning Publications Co., (2018).

[143] How to configure image data augmentation in keras, https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/ (29 December 2020).

[144] F. Gao, T. Wu, J. Li, B. Zheng, L. Ruan, D. Shang, B. Patel, "SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis," Computerized Medical Imaging and Graphics, 70, 53-62 (2018)

[145] S. P. Singh, L. Wang, S. Gupta, B. Gulyás and P. Padmanabhan, "Shallow 3D CNN for Detecting Acute Brain Hemorrhage from Medical Imaging Sensors," in IEEE Sensors Journal, doi: 10.1109/JSEN.2020.3023471 (2020)

[146] CW. Wang, CP. Yu, "Automated morphological classification of lung cancer subtypes using H&E tissue images," Machine Vision and Applications, 24, 1383–1391 (2013).

[147] X. Zhu, D. Dong, Z. Chen, M. Fang, L. Zhang, J. Song, D. Yu, Y. Zang, Z. Liu, J. Shi and J. Tian, "Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer," Eur Radiol, 28, 2772–2778 (2018).

[148] J. Liu, J. Cui, F. Liu, Y. Yuan, F. Guo, G. Zhang, "Multi-subtype classification model for non-small cell lung cancer based on radiomics: SLS model," Medical Physics, 46, 7, 3091-3100 (2019).

[149] Y. Han, Y. Ma, Z. Wu, F. Zhang, D. Zheng, X. Liu, L. Tao, Z. Liang, Z. Yang, X. Li, J. Huang and X. Guo, "Histologic subtype classification of non-small cell lung cancer using PET/CT images," Eur J Nucl Med Mol Imaging 48, 350–360 (2021).

[150] C. Wang, Y. Long, W. Li, W. Dai, S. Xie, Y. Liu, Y. Zhang, M. Liu, Y. Tian, Q. Li and Y. Duan, "Exploratory study on classification of lung cancer subtypes through a combined K-nearest neighbor classifier in breathomics," Sci Rep 10, 5880 (2020).

[151] F. Yang, W. Chen, H. Wei, X. Zhang, S. Yuan, X. Qiao and YW. Chen, "Machine learning for histologic subtype classification of non-small cell lung cancer: A retrospective multicenter radiomics study," Frontiers in Oncology, 10, 608598 (2021).

[152] Y. Guo, Q. Song, M. Jiang, Y. Guo, P. Xu, Y. Zhang, CC. Fu, Q. Fang, M. Zeng, X. Yao, "Histological Subtypes Classification of Lung Cancers on CT Images Using 3D Deep Learning and Radiomics," Academic Radiology (2020).

[153] Y. Ji, Q. Qiu, J. Fu, K. Cui, X. Chen, L. Xing and X. Sun, "Stage-specific PET radiomic prediction model for the histological subtype classification of non-small-cell lung cancer," Cancer Management and Research, 13, 307–317 (2021).

CURRICULUM VITAE

2004 – 2008	B.Sc., Computer Systems Teaching, Selçuk University, Konya, TURKEY
2011 – 2013	M.Sc., Computer and Information Systems, Gazi University, Ankara, TURKEY
2013 – Present	Instructor, Computer Engineering, Hasan Kalyoncu University, Gaziantep, TURKEY
2015 – 2021	Ph.D., Electrical and Computer Engineering, Abdullah Gül University, Kayseri, TURKEY
2019	Entrepreneurship Support, TÜBİTAK, Ankara, TURKEY
2020 – Present	Founder, Indevo Technology LTD., Gaziantep, TURKEY
2020	Start-Up Acceleration Support, GETHAM-GIZ, Gaziantep, TURKEY

SELECTED PUBLICATIONS AND PRESENTATIONS

J1) M. Bicakci, O. Ayyildiz, Z. Aydin, A. Basturk, S. Karacavus and B. Yilmaz, Metabolic Imaging Based Sub-Classification of Lung Cancer published in IEEE Access (Dec. 2020).