

REPUBLIC OF TURKEY  
YILDIZ TECHNICAL UNIVERSITY  
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

USING GENE EXPRESSION PROFILES OF CANCER  
PATIENTS WITH IMAGE-BASED DEEP LEARNING  
APPROACH TO DEVELOP METHODS FOR  
CLASSIFICATION AND PREDICTION OF CANCER WHILE  
REVEALING CRITICAL GENES

**Büşra Nur DARENDELİ**

MASTER OF SCIENCE THESIS  
Department of Bioengineering  
Program of Bioengineering

Supervisor  
Assist. Prof. Dr. Alper YILMAZ

August, 2021

**REPUBLIC OF TURKEY**  
**YILDIZ TECHNICAL UNIVERSITY**  
**GRADUATE SCHOOL OF SCIENCE AND ENGINEERING**

**USING GENE EXPRESSION PROFILES OF CANCER PATIENTS WITH  
IMAGE-BASED DEEP LEARNING APPROACH TO DEVELOP  
METHODS FOR CLASSIFICATION AND PREDICTION OF CANCER  
WHILE REVEALING CRITICAL GENES**

A thesis submitted by Büşra Nur DARENDELİ in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE** is approved by the committee on 13.08.21 in Department of Bioengineering, Program of Bioengineering.

Assist. Prof. Dr. Alper YILMAZ  
Yıldız Technical University  
Supervisor

**Approved By the Examining Committee**

Assist. Prof. Dr. Alper YILMAZ, Supervisor  
Yıldız Technical University

\_\_\_\_\_

Assoc. Prof. Dr. Murat TOPUZOĞULLARI, Member  
Yıldız Technical University

\_\_\_\_\_

Assist. Prof. Dr. Enes Seyfullah KOTİL, Member  
Bahçeşehir University

\_\_\_\_\_

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of Using Gene Expression Profiles of Cancer Patients with Image-Based Deep Learning Approach to Develop Methods for Classification and Prediction of Cancer while Revealing Critical Genes supervised by my supervisor, Assist. Prof. Dr. Alper YILMAZ. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Büşra Nur DARENDELİ

Signature

*Dedicated to my lovely parents  
and my fiancée*



## ACKNOWLEDGEMENTS

---

I would like to thank everyone who made this thesis possible and supported me endlessly.

First of all, I would like to thank my dear advisor, Assist. Prof. Alper Yılmaz, for giving me a direction throughout my undergraduate and graduate life, guiding the insight and knowledge I gained in my studies, and not leaving me alone during my study despite our long working hours. Specially, I would like to apologize to Ömeralp Yılmaz, who could not spend time with his father due to our long working hours, and thank him for his sacrifice and understanding.

I thank to my project supervisor Assist.Prof. Enes Seyfullah Kotil. I cannot express in words how important it is for me to expand my academic perspective and gain experience in the field of laboratory and bioinformatics with the numerous opportunities it provides. I apologize to him for delaying my task in the project for my thesis work. In this process, I am grateful to him for guiding me through all my difficult times.

I would like to thank Kotil Laboratory members; Fatma Zehra Sarı, Ramin Nasheebi, Hüseyin Tunç, Dilan Can and Ghizlane, whom we worked together in project.

I would like to thank Alper Yılmaz Research Group members; Selcen Arı-Yuka, Nilay Yönet, Muhammet Çelik, Elif İrem Keleş and special thanks to Gülce Çelen and Oğuzhan Kalyon for their deep conversations and guidance that put me at ease every time I got stuck.

I would like to thank my precious family who have always been with me throughout this process. I thank my mother, Asiye, for comforting and guiding me whenever I felt stuck or unable to do it, for always supporting me with her prayers. I thank my father, Cengiz, for making me feel safe all my life and giving me the strength to cope with any situation. I would like to thank my sister, Muhlise Kübra, for being with me during all the stressful moments I experienced during my thesis process, for tolerating me and supporting me when I needed it. I thank my mother-in-law, Hülya Kiraz, and father-in-law, Fatih Kiraz, for their support, patience and understanding.

Last but not least, special thank my very precious fiancée, Furkan Kiraz. I cannot express the importance of the countless sacrifices you have made to focus on my thesis during this painful process. Thank you so much for always providing me with a peaceful, safe and inspiring environment. Many things would not have been possible without your support.

Büşra Nur DARENDELİ



# TABLE OF CONTENTS

---

<b>LIST OF ABBREVIATIONS</b>	<b>viii</b>
<b>LIST OF FIGURES</b>	<b>x</b>
<b>LIST OF TABLES</b>	<b>xii</b>
<b>ABSTRACT</b>	<b>xiii</b>
<b>ÖZET</b>	<b>xv</b>
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 Literature Review . . . . .	1
1.2 Objective of the Thesis . . . . .	2
1.3 Hypothesis . . . . .	2
<b>2 CANCER</b>	<b>3</b>
2.1 Cancer and its mechanisms . . . . .	3
2.2 Diagnosis Techniques . . . . .	7
2.2.1 Pathology . . . . .	7
2.2.2 Radiology . . . . .	10
2.2.3 Molecular methods . . . . .	11
2.3 Gene Expression Based Cancer Diagnosis . . . . .	15
2.3.1 Methods for Gene Expression Quantification . . . . .	20
2.3.2 Meta-analysis . . . . .	23
<b>3 DEEP LEARNING AND CANCER</b>	<b>27</b>
3.1 History and Basics . . . . .	27
3.2 Deep Neural Networks . . . . .	30
3.2.1 Multi Layer Perceptron (MLP) . . . . .	32
3.2.2 Convolutional Neural Network (CNN) . . . . .	35
3.2.3 Recurrent Neural Networks (RNN) . . . . .	37
3.2.4 Vulnerability of Deep Learning Techniques: One-Pixel Attack . .	37
3.3 Deep Learning Approach for Cancer Diagnosis . . . . .	38

<b>4</b>	<b>METHODS</b>	<b>41</b>
4.1	Datasets . . . . .	41
4.2	Selection of genes with highest DEG cases . . . . .	41
4.3	Conversion of gene expression data into image . . . . .	41
4.4	Model Architecture used for Deep Learning Training . . . . .	43
4.5	One pixel attack . . . . .	43
<b>5</b>	<b>RESULTS</b>	<b>45</b>
5.1	Gene expression data as image . . . . .	45
5.2	Model Training . . . . .	45
5.3	Performance Measurement . . . . .	46
5.4	One-pixel Attack Results . . . . .	48
5.5	Annotation of Attack Results Based on Real Cancer Patient Expression Levels . . . . .	52
<b>6</b>	<b>RESULTS AND DISCUSSION</b>	<b>57</b>
	<b>REFERENCES</b>	<b>59</b>
<b>A</b>	<b>ONE-PIXEL ATTACK RESULTS</b>	<b>77</b>
<b>B</b>	<b>STATISTICAL DATA OF GENE EXPRESSION VALUES OF AGRIN, CATALASE AND KINESIN-LIKE PROTEIN GENES ACCORDING TO CANCER TYPES</b>	<b>88</b>
	<b>PUBLICATIONS FROM THESIS</b>	<b>92</b>

## LIST OF ABBREVIATIONS

---

AS-PCR	Allele-Specific PCR
AI	Artificial Intelligence
CAT	Catalase
CEAC	Competitive Allele Specific Controller
cAS-PCR	Competitive External Allele-Specific PCR
cDNA	Complementary DNA
CT	Computer Tomography
CNN	Convolutional Neural Network
dNTP	Deoxyribonucleotide
ddNTP	Dideoxyribonucleotides
DEGs	Differentially Expressed Genes
FISH	Fluorescence In Situ Hybridization
GTE <sub>x</sub>	Genotype-Tissue Expression
H&E	Hematoxylin and Eosin
IHC	Immunohistochemistry
ISFET	Ion Sensitive Field Effect Transistor
KIF1C	Kinesin-like Protein
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
MRI	Magnetic Resonance Imaging
CMOS	Metal Oxide Semiconductor
MLP	Multi Layer Perceptron
NGS	Next Generation Sequencing

PCR	Polymerase Chain Reaction
PET	Positron Emission Tomography
PPi	Pyrophosphate
ROC	Receiver Operating Characteristic
RNN	Recurrent Neural Network
RGB	Red Green Blue
ReLU	Restricted Linear Unit
RB	Retinoblastoma
RNA-Seq	RNA Sequencing
SNP	Single Nucleotide Polymorphism
SNVs	Single Nucleotide Variants
SVM	Support Vector Machine
t-SNE	t-Distributed Stochastic Neighbor Embedding
TCGA	The Cancer Genome Atlas
tAS-PCR	Traditional Allele-Specific PCR
UCSC	University of California Santa Cruz
WHO	World Health Organization

## LIST OF FIGURES

---

<b>Figure 2.1</b>	Inflammation vs Cancer . . . . .	4
<b>Figure 2.2</b>	Hallmarks of Cancer . . . . .	5
<b>Figure 2.3</b>	Evolution of Tumor Cell . . . . .	6
<b>Figure 2.4</b>	Clonal Heterogeneity . . . . .	7
<b>Figure 2.5</b>	Workflow of histopathology . . . . .	8
<b>Figure 2.6</b>	Lung cancer prediction results using the CNN algorithm . . . . .	9
<b>Figure 2.7</b>	Demonstration of deep learning-based Gleason classification to the clinician . . . . .	9
<b>Figure 2.8</b>	The process of mutations that occur in the division phase of a normal cell to evolve a cell into a cancer cell . . . . .	11
<b>Figure 2.9</b>	Working principle of Polymerase Chain Reaction . . . . .	12
<b>Figure 2.10</b>	Working principle of Real Time PCR and the probes used . . . . .	13
<b>Figure 2.11</b>	The working principles of tAS-PCR, cAS-PCR and rcAS-PCR . . . . .	14
<b>Figure 2.12</b>	The stages of the Sanger Sequencing method . . . . .	16
<b>Figure 2.13</b>	Overview of short-read sequencing methods . . . . .	17
<b>Figure 2.14</b>	Overview of long-read sequencing techniques . . . . .	19
<b>Figure 2.15</b>	Overview of Microarray . . . . .	20
<b>Figure 2.16</b>	Common RNA-seq experiment . . . . .	22
<b>Figure 2.17</b>	Network of genes involved in colorectal cancer formation . . . . .	25
<b>Figure 3.1</b>	Timeline of AI . . . . .	27
<b>Figure 3.2</b>	MARK I Perceptron . . . . .	28
<b>Figure 3.3</b>	The relation between AI and deep learning . . . . .	30
<b>Figure 3.4</b>	Possibility of Complex Distribution . . . . .	31
<b>Figure 3.5</b>	Relation between Neuron and Perceptron . . . . .	32
<b>Figure 3.6</b>	Architecture of MLP . . . . .	33
<b>Figure 3.7</b>	Sigmoid Neuron . . . . .	33
<b>Figure 3.8</b>	Tanh Neuron . . . . .	34
<b>Figure 3.9</b>	ReLU . . . . .	34
<b>Figure 3.10</b>	CNN Architecture . . . . .	35
<b>Figure 3.11</b>	Schematize of CNN . . . . .	36
<b>Figure 3.12</b>	RNN Architecture . . . . .	37

<b>Figure 3.13</b>	Estimation of cancer stage and molecular properties from the H&E image of breast cancer . . . . .	39
<b>Figure 4.1</b>	Illustration of mapping gene expression values to RGB space . . . .	42
<b>Figure 4.2</b>	Illustration of Numpy 4D arrays . . . . .	42
<b>Figure 4.3</b>	CNN Architecture . . . . .	43
<b>Figure 4.4</b>	Illustration of Differential Evolution algorithm over Ackley function	44
<b>Figure 5.1</b>	Color image representation of samples for 1024 genes . . . . .	45
<b>Figure 5.2</b>	Accuracy and Loss plots of the trained model after 40 epochs. . . .	46
<b>Figure 5.3</b>	The ROC curve . . . . .	46
<b>Figure 5.4</b>	Sample images obtained as a result of the attack . . . . .	51
<b>Figure 5.5</b>	Gene expression values of Agrin gene according to cancer types . .	54
<b>Figure 5.6</b>	Gene expression values of Catalase gene according to cancer types	55
<b>Figure 5.7</b>	Gene expression values of Kinesin-like protein gene according to cancer types . . . . .	56

## LIST OF TABLES

---

<b>Table 4.1</b>	Model architecture . . . . .	43
<b>Table 5.1</b>	Confusion Matrix . . . . .	47
<b>Table 5.2</b>	Comparison model with other studies . . . . .	47
<b>Table 5.3</b>	Identified genes resulted from one-pixel attack . . . . .	48
<b>Table 5.4</b>	Cancers and stage information whose labels are changed by decreasing expression values . . . . .	49
<b>Table 5.5</b>	Cancer type and stage distribution of samples that turned into cancer as a result of an attack when they were normal . . . . .	50
<b>Table A.1</b>	One-pixel attack results applied to each sample . . . . .	78
<b>Table B.1</b>	canSAR Black expression analysis of AGRN gene . . . . .	89
<b>Table B.2</b>	canSAR Black expression analysis of CAT gene . . . . .	90
<b>Table B.3</b>	canSAR Black expression analysis of KIF1C gene . . . . .	91

# Using Gene Expression Profiles of Cancer Patients with Image-Based Deep Learning Approach to Develop Methods for Classification and Prediction of Cancer while Revealing Critical Genes

Büşra Nur DARENDELİ

Department of Bioengineering  
Master of Science Thesis

Supervisor: Assist. Prof. Dr. Alper YILMAZ

Cancer is one of the malignant diseases worldwide. Difficulties in diagnosis and treatment cannot prevent the progression of the disease and cause the death of millions of people. The intra-tumor and inter-tumor heterogeneity characteristic of tumor cells has resulted in cancer being a disease with individual characteristics. Since each individual has a unique tumor and tumor microenvironment, general screening methods make early detection of the disease difficult. Here, we aimed to provide new perspective of cancer diagnosis using deep learning approach on gene expression data. The training of gene expression data, in which the exact results of the changes in the genome are seen, was carried out using the deep learning method. In addition, it is aimed to identify critical genes that are effective in identifying tumor and normal tissues, which deep learning has determined with high accuracy.

In this study, The Cancer Genome Atlas (TCGA) dataset with RNA-Seq data of approximately 30 different types of cancer patients and GTEx RNA-seq data of normal tissues were used. The input data for the training was transformed to RGB format and the training was carried out with a Convolutional Neural Network (CNN). The trained algorithm is able to predict cancer with 97.7% accuracy, based on gene expression data. Moreover, we applied one-pixel attack on the trained model to determine effective genes for prediction of the disease. As a result of the application of this method, 13 critical genes that are effective on the prediction of the deep learning model were determined.

As a result, with the developed deep learning model, a model that can distinguish tumor and normal tissues based on gene expression data has been developed. By examining the prediction mechanism of this model, genes that are candidates to be biomarkers for cancer were determined. When the identified genes were searched in the literature, their relationship with cancer was observed. These genes, which were determined as a result of the study, can be used as a biomarker for cancer by supporting experimental data. In line with the results obtained, it is shown that individual cancers can be examined on the basis of genes, and that individual diagnoses and treatments can also be applied.

**Keywords:**

cancer, TCGA, CNN, biomarker



# Kanser Hastalarının Gen İfade Verileri Kullanılarak Kanserde Kritik Genlerin Tanımlanması, Kanser Sınıflandırılması ve Tahmini için Görüntü-Tabanlı Derin Öğrenme Yaklaşımı

Büşra Nur DARENDELİ

Biyomühendislik Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Dr. Öğr. Üyesi Alper YILMAZ

Kanser, dünya genelinde ölümcül hastalıkların başında yer almaktadır. Teşhis ve tedavi aşamasındaki zorluklar hastalığın ilerleyişini engelleyememekte milyonlarca insanın ölümüne yol açmaktadır. Tümörlü hücrelerin sahip olduğu intra-tümör ve inter-tümör heterojenite özelliği kanserin bireylere özgü özelliklere sahip bir hastalık olması sonucunu doğurmuştur. Her bireyin sahip olduğu tümör ve tümör mikroçevresi kendisine özgü olması sebebi ile genel tarama yöntemleri hastalığın erken tespitini zorlaştırmaktadır. Burada gen ekspresyon verileri üzerinde derin öğrenme yaklaşımı kullanılarak kanser teşhisine yeni bir bakış açısı kazandırmayı amaçladık. Genomdaki değişikliklerin birebir sonuçlarının görüldüğü gen ifade verilerinin derin öğrenme yöntemi kullanılarak eğitimi gerçekleştirilmiştir. Ek olarak, derin öğrenmenin yüksek doğrulukla belirlediği tümör ve normal dokuları belirlemede etkili olan kritik genlerin belirlenmesi amaçlanmıştır.

Bu çalışmada, yaklaşık 30 farklı kanser hastasının RNA-Seq verileri ile Kanser Genom Atlas (TCGA) veri seti ve normal dokuların GTEx RNA-seq verileri kullanılmıştır. Eğitim için girdi verileri RGB formatına dönüştürülmüş ve eğitim bir Evrişimsel Sinir Ağı (CNN) ile gerçekleştirilmiştir. Eğitilmiş algoritma, gen ekspresyonu verilerine dayanarak kanseri %97.7 doğrulukla tahmin edebilir. Ayrıca, hastalığın tahmini için etkili genleri belirlemek için eğitilmiş modele bir-piksel saldırı uyguladık. Bu yöntemin uygulanması sonucunda derin öğrenme modelinin tahmini üzerinde etkili olan 13

kritik gen belirlenmiştir.

Sonuç olarak, geliştirilen derin öğrenme modeli ile tümör ve normal dokuları gen ifade verilerine bağlı olarak ayırt edebilen bir model geliştirilmiştir. Bu modelin tahmin mekanizması incelenerek kanser için biyobelirteç olmaya aday olan genler belirlenmiştir. Belirlenen genlerin literatür araştırılması yapıldığında kanserle ilişkileri görülmüştür. Gerçekleştirilen çalışma sonucu belirlenen bu genler deneysel verilerle desteklenerek kanser için biyobelirteç olarak kullanılabilir. Elde edilen sonuçlar doğrultusunda kişilere özgü kanserler gen bazında incelenerek, kişiye özgü tanı ve tedavilerin de uygulanabileceği gösterilmektedir.

**Anahtar Kelimeler:** kanser, TCGA, CNN, biyobelirteç



### 1.1 Literature Review

The deep learning approach has emerged by designing computer models that can perform the learning process as a result of interconnected layers based on the human brain, such as neurons. As a result of the development of data science and especially the rapid increase in biological data in the last decade, the designed neural networks have begun to play important roles in the interpretation of biological data for the diagnosis and treatment of diseases [1]. Cancer, which is one of the biggest health problems in the world, is one of the diseases in which deep learning approaches are widely applied.

Since cancer is a disease with high genomic heterogeneity and phenotypic plasticity, its diagnosis and treatment involve various difficulties [2]. By virtue of the developing technology, many medical data of cancer patients are available. As a result of the processing of these medical data with deep learning approaches, the stages of diagnosis and treatment have improved.

Deep learning approaches used for cancer diagnosis and treatment can be examined under two headings according to the inputs. These are generally image-processing methods using image-based data such as CT, histopathology, radiology and gene expression signature-based approaches using gene expression data as input. In both approaches, the aim is to produce solutions that will facilitate the diagnosis and treatment of cancer.

The gene expression signature-based approach has created a perspective that provides more opportunities for personalized approaches in cancer diagnosis and treatment. The use of data obtained as a result of next generation sequencing, which is one of the developing molecular biology techniques, through machine learning algorithms strengthens this approach. Estimation of disease progression and survival time of cancer patients using gene expression data, generating solutions to various

cancer-specific classification problems (such as molecular cancer classification [3], molecular subtype classification [4], specific cancer classification [5]) Solutions to such problems were produced with a deep learning approach.

## **1.2 Objective of the Thesis**

The increase in deaths due to cancer shows that studies that will improve the diagnosis and treatment of cancer are important. Although there is a lot of unknown for the aforementioned disease, data that can provide clues about the critical points of the disease are also available thanks to the developing experimental methods. There is also a developing artificial intelligence technology for the processing and interpretation of this data.

In this study, by combining these possibilities, it is aimed to enable the machine to distinguish between tumor and normal tissue, and then to find genes that may be effective in cancer by examining how it makes this distinction. In this way, it is aimed to perform the cancer biomarker determination process in a computerized environment and to shorten the time-consuming and expensive experiment process.

## **1.3 Hypothesis**

In this study, an algorithm that can separate tumor and normal tissues will be implemented using publicly available datasets The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) [6]. The TCGA includes RNA-seq data from 10,477 tumor and normal tissue, while the GTEx dataset includes RNA-seq data from 7429 healthy tissues. These two datasets will be converted into an image format without any normalization in order not to lose the information they have in the gene expression data, and will be trained with the Convolutional Neural Network (CNN) model, which is created because it is thought that it can learn more information from the input with feature extraction capability. If high success is achieved after the training, it is aimed to determine the critical genes by applying the one-pixel attack method to the images created using gene expression data.

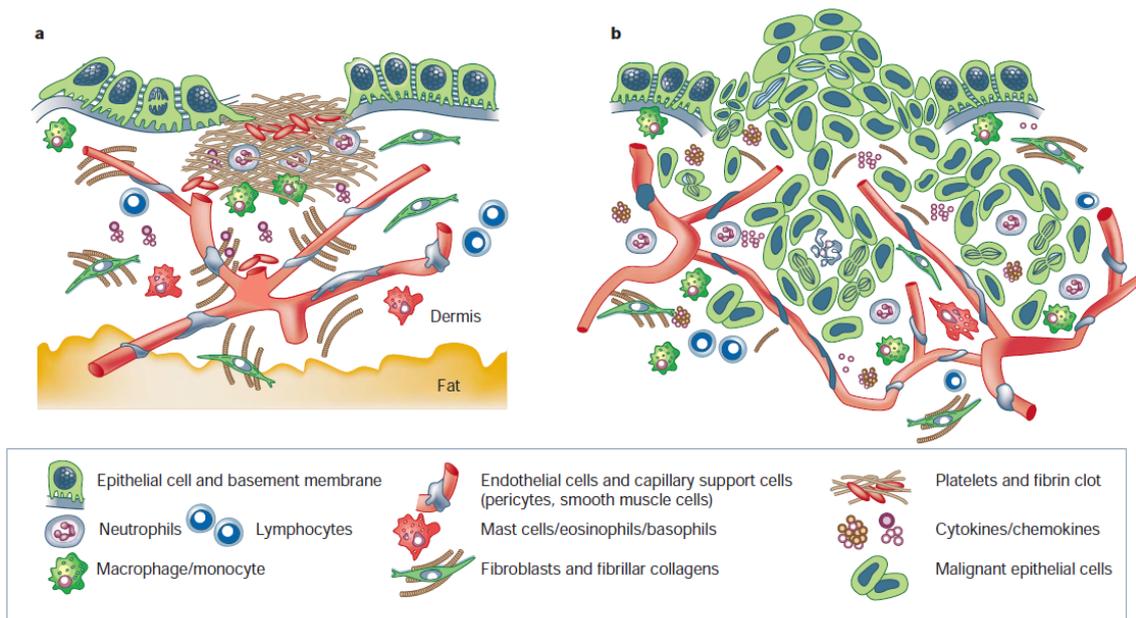
Cancer ranks second in the list of diseases that cause death in the world. According to the World Health Organization (WHO) data that 9.6 million people have died from cancer worldwide since 2018 have been shared. This disease, which causes a high incidence, occurs as a result of changes in normal cells. It spreads as a result of uncontrolled division of normal cells and invasion of other cells. In this section, cancer mechanism and factors affecting this mechanism will be mentioned. Next, cancer diagnosis techniques and gene expression based cancer diagnosis techniques will be discussed.

### 2.1 Cancer and its mechanisms

Cancer is the name given to the types of diseases in which cells display the same behavior in different tissues. No matter what tissue the cells are in, they tend to divide uncontrolled and spread to the surrounding tissues, causing tumor development. In a normal body cycle, cells have the characteristics of division. In fact, there is a mechanism in which normal body cells are constantly dividing like tumor cells, which is known as inflammation. In the case of inflammation, chemotactic factors in cells organized in the epithelial tissue activate and stimulate the signal transduction system required for the regeneration of the extracellular matrix. With the end of the inflammation, the signal transmission decreases and the normal cycle is restored (Figure 2.1.a). Tumor formation also follows a similar sequence to inflammation, but as a result of genomic instability in cells during division, the signal transmission is lost and the cells continue to divide and spread to different tissues (Figure 2.1.b) [7].

There are many biological mechanisms behind the proliferation mechanism of tumor cells. These mechanisms that distinguish between normal cells and tumor cells can be listed as follows:

- Continuity of proliferative signals; Tumor cells ensure the continuity of the



**Figure 2.1** Inflammation vs Cancer [7]

signal production that occurs in normal tissues and supports cell growth, thus ensuring the continuity of division processes.

- Avoiding growth suppressing signals; There are proteins such as RB (associated with retinoblastoma) and TP53, which are naturally present in the body mechanism and enable to suppress the proliferation of tumor cells. However, tumor cells can eliminate factors that limit division by bypassing these mechanisms.
- Avoiding the immune system; Although not in all types of cancer, it has been observed that tumor cells have systems to escape from the immune response in viral cancer types. The immune system seems to play a role in tumor formation and progression in some non-viral cancer types.
- Immortality due to replication; Telomere sequences found in normal tissue cell DNA shorten as a result of replication, leading to aging of the cell and death after a stage. This is the opposite in tumor cells. They tend to maintain telomere length to avoid aging and apoptotic effects. Thus, they can maintain their unlimited reproductive abilities.
- Tumor development caused by inflammation; Although there are studies showing that inflammation responses are tumor suppressing in the first studies, it has been observed that the effects resulting from inflammation in the later stages help tumor cells to gain distinctive features. [8–11]
- Invasion and metastasis; It has been observed that tumor cells can gain the

ability to resist apoptosis and spread to surrounding tissues through the epithelial-mesenchymal transition program. [12–16]

- Stimulate angiogenesis; Like normal tissue cells, tumor cells also show vascularization due to their need for nutrients and oxygen and their need for waste disposal.
- Genome instability & mutation; All the mentioned features arise as a result of mutations in the genome of tumor cells. Apart from this, some features can also occur with epigenetic changes.
- Cell death resistance; apoptosis is the most common mechanism for preventing cancer development. Tumor cells have developed various strategies to limit and prevent apoptosis.
- Disorganization of cellular energy ; Uncontrolled proliferation of tumor cells also affects the amount of glucose needed. It has been observed that even under aerobic conditions, tumor cells limit their energy production by glucose metabolism and reprogram their energy metabolism. [17–19]

The 10 features mentioned above are the distinctive features of tumor cells (Figure 2.2) [20] In the light of all this information, the general mechanism can be determined by the contribution of the tumor microenvironment to tumorigenesis as well as transferring all tumor characteristics one by one.

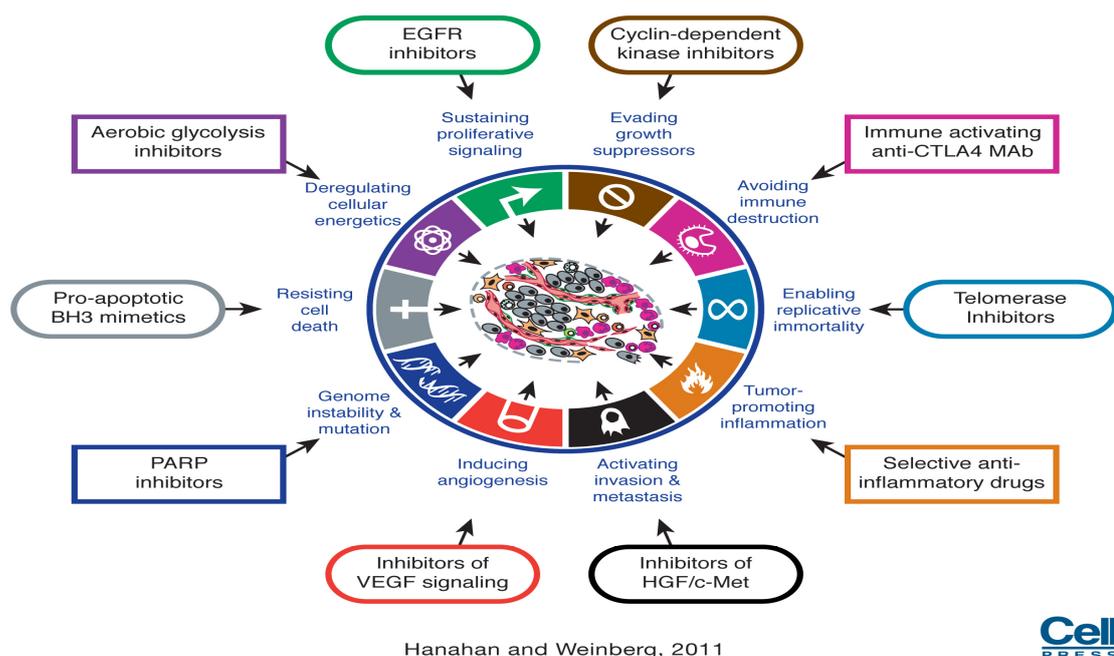
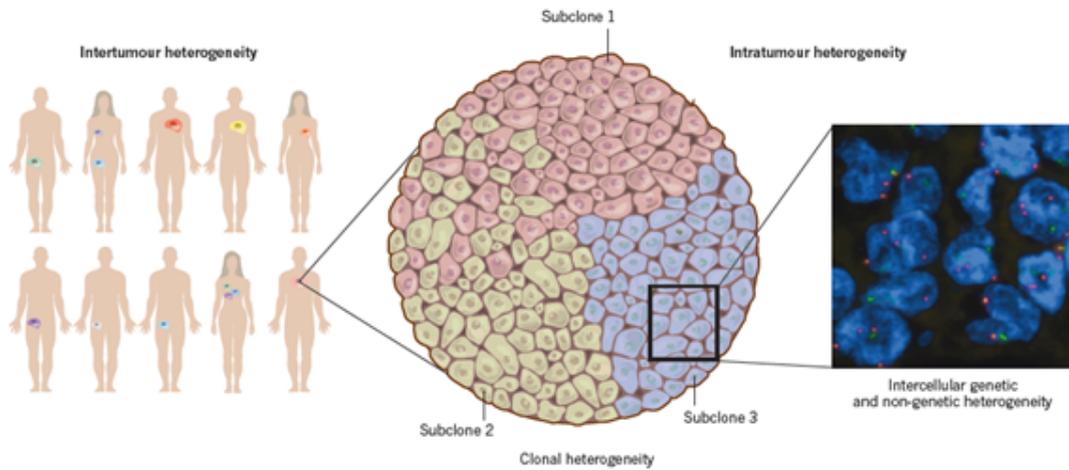


Figure 2.2 Hallmarks of Cancer [20]



determined. The complex clonal heterogeneity system and the evolution of the tumor in heterogeneity were determined by sequencing technology (Figure 2.4) [25–29].



**Figure 2.4** Clonal Heterogeneity [29]

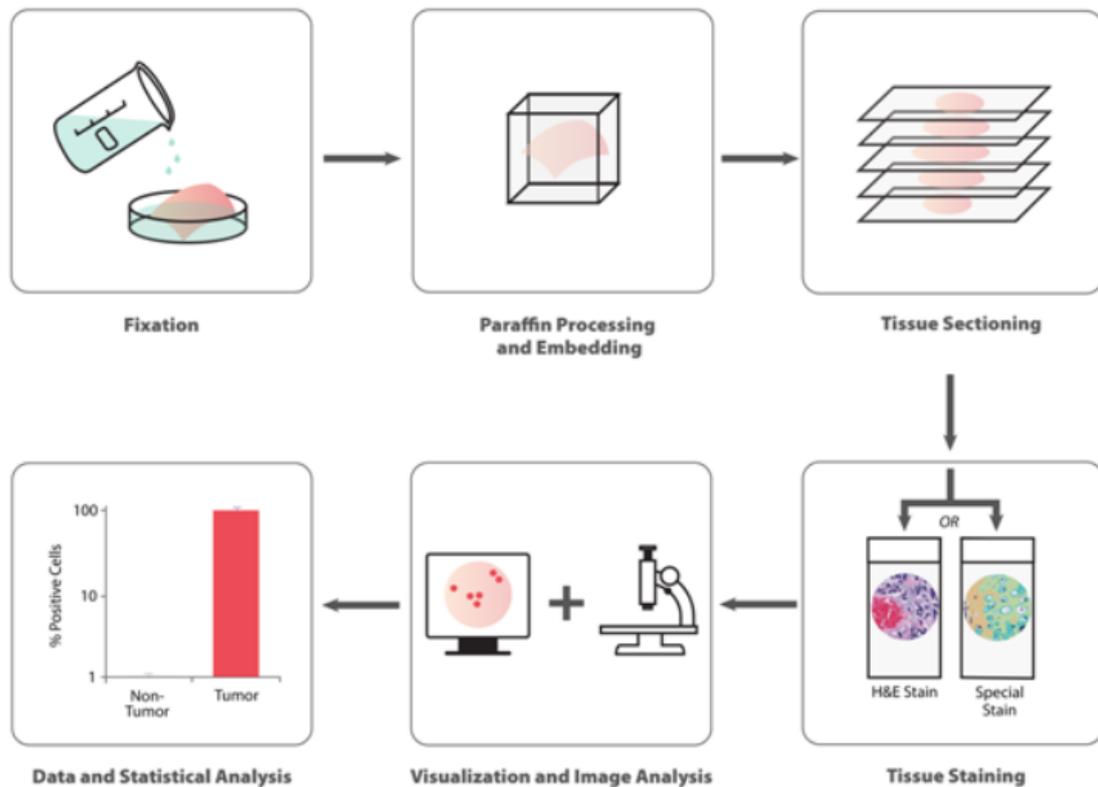
## 2.2 Diagnosis Techniques

Countless methods have been developed for cancer diagnosis. These can be grouped under 3 titles in general; including laboratory tests, imaging tests and biopsy. Laboratory tests are generally based on tests performed on samples taken from tissue or blood. Mutation-based techniques based on tumor markers will be discussed in the molecular methods section. Techniques based on gene expression are included in section 2.3. Imaging tests; It includes CT Scan, MRI, Bone Scan, PET Scan, Ultrasound and X-ray methods. In the radiology section, diagnoses made using these methods are explained in detail. Biopsy, on the other hand, is a method that provides diagnosis as a result of applying various procedures to tissues taken from areas containing tumor suspicion. Biopsy procedure can be summarized under 3 headings; by needle, by operation, by endoscopy. It is classified according to the process of tissue removal from the body. In this section, pathology, radiology and molecular methods used for cancer diagnosis will be emphasized. In addition, it will be mentioned about the studies where these methods facilitate the diagnosis phase through machine learning methods using the technical data they have.

### 2.2.1 Pathology

Pathology techniques that enable the diagnosis to be made according to cell morphology under microscopy by staining after the surgical removal of tissue from the body. As mentioned before, many methods are available for cancer diagnosis

from tissue. The most common of these and the most used method in the diagnosis of the data obtained after the application process is the staining of the tissues with hematoxylin & eosin dye, which is a histopathological method. After the process of fixing the tissue section to the plate, the dye application allows the nuclei, cytoplasm and extracellular matrix in the tissue to be painted in different colors. By means of this procedure, the diagnosis process is performed according to the changes in cell morphologies [30] (Figure 2.5)



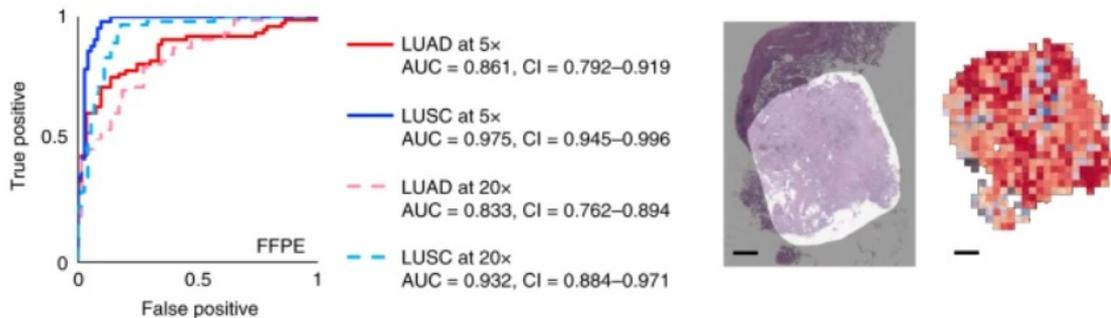
**Figure 2.5** Workflow of histopathology

With the development of the method specified in the later years, the term digital pathology has emerged. Digital pathology is a sub-branch of pathology that enables the processing of data in computer environment by digitizing the stained tissue slides. Even if the process was facilitated by digital pathology, the time-consuming burden of the diagnostic phase continued. Developing artificial intelligence technologies enabled the processing and automatization of the acquired image data. Many pathology data using a learning algorithm after image processing have achieved results close to clinicians in cancer diagnosis [31–34].

In this area, histopathology images have been widely used in tissue classification [35–38], mitosis detection [39–41], grading detection [42, 43], segmentation [44–46].

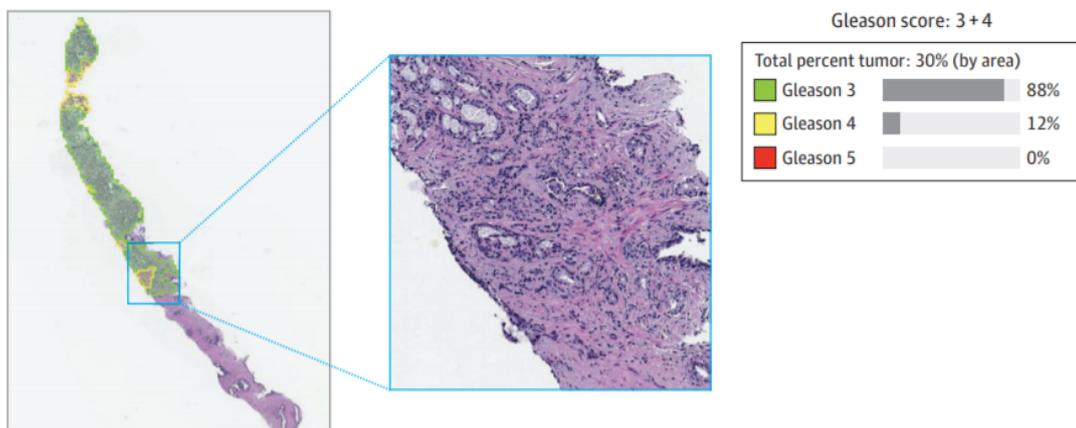
In the study by Coudray et al [38], slide images of lung cancer tissues were trained

using the CNN algorithm. A training dataset containing tissue images of two different types of lung cancer (LUAD, LUSC) was used. The accuracy achieved after training is 0.97. It has been shown that the algorithm gives results as nearly as accurate as a pathologist (Figure 2.6).



**Figure 2.6** Lung cancer prediction results using the CNN algorithm [38]

The problem of determining the grade groups present in prostate cancer has become automatized with the help of deep learning algorithms. In prostate cancer, evaluation of images according to cell differences and classification according to Gleason score generally results in subjective results and the result may differ from expert to expert. In order to standardize this situation, a deep learning algorithm was developed by Nagpal et al [43] to determine biopsy samples according to Gleason score. According to the results obtained from the algorithm, while the rate of experts in the Gleason score was 94.7%, the results of the algorithm were almost the same and gave an accuracy of 94.3%. There is a difficulty in Gleason scaling due to the fact that the tumor tissue does not have a single grade, and the complexity created by this difficulty has been tried to be eliminated by the algorithm. The grade detection accuracy of the algorithm has been shown to be 94.3% (Figure 2.7).



**Figure 2.7** Demonstration of deep learning-based Gleason classification to the clinician [43]

### 2.2.2 Radiology

Another widely used method for cancer diagnosis is radiology. Radiology is mainly an area where medical imaging techniques are used. These imaging techniques; computed tomography (CT) scanning, X-ray, ultrasound, mammogram, magnetic resonance imaging (MRI). Imaging is performed by choosing one of the appropriate medical imaging methods for patients with suspected cancer. The clinician diagnoses the disease through the image obtained. The working mechanisms of the mentioned techniques differ among themselves. X ray; It provides image formation by bombarding the tungsten target of electron beams in an x-ray tube [47]. CT; By using a 360 degree rotation angle, it obtains a cross-sectional view of the body with x-ray transmission measurements. Mathematically transfers the obtained x-ray images to the computer environment [48]. Ultrasound; It enables the image to be taken by converting electrical energy into high frequency sound energy that will pass through the patient tissue [49]. MR; It enables the production of images by using magnetic fields and radio waves [50].

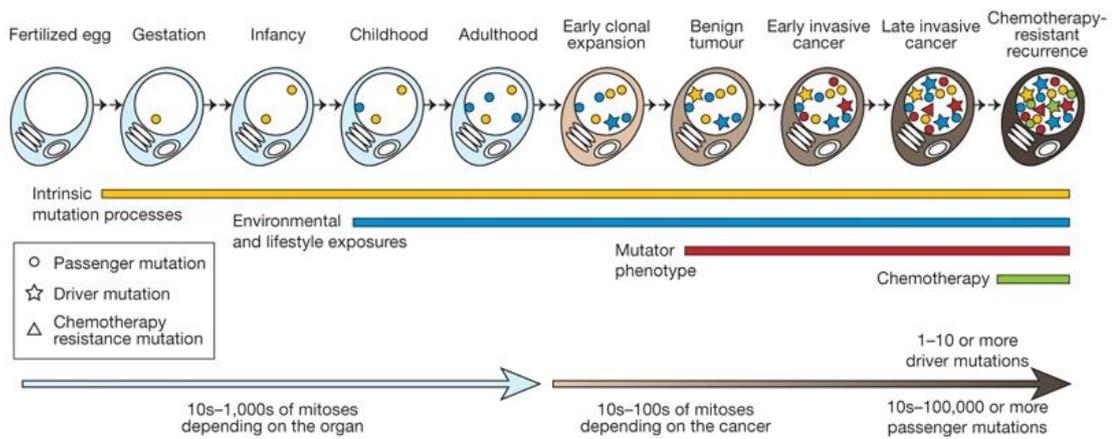
All imaging techniques used generate data for clinicians. Diagnoses about diseases are made by making sense of the data produced. The biggest problem encountered in interpreting radiology images is that there are not enough professionals to make sense of the data produced [51]. Especially, these increasing data put a lot of workload on radiologists. Due to the workload they have, they cause overlooked situations during the diagnosis stage [52, 53]. Overlooked situations or errors in diagnosis lead to fatal consequences for patients [53].

Deep learning approaches have been used in order to overcome these situations that cause fatal consequences [54]. In the field of radiology, deep learning algorithms are used in cancer diagnosis [55], image processing [56] and data mining [57].

Many competitions have been organized regarding the active use of deep learning applications in the field of radiology [58, 59]. Apart from these competitions, many radiology data are shared as open source data for the application of deep learning algorithms [60]. Using the Wisconsin Breast Cancer dataset, a deep learning application was carried out with 98% success in classifying breast cancer [61]. The aforementioned studies have proven that deep learning methods are an application that helps to eliminate many difficulties encountered in the field of radiology. It has been observed that many processes that create a workload for doctors can be mitigated by using deep learning methods.

### 2.2.3 Molecular methods

Genome alterations are the major underlying cause of cancer occurrence. These changes in the genome cause an increase or decrease in the expression of proteins necessary for the cell [62]. In this case, the cell loses many of its previous features. It begins to give different responses, such as uncontrolled proliferation, an increase in angiogenesis [20]. These mutations in the genetic material have an important power in cancer progression. Mutations, which lead to the occurrence of cancer, emerge during human life occurring normal cell division and become capable of causing disease by changing over time Figure 2.8 [63].



**Figure 2.8** The process of mutations that occur in the division phase of a normal cell to evolve a cell into a cancer cell. [63]

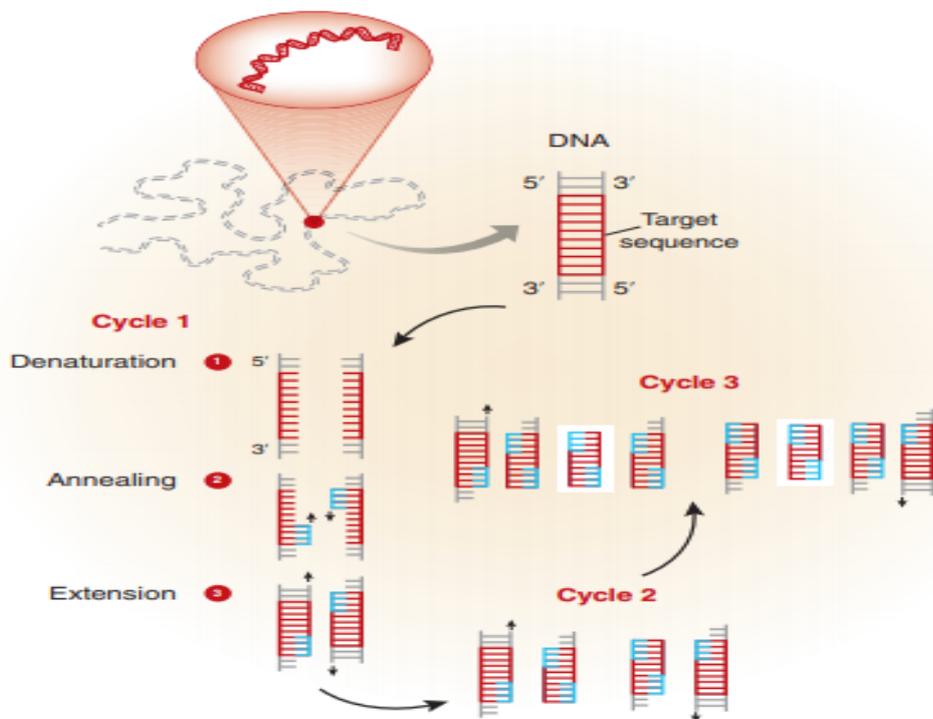
Existence mutations cause tumor formation in the cell over time. Cancers caused by certain mutations (such as KRAS, EGFR, BRCA 1-2) are known. By identifying these mutations, it is possible to determine the cancer the patient has. Generally, there are certain types of mutations that occur in the cancerous cell. These; Single Nucleotide Variants(SNVs) are events such as small duplications, deletions or insertions that occur, changes in gene copy number or exon, and finally translocation on the chromosome called Structural Variants(SVs). PCR methods [64], Sanger sequencing methods [65], Fluorescence In Situ Hybridization (FISH) [66] and next generation sequencing techniques (NGS) [67] are used to identify such mutations [68]. NGS and gene expression based methods will be mentioned in section 2.3. In this section, tests on mutation-based approaches will be explained.

#### 2.2.3.1 PCR Methods

It has been mentioned many times that cancer has a progressive disease because of arising the mutations. Based on this knowledge, most tests required for cancer diagnosis are focused on diagnosis by performing mutation analysis from tumor tissue.

Polymerase Chain Reactions (PCR) are among the most used methods for mutation analysis tests.

PCR was used for the first time in 1983 to identify *in vitro* amplification of genetic material [69]. By designing primers specific to the regions to be replicated in the genome, DNA amplification with unlimited copies is achieved. Generally, the basic principle of the Polymerase Chain Reaction is to create an unlimited copy of the desired region in the genome. It includes 3 steps on the basis of the reaction to accomplish this. These are respectively; denaturation, annealing, and elongation. In order to amplify the desired region in the genome, the double helix of DNA must be opened. This step is performed during the denaturation process. Heating is performed above the melting point of the target DNA to separate the double strand from each other. The DNA strands separated from each other by this heating process become open for the joining of the desired regions and the complementary primers. In the second step, called annealing, the temperature is lowered so that primers designed to reproduce the desired area can cling to the target areas. In the last step, the Extension step, the temperature is raised again and the regions of the DNA Polymerase enzyme established between the primer and the target DNA are completed with nucleotides. This process is defined as a cycle in every Polymerase Chain Reaction. These processes are repeated for the specified number of cycles, allowing the number of amplified DNA to be doubled Figure 2.9 [70].

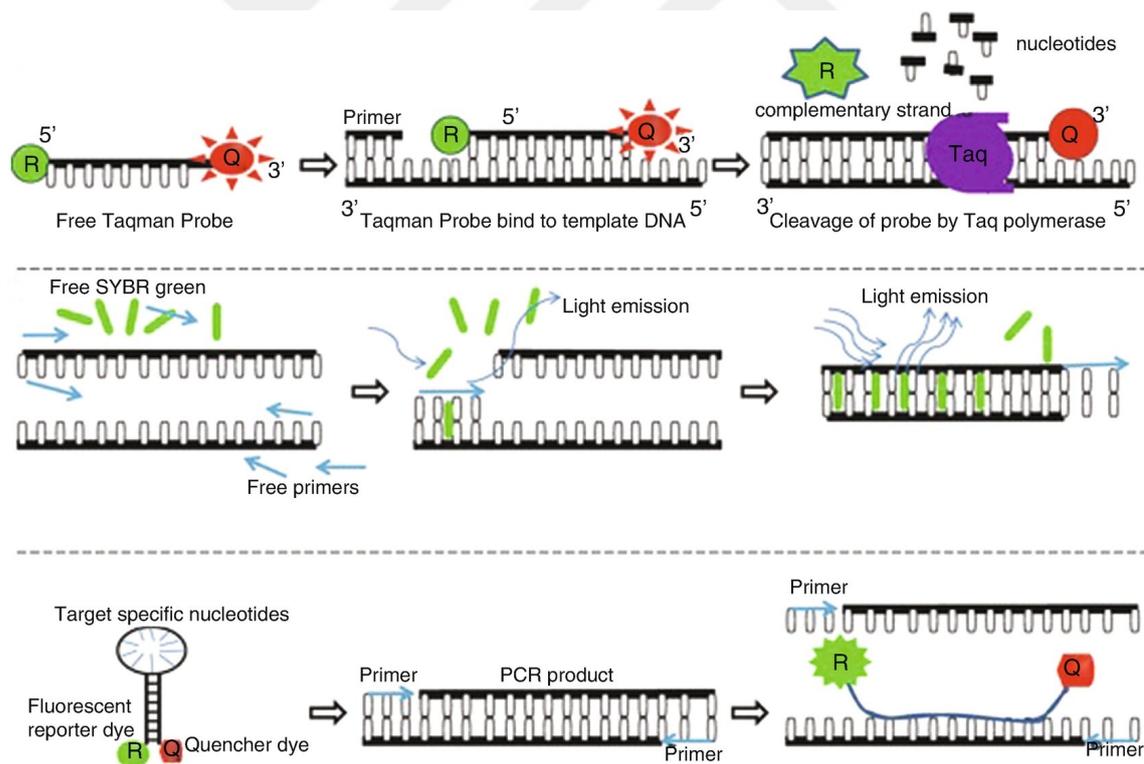


**Figure 2.9** Working principle of Polymerase Chain Reaction [70]

This is the general working principle of Polymerase Chain Reactions (PCR). In most cancer studies, different PCR techniques are used depending on the developing PCR technologies. Real-Time PCR and Allele-Specific PCR are examples of these PCR techniques.

### Real-Time PCR

Real-Time PCR is one of the PCR techniques used for cancer diagnosis. Real-Time PCR differs from PCR in that the resulting PCR products can be observed instantly via a computer [71]. Apart from that, Real Time PCR technique can be classified according to the use of different reporter depending on the imaging process. These can be arranged as Fluorescence-based Real Time PCR, DNA probe-based Real Time PCR, Taqman probe Real Time PCR. Fluorescence probes used in fluorescence-based Real Time PCR method bind non-specifically to double stranded DNA. Thus, as the amount of PCR product double stranded DNA increases, the fluorescence irradiance intensity increases [72]. It is the most commonly used reporter fluorescence dye in Real Time PCR method, but there are methods using Free Taqman dye and Quencher dye Figure 2.10 [73].



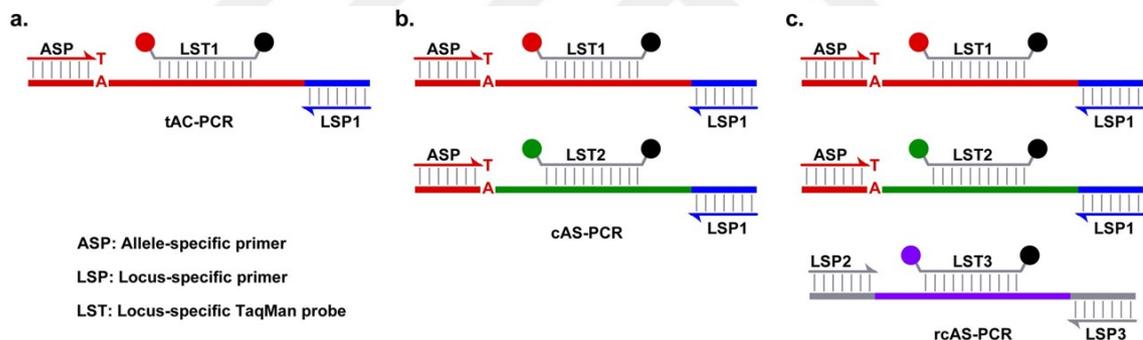
**Figure 2.10** Working principle of Real Time PCR and the probes used [73]

Mutations affecting oral cancer [74], breast cancer [75], colorectal cancer [76], lung cancer [77], pancreatic cancer [78], lymphoma [79] and ovarian cancer [80] were examined using the Real Time PCR method.

## Allele-Specific PCR

Allele Specific Polymerase Chain Reaction is a method used to detect single nucleotide polymorphisms (SNP) that occur as a result of single mutation changes. With this method, genetic diseases and molecular diagnostic analysis can be done [81].

The Allele-Specific PCR (AS-PCR) method developed by Yang et al. [82] for the detection of the *BRAF V600E* mutation has shown that it is sensitive enough to identify the allele in a single copy. In the study, it was aimed to eliminate the weaknesses of traditional Allele-Specific PCR (tAS-PCR). These vulnerabilities are low replication of the particular mutation and low mutation selectivity. To eliminate this situation, three different allele-specific primers have been designed. The designed primers are competitive external allele-specific controller (cAS-PCR), traditional allele-specific PCR (tAS-PCR) and referenced internal positive controller (rcAS-PCR) in the cAS-PCR. Plasmids named competitive allele specific controller (CEAC) were used in cAS-pCR containing the same sequence as primers targeting the human *BRAF V600E* MT-allele. With the help of these plasmids, it is aimed to eliminate non-specific binding. The working principles of allele-specific PCR specific to 3 primers designed in Figure 2.11 are shown.



**Figure 2.11** The working principles of tAS-PCR, cAS-PCR and rcAS-PCR are shown respectively [82] [73]

According to Figure 2.11, there is a single primer traditionally targeting *BRAF V600E* in tAS-PCR. Figure 2.11 shows the working principle of cAS-PCR. Accordingly, there are primers originating from CEAC plasmids in addition to tAS-PCR. In Figure 2.11, there is a primer targeting the *leptin* gene in addition to rAS-PCR. As a result of the study, it was observed that the designed rAS-PCR and crAS-PCR methods have higher mutation selectivity compared to the tAS-PCR method. By means of this method, even cancer individuals with low levels of *BRAF V600E* gene mutation can be detected. Allele Specific PCR has clearly demonstrated the importance of cancer diagnosis. Increasing the sensitivity of these methods gives hope that the challenges encountered during the diagnosis phase can be eliminated.

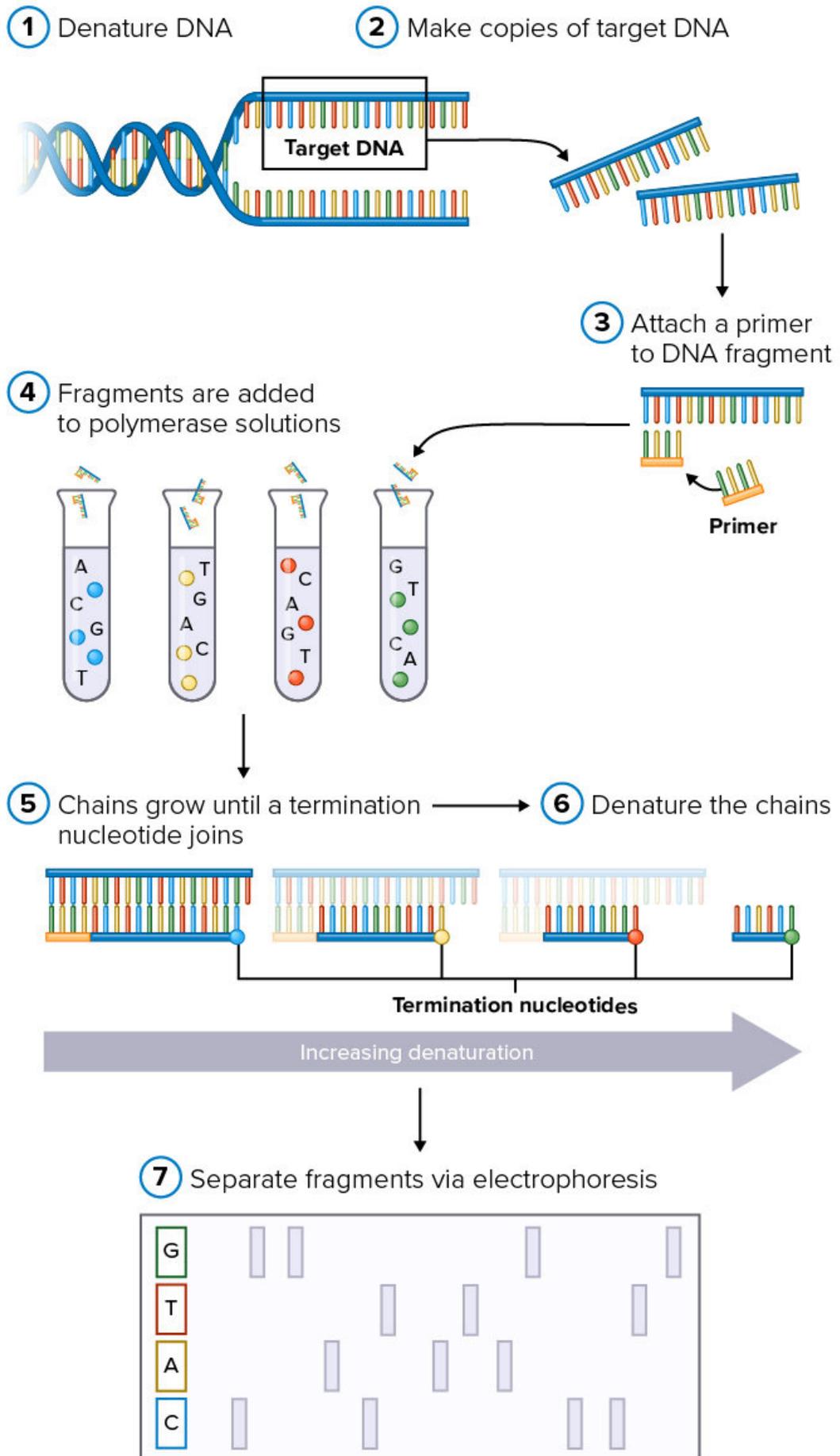
### 2.2.3.2 Sanger Sequencing

The Sanger Sequencing method was developed by Frederick Sanger in 1977 [83]. This method is also called chain-termination PCR. In addition to the materials required for PCR in Sanger sequencing, there are dideoxynucleotides (ddNTPs) that do not contain 3'-OH groups. It is found in lesser amounts than deoxynucleotides (dNTP). When DNA polymerases add ddNTPs to the sequence in the extension stage in the PCR method, chain extension is terminated. ddNTPs are also labeled with fluorescent probes. After the chain-termination PCR process is completed, the PCR products obtained using gel electrophoresis method are separated according to their sizes. In the last stage, the results obtained with the help of laser excitation or sequencing devices are interpreted. Figure 2.12.

In the PCR Methods section 2.2.3.1, a study was mentioned in which an AS-PCR method was developed for the detection of the *BRAF V600E* mutation using the Allele-Specific PCR method [82]. A study was conducted to compare the detection sensitivity of *BRAF V600E* mutation from FFPE tissue using Sanger Sequencing method, immunohistochemistry (IHC), droplet PCR and NGS [65]. In this study performed by Cheng et al., the accuracy of detecting *BRAF* mutations of tissue samples taken was compared. As a result of these comparisons, it has been shown that IHC results produce false negative results, while other methods have high accuracy in detecting mutations. They showed that Sanger Sequencing can be used as a rapid identification method for the detection of variant allele-specific *BRAF* mutations.

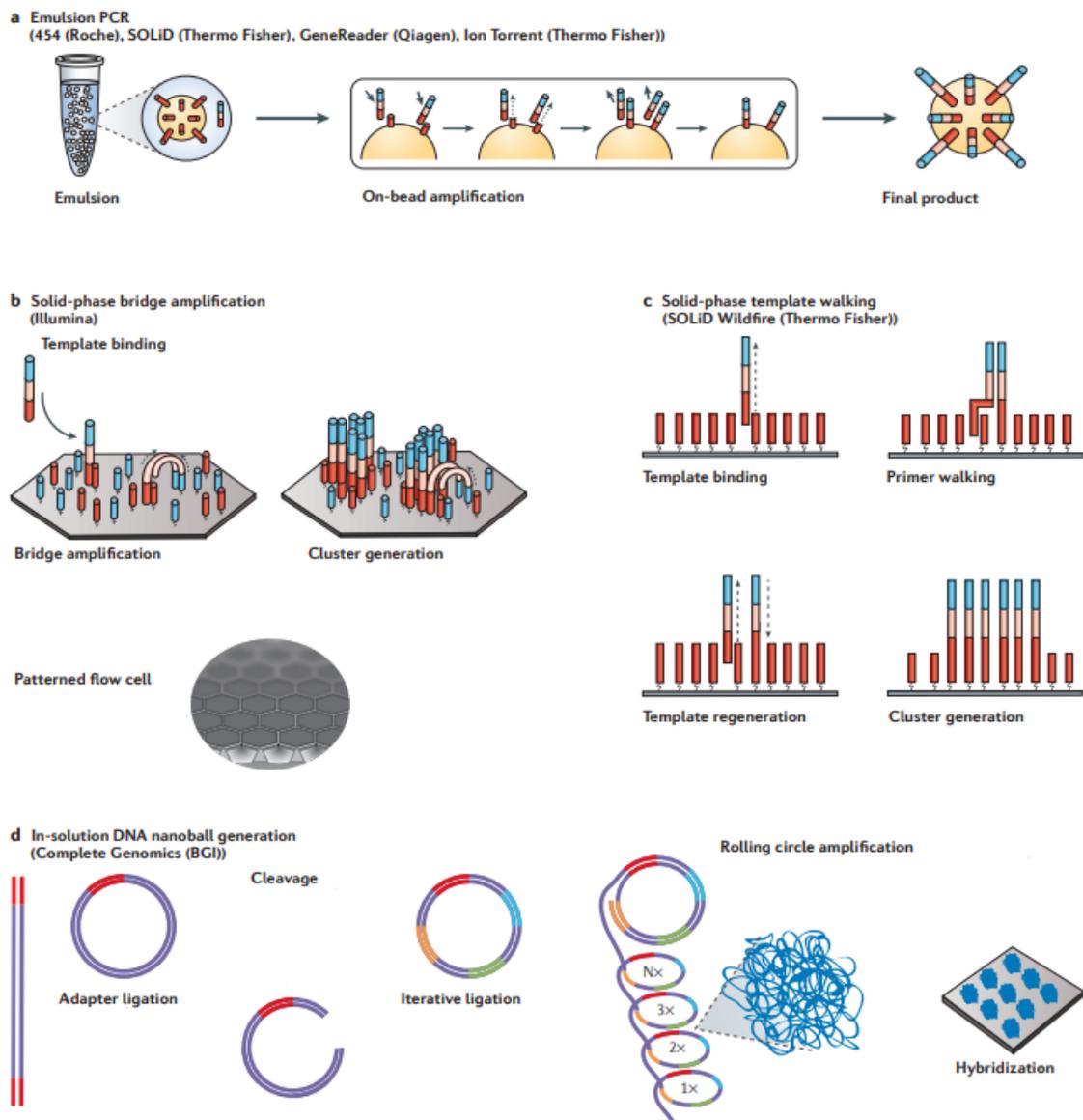
## 2.3 Gene Expression Based Cancer Diagnosis

It is known that cancer is a disease with a mutation-based progression process [63]. For this reason, changes in the genome are analyzed to identify cancer and many other mutation-based diseases. The most important of these are sequencing methods. After the sequencing method developed by Frederick Sanger in 1977, studies were carried out to eliminate the difficulties (time-consuming, high-cost etc.) of sequencing studies. The first of these studies is the work developed by Martin et al. [85], which prevents the loss of time by automating the repetitive processes of Sanger sequencing. By virtue of this work, the time for a complete sequencing procedure has been reduced to 45 minutes. This has largely eliminated the time-consuming part of the sequencing process. Although this process solved the time-consuming problems of the sequencing process, it could not meet the needs of increasing knowledge about the nature and biology of diseases and sequencing the whole genome for new analyzes [86]. This has led to the development of many groundbreaking methods in the field of sequencing [87]. These methods are generally called Next Generation Sequencing



(NGS) methods.

NGS methods are classified as short-read sequencing and long-read sequencing. Short-read sequencing is also known as the second generation sequencing method. This method is the first development in the field of sequencing after the Sanger sequencing method. It is generally used for sequencing short base sequences (about 750 – 800 kb). They are also called parallel sequencing because more than one array is processed at the same time [88]. The short-read sequencing method has 3 stages in itself. These; library preparation for sequencing, sequencing and data analysis. These steps are repeated for different types of samples to be sequenced.



**Figure 2.13** Overview of short-read sequencing methods. a. emulsion PCR, b. solid-phase bridge amplification, c. solid-phase template walking, d. DNA nanoball [89].

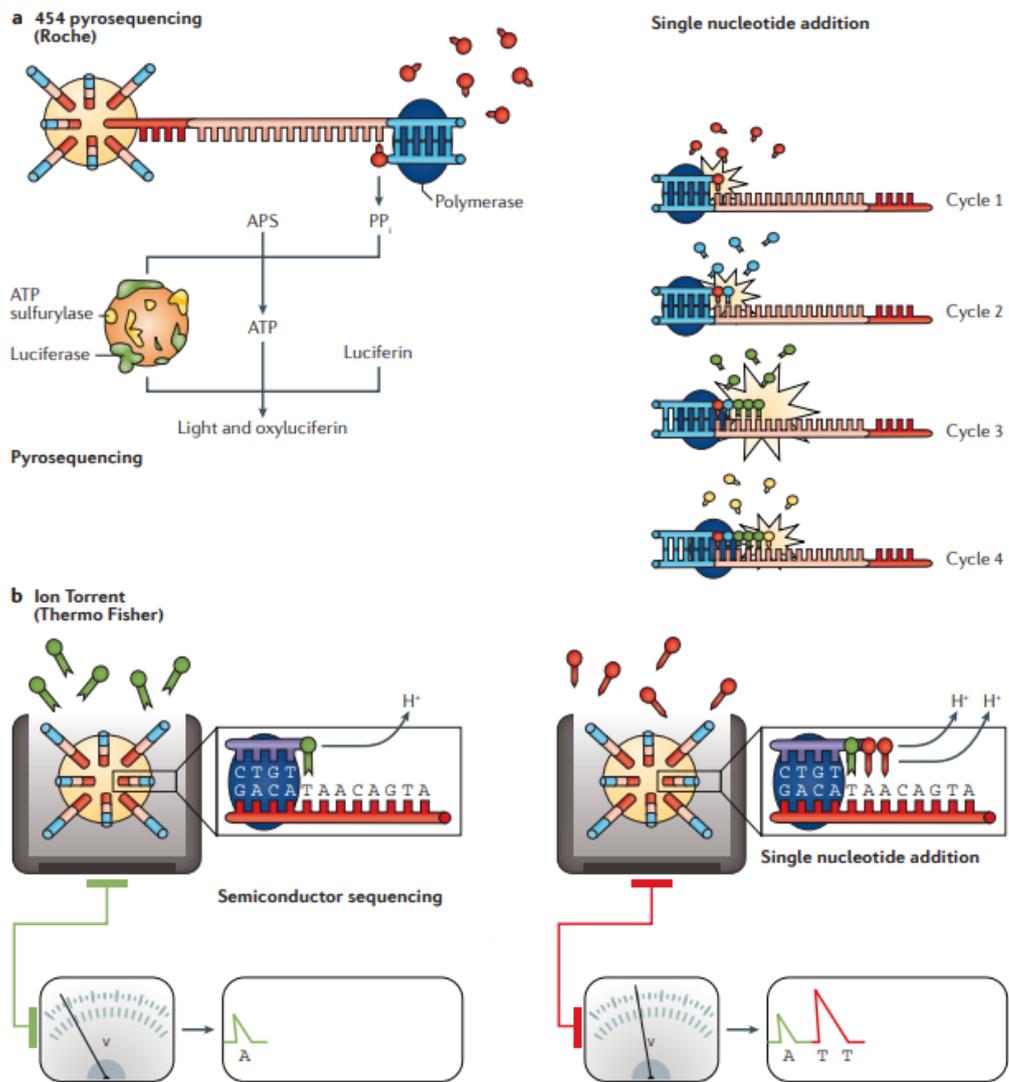
Figure 2.13 shows an overview of short-reading sequencing methods. Figure 2.13.a

shows the Emulsion-PCR method. This method basically converts each of the duplicated DNA sequences into magnetic beads. DNA molecules transformed into magnetic beads are stained with the help of fluorescence probes and measured using flow cytometry. In this way, counting of DNA molecules containing the variant can be performed [90]. Figure 2.13.b shows the solid-phase bridge amplification method developed by Illumina. According to this method; the primers are fixed to the developed plate. Fragmented DNAs are linked to fixed primers to form double-stranded DNA. Figure 2.13.c shows the solid-phase template walking method developed by Thermo Fisher. In this method, similar to the method developed by Illumina, primers fixed to the solid surface were used. However, unlike solid-phase bridge amplification, the double-stranded templates are partially denatured, allowing them to bind to different primer sequences. Thus, reverse primers are formed. These primers create templates that can bind to a new primer. Figure 2.13.d shows DNA nanoballs. In this method, adapter arrays are created. The cleaved DNA sequences bind to the first of the adapter sequence and amplification occurs. Then it is divided with the help of endonuclease and a second set of adapters is added and the process is repeated. These resulting DNA nanoballs are then analyzed using flow cytometry [89].

Third generation sequencing methods are also known as long-read sequencing. Long-read sequencing has emerged as a result of the fact that short-read sequencing does not meet the needs, based on the knowledge that the complex structure of the genome and the relationship of diseases with the genome. With this method, sequences larger than 10 kb can be generated from the DNA sequence [91].

Figure 2.14 shows two methods as examples of long-read sequencing. Pyrosequencing (Figure 2.14.a) uses single nucleotides to synthesize complementary chains. Pyrophosphate (PPi) is released into the environment for each complementary compound. PPi is then converted with the aid of enzyme-catalyzed reactions to determine the number of successful couplings. Thus, the DNA sequence is obtained [92]. In the Ion Torrent method (Figure 2.14.b) developed by Thermo Fisher, the hydrogen ion release (H<sup>+</sup>) is measured by a metal oxide semiconductor (CMOS) and an ion sensitive field effect transistor (ISFET). After each nucleotide addition, washing is done to remove unmatched nucleotides and sequencing is performed by adding the next [93].

In addition to all these methods, there are also lower cost methods (microarray, qPCR etc.). This section will focus mainly on Microarray and RNA-seq method.



**Figure 2.14** Overview of long-read sequencing techniques. a. pyrosequencing, b. Ion Torrent [89].

### 2.3.1 Methods for Gene Expression Quantification

In this section, the gene expression quantification methods Microarray and RNA-Seq will be discussed. The applications of these methods in cancer studies will be reviewed.

#### 2.3.1.1 Microarray

Microarray technologies have a technology that allows sequencing long genome sequences in prompt [94]. It has been proven with the development of sequencing technologies that the information obtained through the genome is much more beneficial in terms of the process to be followed in the diagnosis and treatment of diseases. However, genome sequencing was a time-consuming process, as in the example of Sanger sequencing, and only certain regions of the genome could be sequenced. Based on this need, DNA microchips were developed, in which the entire genome can be sequenced at once. These glass chips were called spots. Each spot contained gene-specific probes fixed to its surface. In this way, after an isolated mRNA sample is converted into complementary DNA, it is placed in these spots and the gene sequences present in the sample are detected by fluorescence [95] (Figure 2.15).

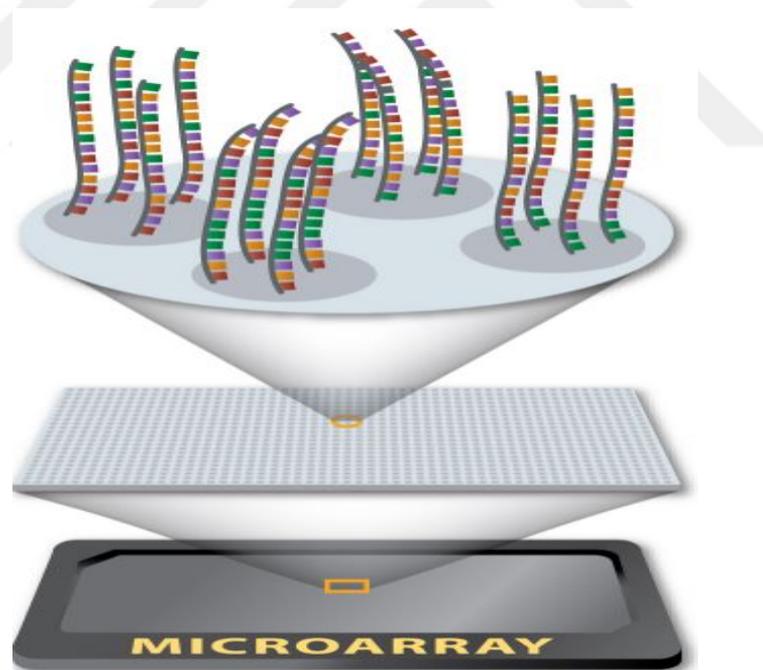


Figure 2.15 Overview of Microarray [96]

In cDNA microarrays, mRNAs were stained with two different fluorescent dyes (Cy3 and Cy5), making it possible to perform sequencing from 2 different cell or tissue samples at the same time. In high-density oligonucleotide arrays, oligonucleotide sequences are selected from the reference mRNA sequence of each gene and oligonucleotide syntheses specific to these regions are performed. In the transcription

step, cRNAs are created using biotin-labeled nucleotides. The bindings that take place are detected by fluorescent dyes coupled with streptavidin. The signal intensity from the dyes gives information about the mRNA density for the genes in the sequence under investigation [97]. The fact that the change in products resulting from a change in a single gene such as cancer affects the course of the disease has led to the development of new microarray strategies. A gene product whose expression is increased by a gene present in tumor cells provides a lot of information about cancer biology. This approach has allowed the development of tissue microarrays. Many tumor markers have been detected through microarray analysis [98, 99].

By virtue of these developed features of microarray technology, tumor microarrays have started to be used a lot in the field of cancer to obtain information about biomarkers and the cancer process. Hickey et al. (2021), examining the androgen receptor as a biomarker for ER+ breast cancer, using tumor microarray, it was proven that the androgen receptor is a tumor suppressor. They also performed a tumor microarray experiment to determine the stages of breast cancer [100].

Microarray experiment does not only consist of in vitro steps. There are many genes on the spot used in the experiment phase, and as a result of the experiment, data belonging to these genes occur. Meta-analysis methods are used to make sense of this dense data. These methods will be described in section 2.3.2.

### 2.3.1.2 RNA-Seq

With the development of next-generation sequencing (NGS) techniques, there have been many improvements in transcriptome profiling, that is in the analysis of gene expression data. The most important of these is RNA sequencing (RNA-seq). The difference of RNA-seq from microarray is that it performs direct sequencing after converting RNAs to cDNA. The cost of this process, which was very expensive in the beginning, started to decrease over time, so it became a candidate to be used instead of microarray [101, 102].

In the RNA-seq experiment (Figure 2.16); long sequences are converted into cDNA fragments by RNA or DNA fragmentation. Sequencing adapters are added to each cDNA fragment, the read sequences are aligned with the reference genome or transcriptome. In some cases, there is no reference genome or transcriptome. In such experiments, it is aimed to find similar genomes by *de novo* genome assembly. After alignment with the reference genome or transcriptome, sequencing reads are aligned to profile expression. As a result of this process, the expression profile of the genes in the sequenced cell or tissue is created [101].

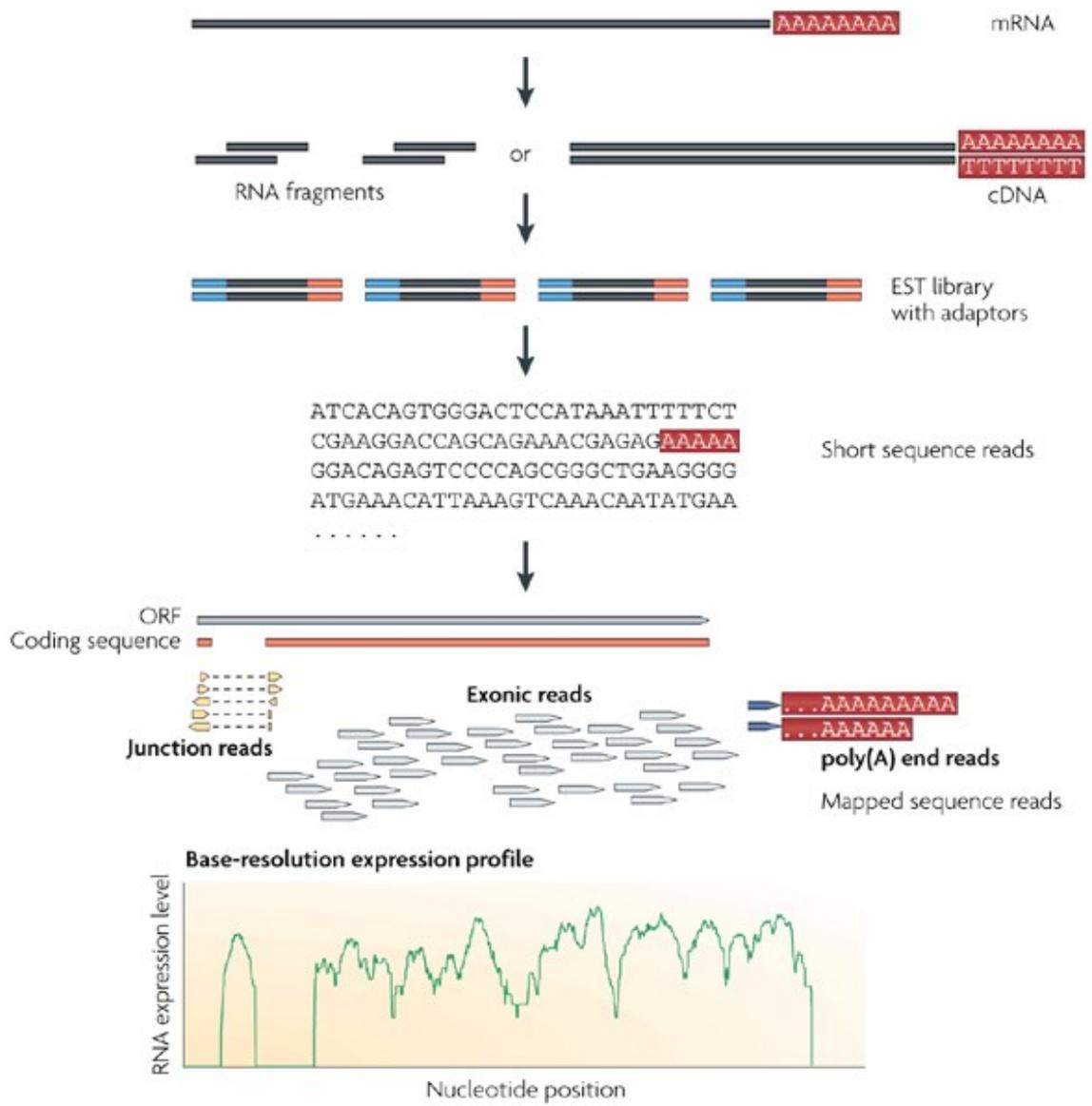


Figure 2.16 Common RNA-seq experiment [101].

Although RNA-seq seems to have many advantages over microarray, many researchers have preferred to examine RNA-seq and microarray comparatively. In some cases, microarray experiments have even been performed to validate RNA-seq results [103, 104]. RNA-seq microarray comparison experiments performed on cancer have shown that RNA-seq is superior to the detection of differentially expressed genes (DEGs) compared to microarray [105].

Xu et al. (2013) [105], the detection of changes in gene expression values using by RNA-seq and microarray in the treatment of HT-29 colon cancer with 5-Aza was investigated. First, its sensitivity in detecting these changes on its two platforms was investigated. As a result, RNA-seq was found to be more successful in detecting genes with low expression levels. It was also observed that RNA-seq was superior in the comparison of RNA-seq and microarray programs used to identify differentially expressed genes. These results prove that RNA-seq gives more sensitive results than microarray. RNA-seq method also has many advantages such as identifying biomarkers in cancer and providing information about genes that are effective in cancer progression.

The data obtained as a result of the RNA-seq process goes through an analysis process as it is microarrayed. These analysis methods will be described in section 2.3.2.

### **2.3.2 Meta-analysis**

Meta-analysis is an approach that provides a holistic conclusion by analyzing the results of independent but interrelated studies [106]. Many medical data were processed using meta-analysis method and inferences were obtained [107]. Many inferences were obtained by using meta-analysis methods for cancer. Determination of circular RNAs that can be used as a biomarker in cancer by compiling the results of the research carried out, [108], the cancer prevention feature of statins, which is a cholesterol-lowering drug [109], the effect of obesity on pancreatic cancer [110]. The data of the studies conducted were interpreted using meta-analysis methods and a conclusion was reached. The gene expression quantification methods (such as Microarray and RNA-seq) mentioned in Section 2.3.1 also generate a lot of data that can be analyzed by meta-analysis methods. In particular, meta-analysis methods are used to extract cancer-based gene expression signatures [111].

In this section, differentially expressed genes (DEGs) and network-based approaches, which are meta-analysis methods of data obtained as a result of gene expression quantification techniques such as microarray and RNA-seq, will be discussed.

### **2.3.2.1 Differentially Expressed Genes (DEGs)**

The most important step in the diagnosis and treatment of cancer is the identification of mutated genes. These genes, which are called biomarkers, can be easily identified by the development of sequencing technologies and the analysis of gene expression data [112]. Differentially expressed genes (DEGs) include various statistical methods used to identify genes whose expression varies significantly in diseases with genomic instability such as cancer [113]. By determining the DEGs of the cancer type under examination, biomarkers specific to that cancer types are determined.

In general, DEGs are obtained as a result of analysis of microarray and RNA-seq experiments. Obtaining the expression of thousands of genes simultaneously, such as microarray, causes the analysis of experimental results to be very time-consuming. Therefore, there is a need to separate the genes that are important from the gene expression results obtained. For this purpose, genes showing a change in expression are selected and ranked by using statistical methods [114–116].

Using gene expression profiles produced for cancer, identification of critical genes and pathways effective in colorectal cancer [117], determination of biomarkers for gastric cancers [118] and testicular cancer [119] in many areas of cancer biology, diagnosis and the information needed for its treatment was obtained. This information obtained is generally obtained from microarray test results or statistical analysis of publicly shared microarray test data.

Determination of DEGs is not only possible by statistical analysis of experimental results. Through the advancements in bioinformatics, deep learning techniques are also used to find the genes responsible for cancer over experimental data. The time-consuming process in the interpretation and analysis of data obtained using deep learning methods has been shortened and effective results can be obtained [120].

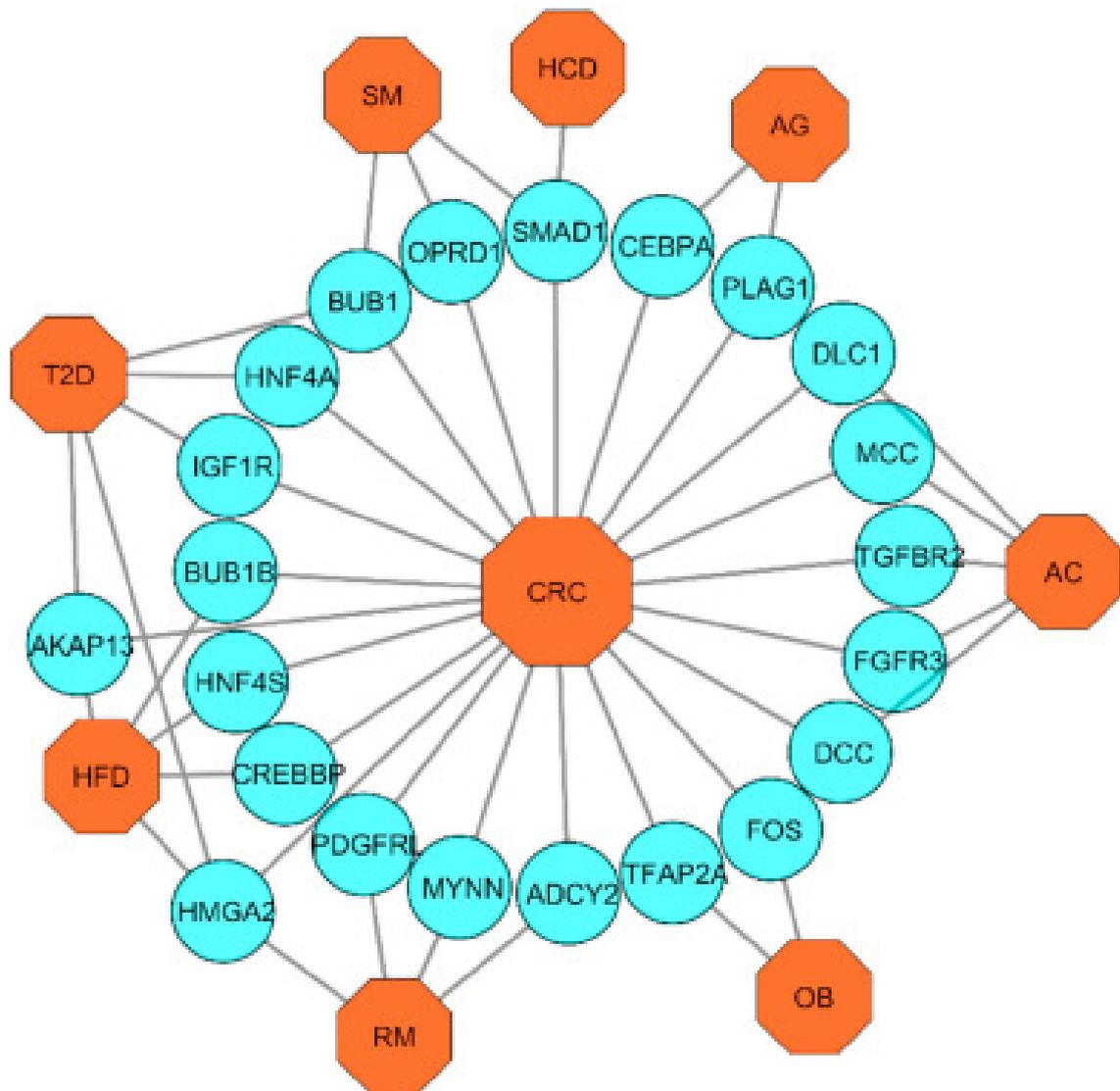
### **2.3.2.2 Network Based Approaches**

In diseases of high heterogeneity and complexity, such as cancer, the analysis of the influence of genes on the occurrence or course of the disease poses great challenges. Network concepts are used to overcome this existing complexity and analyze the relationship between genes [121].

Network-based approaches have been used to identify biomarker genes for breast cancer [122], colorectal cancer [123] and thyroid cancer [124]. In addition to these, it is seen to be used in approaches such as [125] in determining the relationships of cancers with other diseases that may be related, and in determining biomarkers for

drug therapy that can be used effectively in cancer types [126].

Hossain et al. (2021), in the study conducted by [127], they found that the effect of risk factors that cause colorectal cancer formation on survival time may occur as a result of changes in the gene expression of cells. Starting from this hypothesis, they used a network-based approach, using machine learning techniques to explain the molecular mechanism. As a result of this study, they identified seven critical genes that act on patient survival.



**Figure 2.17** Network of genes involved in colorectal cancer formation [127].

The figure 2.17 shows the DEG network of 8 risk factors causing colorectal cancer. By using machine learning and statistical analyzes on microarray data, critical genes showing activity in survival time on colorectal cancer were determined by network-based approach.

These and similar methods, especially by combining machine learning - deep learning

methods, help to illuminate the complex structures of diseases such as cancer and to provide treatment opportunities.



Deep learning is a branch of artificial intelligence technologies. With the developing technologies, the improvements in neural networks have also accelerated and the layers used to increase the success of training in machine learning have been increased. In this section, the general architecture and learning algorithms of multi-layer deep learning techniques from single-layer machine learning techniques will be discussed. In the last part, deep learning examples used in cancer diagnosis will be mentioned.

### 3.1 History and Basics

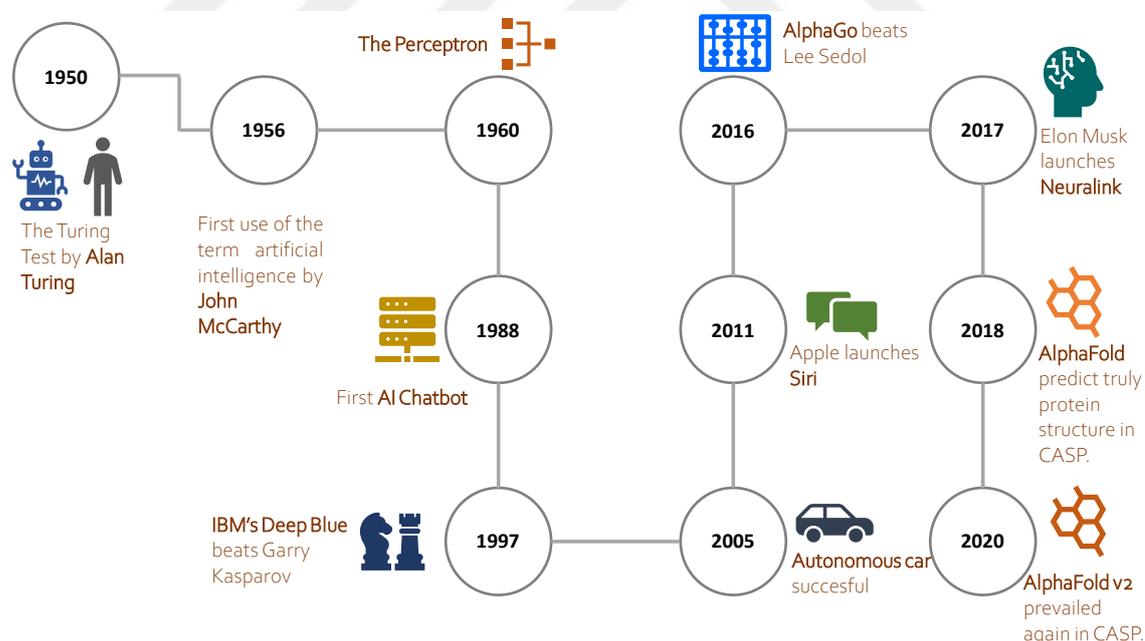


Figure 3.1 Timeline of AI

Deep learning can basically qualify as one of the sub-branches of artificial intelligence (AI). AI is a phenomenon that humanity has been curious about and wanted to realize since ancient times. "Can machines think?" by Alan Turing for the first time in history in 1950. Based on the question, the concept of AI came into being. He developed the Turing Test to prove this theorem. In the Turing Test, answers given to various

questions are transmitted to an interrogator and the interrogator tries to distinguish whether these are human or machine answers. According to Turing's hypothesis, he argued that if the questioner cannot distinguish the answers given by the machine from the human responses, it can be seen as evidence that machines can think, and he succeeded. The Turing test is one of the methods used to distinguish machines from humans even today. It is seen that intelligence is evaluated in connection with linguistics in the Turing test. The perception and response capacity of the machine is related to its linguistic functionality [128].

Following Turing's study, the term AI was first used by John McCarthy in the Dartmouth Conference in 1956 and was described as the birth of AI. [129, 130]

In 1960, the Perceptron entered the history of AI as the first supervised algorithm designed. This single layer binary algorithm was originally designed as a device. This device for image recognition has an array of 400 photocells and electric motors that perform weight updates during learning (Figure 3.2).



**Figure 3.2** MARK I Perceptron [131]

This approach, which is the first learning algorithm in history, has become dysfunctional over time as it has difficulty in detecting tasks that require multi-layer algorithms. This failure in the Perceptron has caused stagnation in the history of AI for many years [132].

In 1988, the first AI-based chatbot was launched by Jabberwacky company, and some transactions were gradually automated [133].

In 1997, with the developing computer technology, there was a period when games were also used. AI has found a place in this field. Deep Blue, an algorithm developed by IBM for chess, defeated world chess champion Gary Kasparov and gave a strong answer to the question of whether machines can think [134].

Developed by Stanford University, Stanley was successful by driving 132 m in driverless car races in 2005, which was conducted by the Defense Advanced Research Project Agency of the US Department of Defense, based on the production of technology for use by the US military [135].

In 2011, another AI with linguistic competence was launched. Apple's Siri application has come to the fore as a voice assistant with an AI algorithm [136].

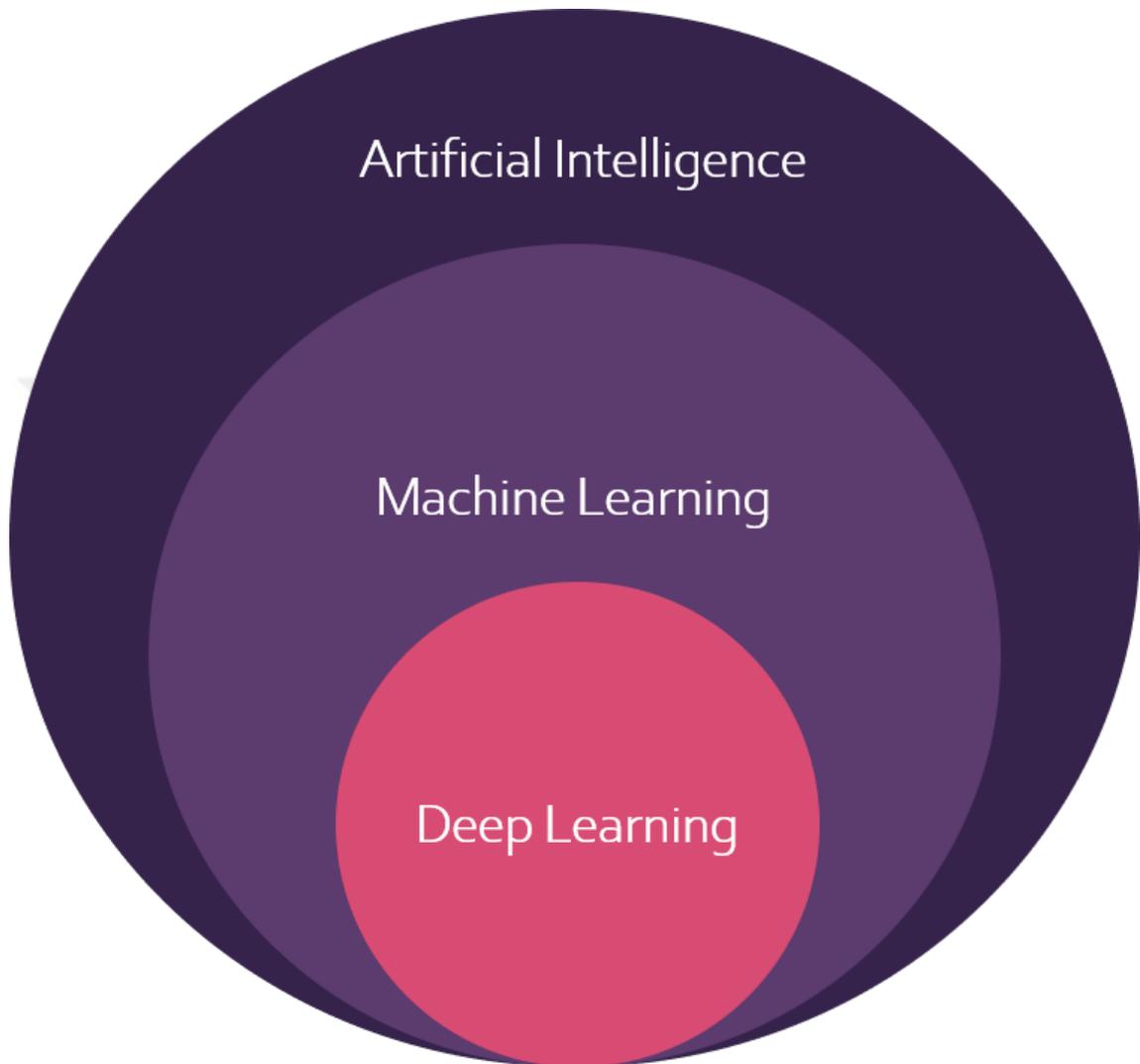
In 2016, an algorithm developed for Go, another strategy game similar to chess, defeated the human world Go champion. In the fight between AlphaGo and Go world champion Lee Sedol, developed by Google's DeepMind team, the algorithm succeeded by taking 3 out of 5 matches [137].

In 2017, the project that made a big repercussion and gave the impression of science fiction for the solution of diseases, was put into the brains of people with diseases such as Alzheimer's and Parkinson's, and the project was announced by Elon Musk. With this project called Neuralink, it is aimed that individuals with diseases that cannot fulfill their vital functions can overcome these deficiencies [138].

Finally, if we come to the present from the history of artificial intelligence, it is seen that AI technologies produce solutions on biological problems and are successful. In the competition held by The Critical Assessment of Protein Structure Prediction in 2018, algorithms that can predict the closest protein folding by comparison with the experimental results competed. In these competitions held in 2018 and 2020, AlphaFold, which was developed by Google's DeepMind team, was successful as the algorithm with the closest predictive values [139].

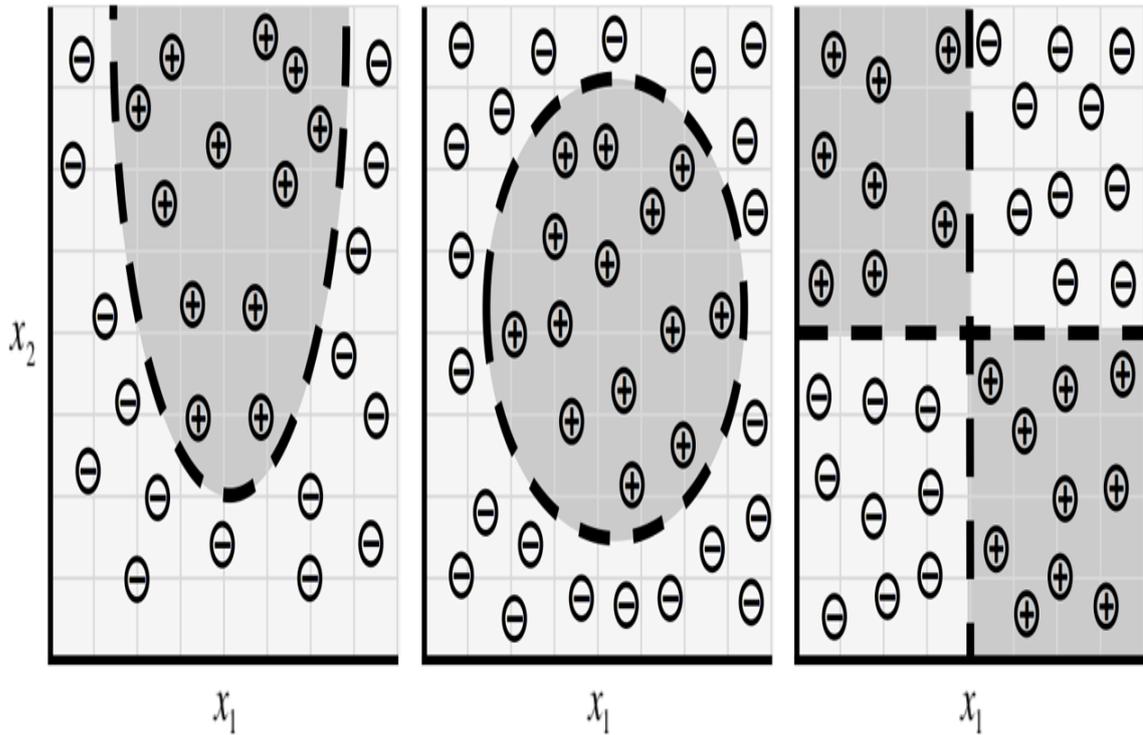
### 3.2 Deep Neural Networks

Developed in 1960 for single-layer binary classification, the Perceptron formed the basis of neural networks [132]. Deep learning is based on AI, learning algorithms with multi-layered neural networks (Figure 3.3).



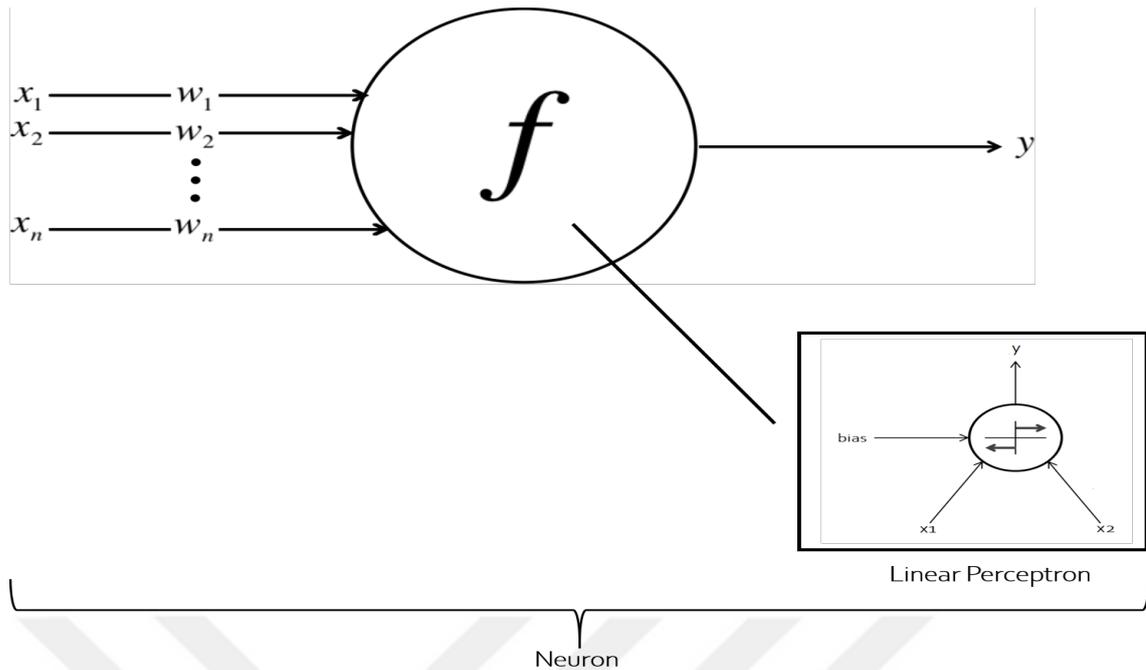
**Figure 3.3** The relation between AI and deep learning

Neural networks are based on the Perceptron algorithm, which consists of a single neural network. This single-layered neural network is generally called a linear Perceptron. It has already been mentioned that it does binary classification. In this context, it is necessary to have a linear distribution for the data to which Linear Perceptron can be applied. If the dataset in question has a distribution as in Figure 3.4, it is not possible for the linear perceptron to classify these data [140].



**Figure 3.4** Possibility of Complex Distribution [140]

Deep learning comes into play here. Said linear perceptrons can be perceived as a single neuron. But for such a complex system, more than one neuron is needed. The connection between the Neuron and Perceptron is shown in Figure 3.5. Each neuron can represent different models depending on the function it has, and this model may contain different classification algorithms. But the perceptron is only one of the model's algorithms. While the linear perceptron is limited by its algorithm, neurons may differ on the basis of their models.



**Figure 3.5** Relation between Neuron and Perceptron

More than one neurons are needed for the learning process, which has complex inputs, to take place. Neurons combine to form layers, and the human brain is known to have 6 layers when it is taken as an example of the human brain [141].

In the light of this information, various deep learning algorithms with different layers and different features have been developed. In this section, first of all, the Multilayer Perceptron network, which is the first of multiple layers, will be mentioned and then Convolutional Neural Network and Recurrent Neural Networks will be detailed.

### 3.2.1 Multi Layer Perceptron (MLP)

The transition to the multilayer algorithm has been with multilayer perceptrons. Multilayer perceptron basically contains an input layer, a hidden layer and an output layer. Nodes other than input nodes in the algorithm have a nonlinear activation function, and use a backpropagation algorithm to reduce errors during training. MLP can be used to separate datasets that are not linearly distributed [142].

A neural network structure with a single hidden layer is shown in Figure 3.6. It has been mentioned before that nonlinear data are difficult to solve with Perceptrons. For the learning phase of non-linear data, 3 different neurons are used in the MLP approach.

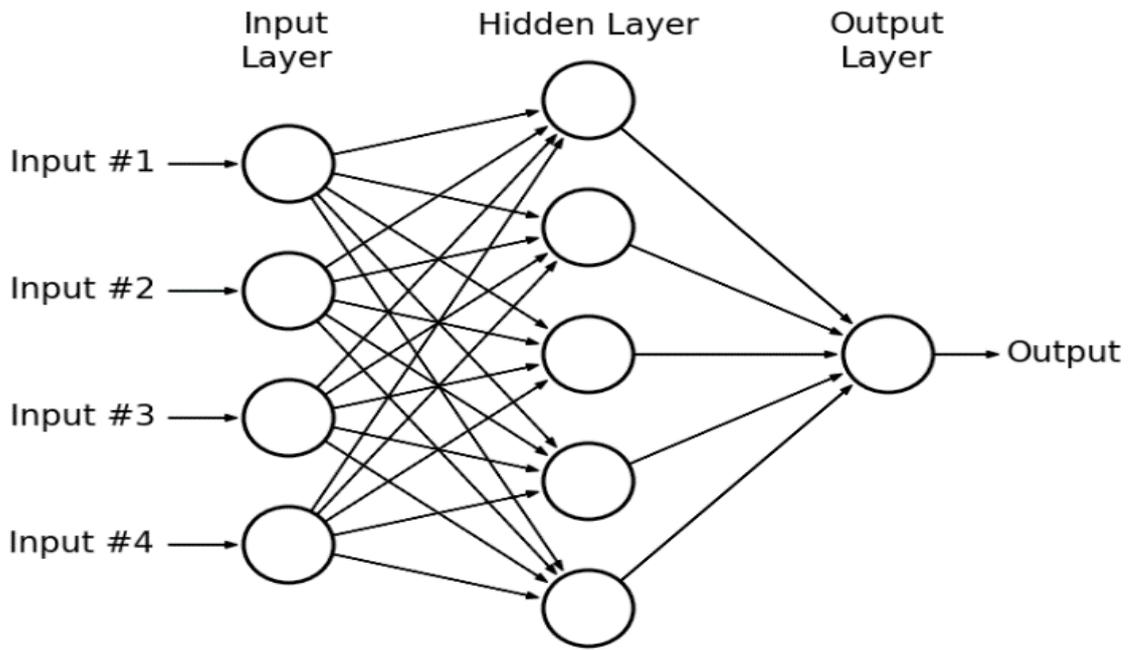


Figure 3.6 Architecture of MLP

- Sigmoid Neuron

It is used in the classification of values between 0 and 1, the closer the data is to zero, the smaller, the closer to 1, the greater the meaning. (Figure 3.7)

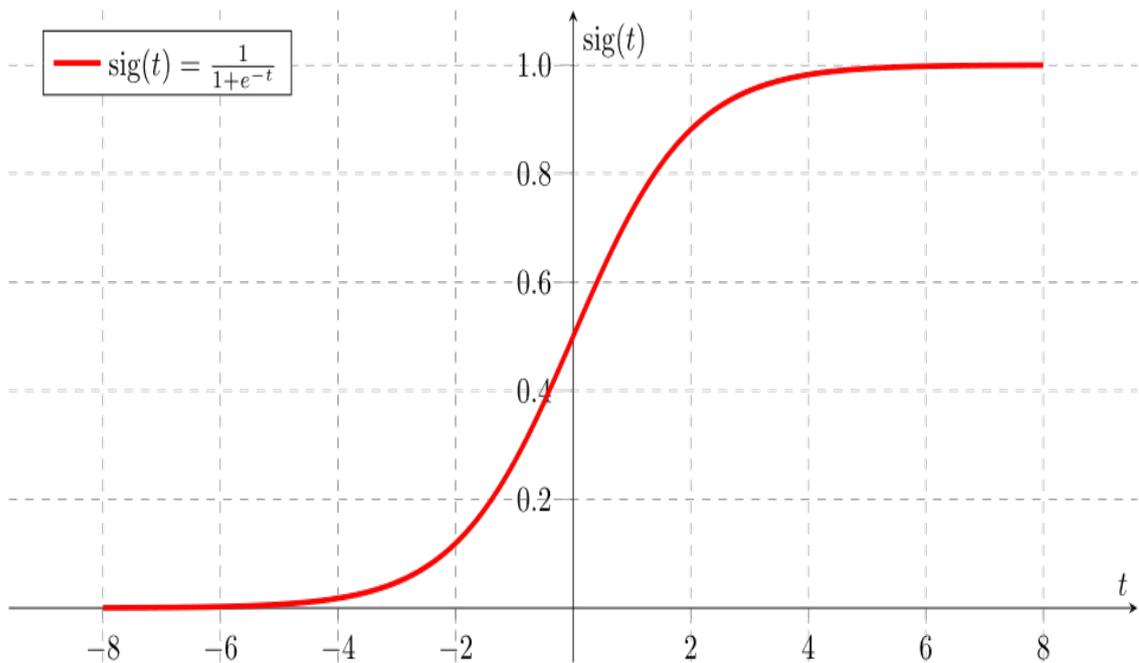


Figure 3.7 Sigmoid Neuron

- **Tanh Neuron**

Tanh neuron is used for data in the range of -1 to 1. (Figure 3.8)

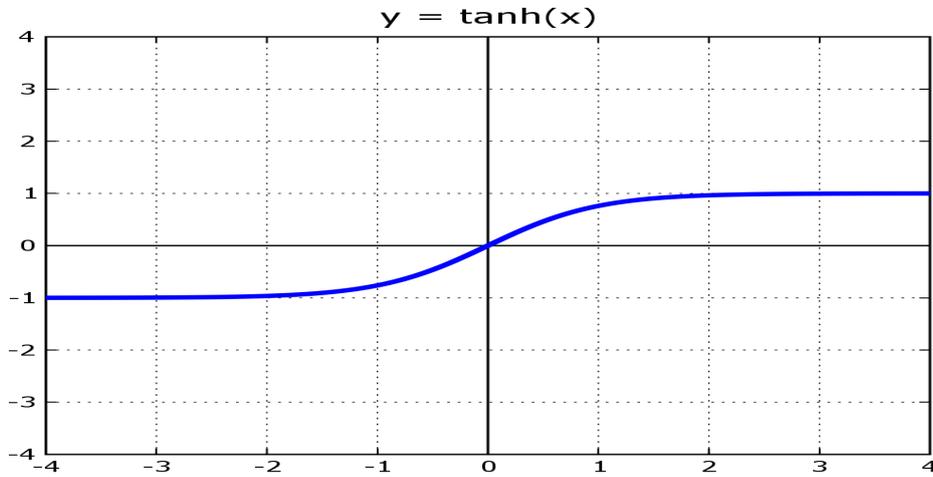


Figure 3.8 Tanh Neuron

- **Restricted Linear Unit (ReLU) Neuron**

The ReLU function is the most preferred layer in nonlinear distributions. performs operations according to the equation 3.1.

$$f(x) = \max(0, x) \quad (3.1)$$

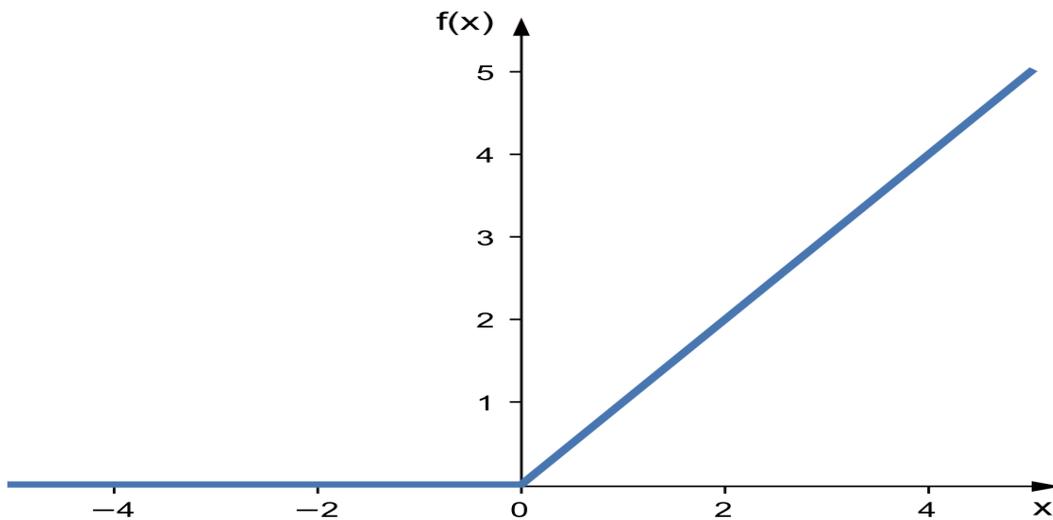
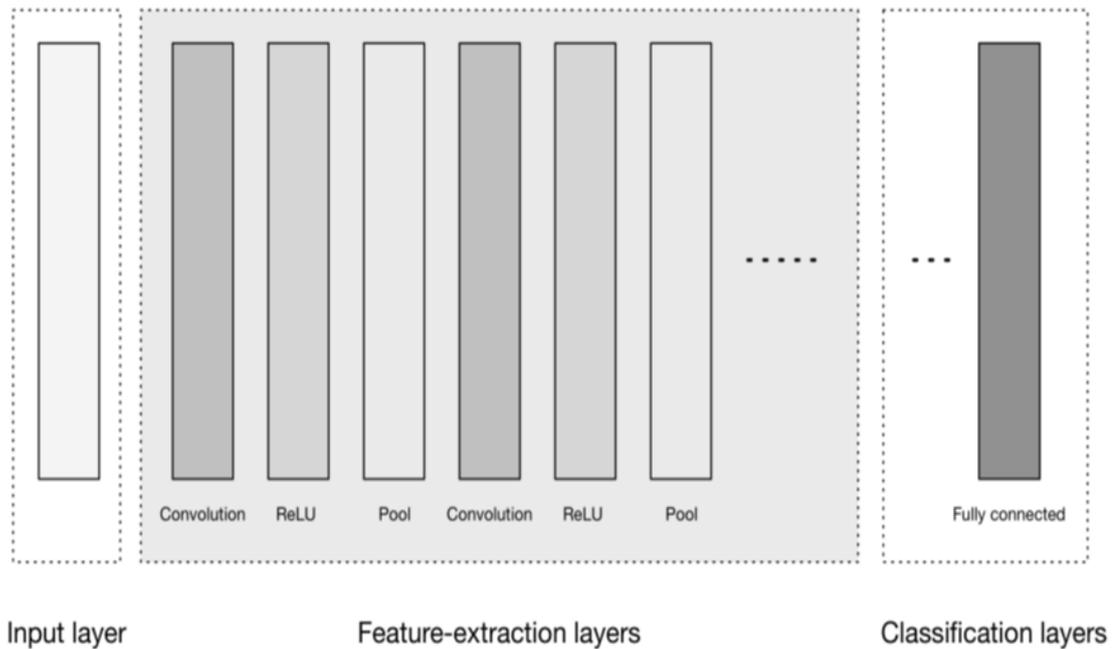


Figure 3.9 ReLU

### 3.2.2 Convolutional Neural Network (CNN)

One of the deep learning neural networks is the Convolutional Neural Network. The CNN algorithm takes the images as input data and performs the training process by extracting features from them.

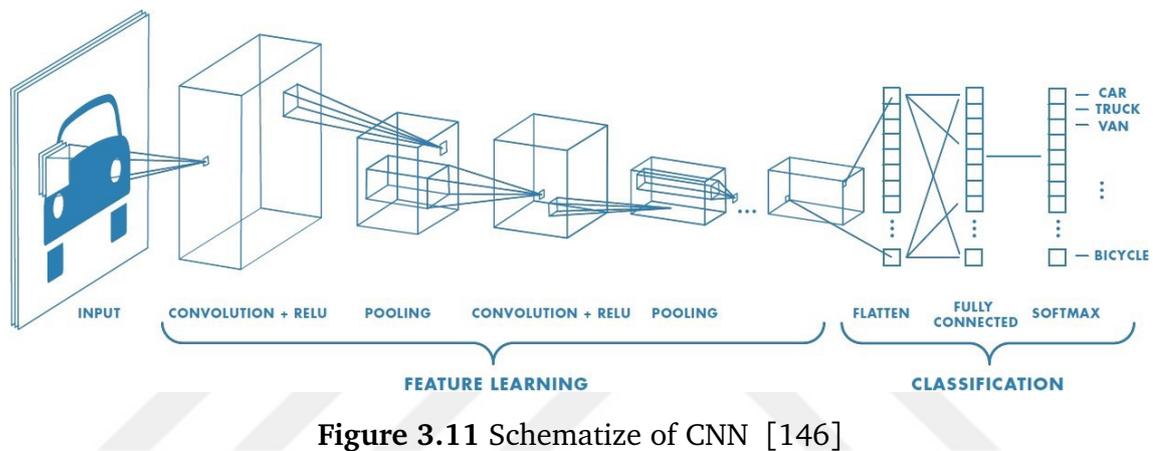
The CNN algorithm is most commonly used in image recognition, but is also suitable for input data of different sizes, such as signals and sequences (1D), images and sound spectrograms (2D) and color images in RGB format (3D). The architecture of the algorithm is shown in Figure 3.10.



**Figure 3.10** CNN Architecture [143]

Each layer has a specific task (ReLU, pool etc.). Through these tasks, the algorithm is able to make feature extraction from the input data, thereby demonstrating high learning properties [144].

Generally, a picture is needed as an input for convolution neural networks. convolution options are determined in accordance with the dimensions of these pictures and the convolution process is performed first, then the pool and the convolution and pool layers can be added once more if desired, according to the data we have. The number of layers is determined according to the data. In the last layer, the output is obtained by transforming these 3-dimensional layers into a one-dimensional and fully connected layer (Figure 3.11) [145]



**Figure 3.11** Schematize of CNN [146]

### 3.2.3 Recurrent Neural Networks (RNN)

Sequential information flow is used in a recursive neural network. In a general neural network, all inputs and outputs are assumed to be independent from each other. RNNs are called repeatedly, they do the same task for each element of an array. The output depends on previous calculations. Another definition of RNNs is that they have a memory that collects information about what has been calculated so far. In theory, RNNs can use information in arbitrarily long strings, but in practice they can only look back a few steps.

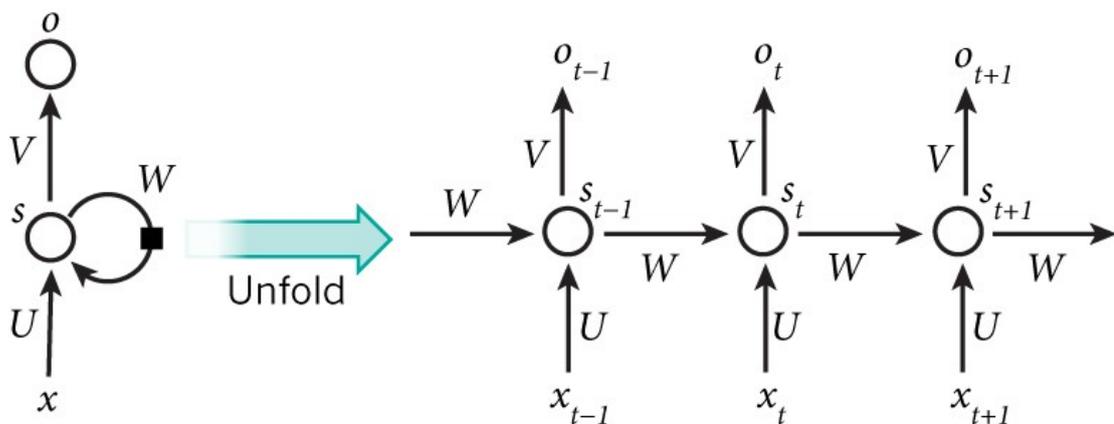


Figure 3.12 RNN Architecture

Figure 3.12 shows an RNN opened to a full network. For example, if the predicted sequence is a 5 word sentence, the neural network creates a 5-layer neural network for each word, each layer expressing one word.

Creates long time gradients for consecutive long sequences. Therefore, there are density losses. For this, it is tried to find a solution to this problem by expanding it as Long Short Term Memory (LSTM) cell. The main idea is to initiate another process that controls the LSTM cell that controls the flow of information throughout the array [145].

### 3.2.4 Vulnerability of Deep Learning Techniques: One-Pixel Attack

Deep learning methods have been actively used to solve complex problems with the help of deep neural networks. It has been seen that, deep learning algorithms used in every field have large security vulnerabilities. Attacks against these vulnerabilities are examined in two classes. The first one is an exploratory attack to gather information about the attacker's machine learning which under attack, and the other one is an evasion attack directly aimed at the algorithm's decision mechanism, which causes it to give wrong results [147].

The most well-known examples of these attacks on deep learning methods are the algorithm predicting the picture as a gibbon as a result of the attack on the panda picture in the GoogLeNet dataset [148]. It has been proven that the final decision of the algorithm can be changed by performing a one-pixel attack with the differential evolution method on the image data in the CIFAR-10 and ImageNet datasets, which is another example of security limitations in machine learning algorithms. [149].

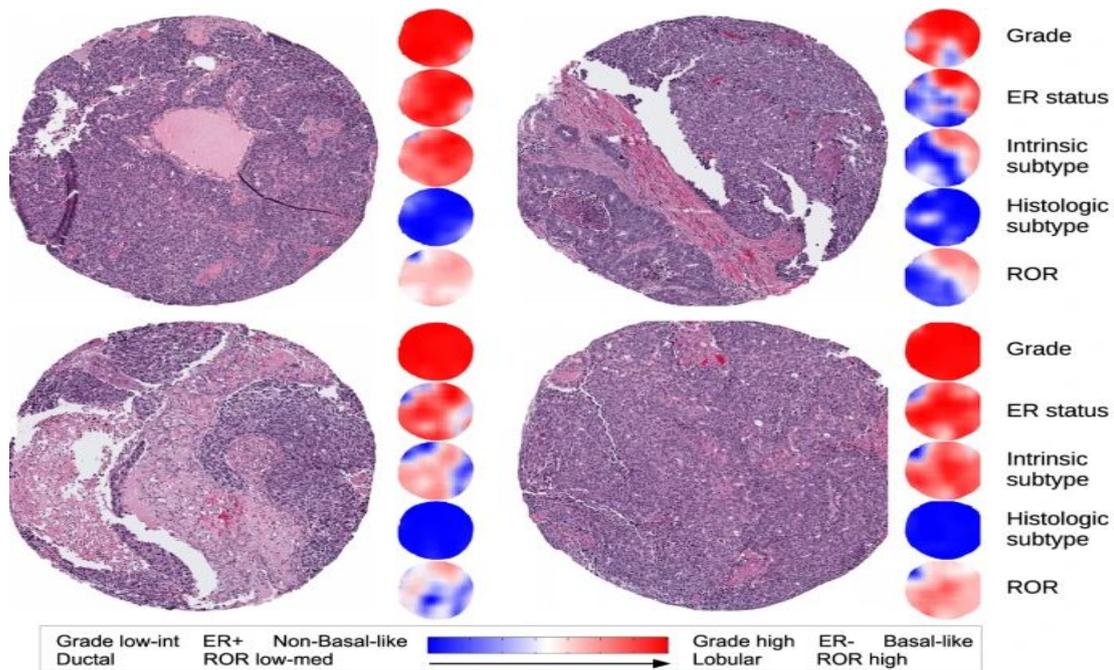
All the studies carried out include problems arising from security constraints in deep learning algorithms, misleading the prediction mechanism of the algorithm or stealing sensitive content about the algorithm. This adverse situation can be turned into an advantage in making sense of biological data that is difficult to understand and analyze.

In this study, transforming the gene expression data into images in RGB format without any loss on the data and training it with the CNN algorithm provides an algorithm that can be the target of these attacks. The one-pixel attack method developed by Su et al. [149], when applied to an image containing completely gene expression data and causes a change in the decision mechanism, raises the possibility of giving information about the effect of the gene contained in the relevant pixel on cancer. The attack mechanism, which has a high disadvantage for the generally used deep learning algorithms, can become advantageous when evaluated on biological data.

### **3.3 Deep Learning Approach for Cancer Diagnosis**

The increase in studies on neural networks and the fact that deep learning techniques produce meaningful results have enabled the data produced in the field of health to be interpreted by using deep learning methods. It is aimed to reduce the workload by automating many diagnostic processes that are time-consuming for clinicians, especially in the field of diagnosis of diseases. For this purpose, CT and histopathology data of cancer patients are used to facilitate diagnosis. In these diagnostic approaches, which are generally based on image processing, convolutional neural networks are preferred in the processing of image data. By using CNN, the diagnosis stage of many cancer types such as breast cancer [150–153], prostate cancer [154–158], lung cancer [159–161], head and neck cancer [162], skin cancer [1, 163] is performed.

CNN algorithm developed by Couture et al. (2018) [153]; makes predictions about breast cancer stage of patients, ER-positive/HER-2 negative results determined with the help of genomic tests, basal-like or non-basal-like status using breast cancer histopathology data.



**Figure 3.13** Estimation of cancer stage and molecular properties from the H&E image of breast cancer [153]

Figure 3.13 shows the prediction obtained by CNN algorithm on a section of tissue from a breast cancer patient stained with hematoxylin and eosin (H&E) dye. Based on the results obtained by staining the cancerous tissue, the analysis of the situation examined by many different tests such as the cancer stage of the person, the ER status, which is one of the molecular genetic features of the cancer, and the classification of tumor-specific conditions (such as ductal, basal) were performed. In addition to being able to perform many analyzes, it also provides the detection of heterogeneity, which makes the diagnosis of cancer difficult. Figure 3.13 is H&E stained images of four different nuclei from the same patient. According to the estimation results, three of these four different images are ER negative and basal-like intrinsic subtype, while the fourth one is predicted as ER negative non-basal-like, which also indicates intra-tumor heterogeneity between nuclei. This and similar studies not only enable clinicians to overcome time-consuming situations they encounter during diagnosis, but also make it easier to understand the diagnosis stage and biology of cancer by making inferences from images.

In addition to image-based approaches, biological data is also used extensively in cancer diagnosis [Jiao2020, 164] and even treatment [165, 166]. Gene expression data and deep learning approaches are integrated to tackle various challenges such as estimation of survival times of individuals with cancer [167], determination of biomarker genes [168], determination of effective therapeutics for cancer treatment [169], classification of cancer subtypes [170–172]. Ahn et al. developed a deep

learning algorithm using publicly available gene expression databases to classify the samples as normal or tumor and high predictive scores were obtained.



### 4.1 Datasets

Raw expression data was retrieved from UCSC Toil [173] which integrates The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) projects [6]. TCGA contains gene expression data for thousands of cancer patients and GTEx harbors gene expression data for normal tissues. Differential gene expression information for matching TCGA signatures of differentially expressed genes for tumors were collected from Harmonizome knowledge base [174].

### 4.2 Selection of genes with highest DEG cases

Harmonizome data was processed to count DEG cases throughout the TCGA dataset for each gene and then top 1024 genes were selected to be used in subsequent steps. Expression data for selected 1024 genes were extracted from Toil dataset. Also, sample labels (Tumor vs Normal) were extracted using phenotype data. Finally, a gene expression matrix of 17,884 samples X 1024 genes were constructed.

### 4.3 Conversion of gene expression data into image

Gene expression matrix for 17,884 samples were split into Train (14,308 samples) and Test (3,578 samples) with 80:20 ratio. Using Numpy [175], datasets were reshaped into multidimensional array suitable for training with Tensorflow [176]. During reshaping, gene expression values were mapped to RGB (red, green, blue) space. Conversion was done by converting the gene expression value into 24 bit long binary and then using first 8 bits for R (red), second 8 bits for G (green) and third 8 bits for B (blue). Figure 4.1 summarizes the concept with an sample conversion.

After the conversion, 14,308 x 1,024 training data becomes [14,308 x 32 x 32 x 3] Numpy array. Accordingly, test data becomes [3,578 x 32 x 32 x 3] Numpy array

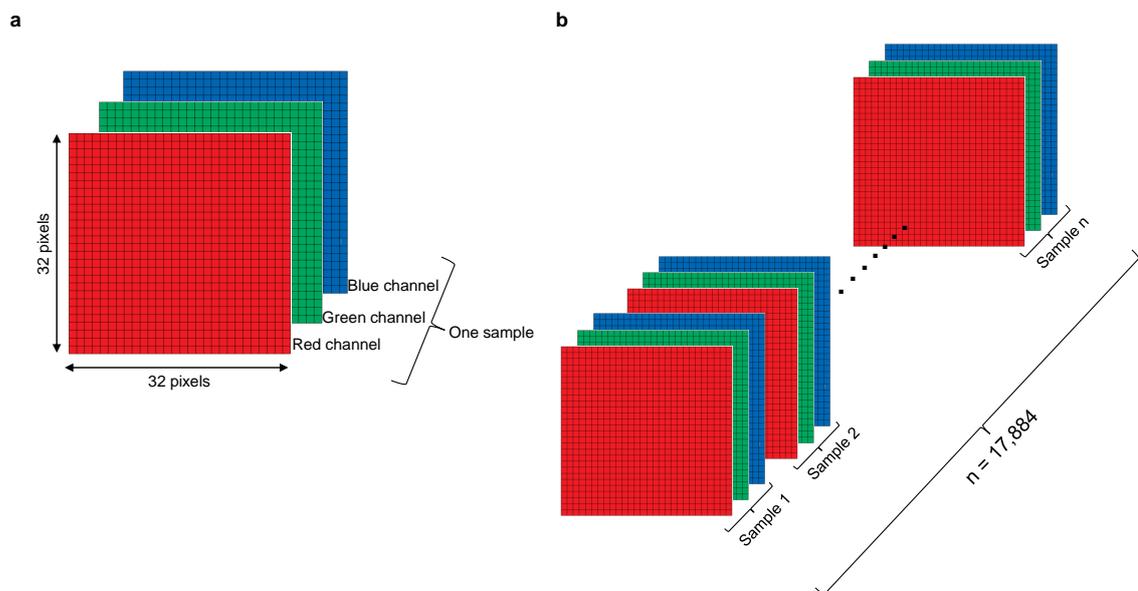
	Gene	Expression Level	RGB Value
	ENSG00000129250	131040	(1,255,224)

	R								G								B									
2-bit system	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Decimal	1								255								224									
Pixel color	#01FFE0 (1,255,224)																									

**Figure 4.1** Illustration of mapping gene expression values to RGB space. Decimal gene expression value is converted to 24 bits and then 8bits from beginning, middle and end are converted back to decimal values, resulting in three integers each smaller than 255. The color of the bottom row corresponds to mixture of individual R,G,B values. #01FFE0 is hexadecimal representation of RGB(1,255,224).

suitable for batch processing by Tensorflow. Final layer represents R,G,B values, the square 32x32 shape corresponds to 1,024 genes.



**Figure 4.2** Illustration of Numpy 4D arrays. (a) For each sample, 1024 genes are shaped as 32x32 pixels. Due to RGB mapping, 3 layers of color channels were used per sample. (b) This shape was imposed on whole dataset

## 4.4 Model Architecture used for Deep Learning Training

The convolutional neural network used for training contained 9 Convolution layers having dropout layers in between and ReLU as activation function. Figure 4.3 summarizes the model visually and Table 4.1 lists each layer of the model.

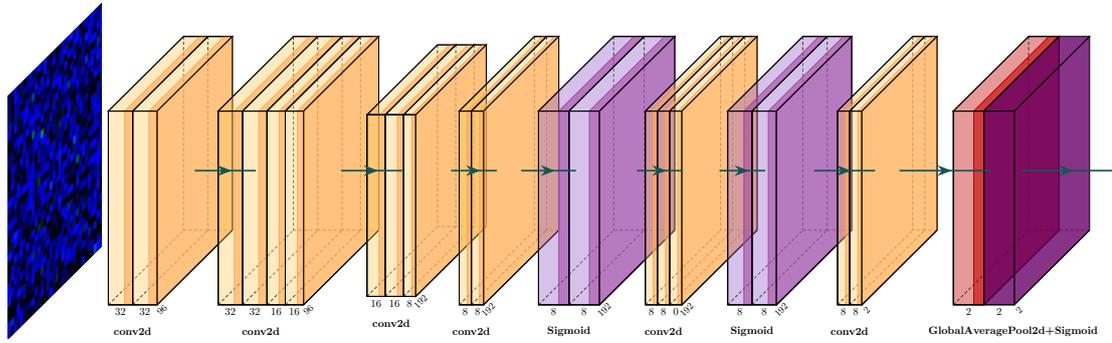


Figure 4.3 CNN Architecture

Table 4.1 Model architecture

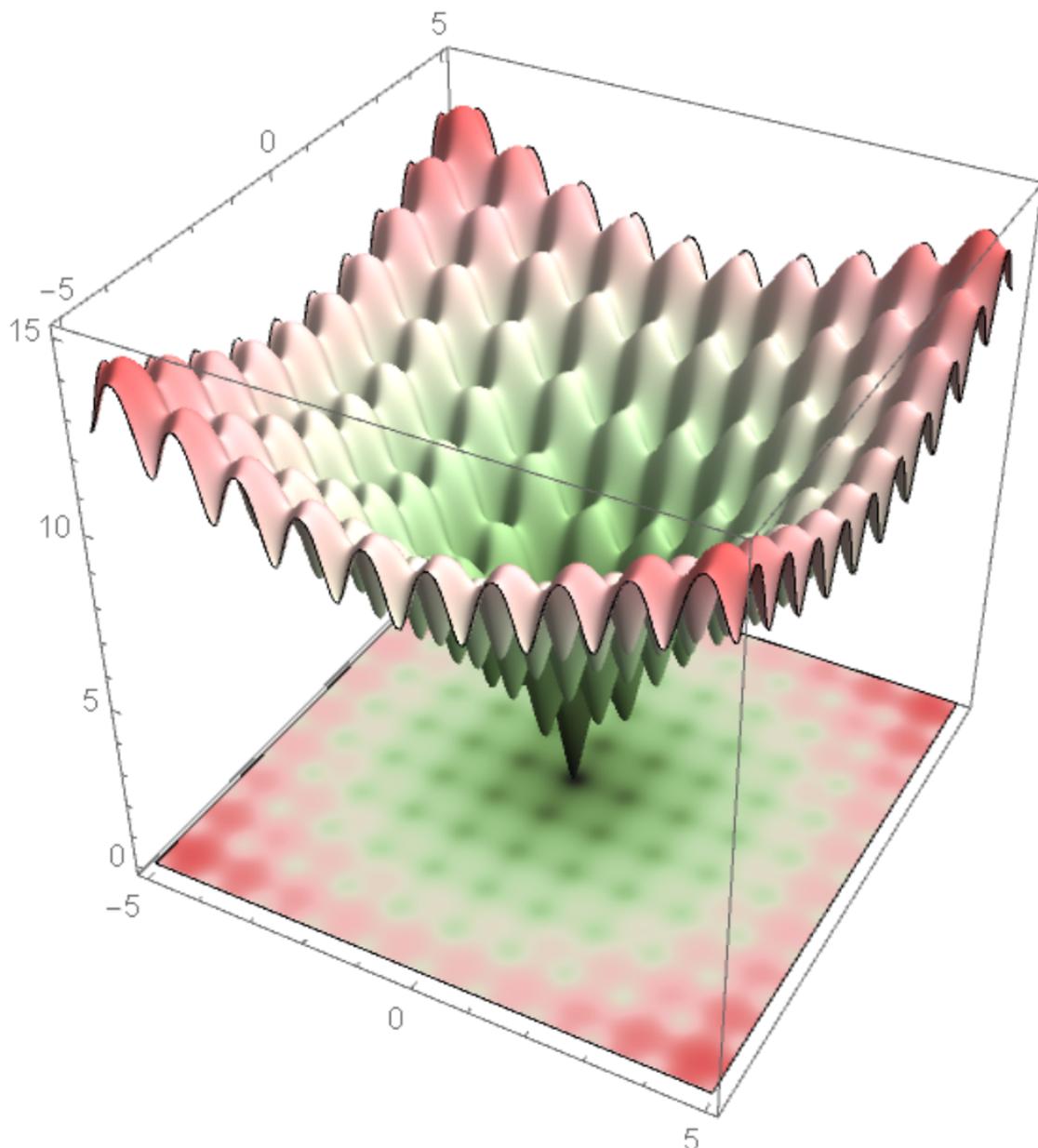
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 32, 32, 96)	2,688
dropout (Dropout)	(None, 32, 32, 96)	0
conv2d_1 (Conv2D)	(None, 32, 32, 96)	83,040
conv2d_2 (Conv2D)	(None, 16, 16, 96)	83,040
dropout_1 (Dropout)	(None, 16, 16, 96)	0
conv2d_3 (Conv2D)	(None, 16, 16, 192)	166,080
conv2d_4 (Conv2D)	(None, 16, 16, 192)	331,968
conv2d_5 (Conv2D)	(None, 8, 8, 192)	331,968
dropout_2 (Dropout)	(None, 8, 8, 192)	0
conv2d_6 (Conv2D)	(None, 8, 8, 192)	331,968
activation (Activation)	(None, 8, 8, 192)	0
dropout_3 (Dropout)	(None, 8, 8, 192)	0
conv2d_7 (Conv2D)	(None, 8, 8, 192)	37,056
activation_1 (Activation)	(None, 8, 8, 192)	0
dropout_4 (Dropout)	(None, 8, 8, 192)	0
conv2d_8 (Conv2D)	(None, 8, 8, 2)	386
global_average_pooling2d	(None, 2)	0
activation_2 (Activation)	(None, 2)	0
Total params: 1,368,194		

## 4.5 One pixel attack

One pixel attack algorithm was adopted from an earlier study [177] which utilizes "differential evolution" algorithm from SciPy Python library and illustrated in Figure

4.4. The attack algorithm picks random locations  $(x,y)$  where  $x < 32$  and  $y < 32$  and random RGB colors. Although blue and green values are picked within  $(0,255)$  range, the red color was only picked within  $(0,2)$  range since gene expression values are mostly below 196,607 corresponding to  $(2,255,255)$  RGB value.

One pixel attack provides pixel location, the new color value which causes label to change in trained model (from Normal to Tumor or vice versa). Since attack is random, we performed many attacks (10 times to be exact) to the test dataset. The resulting attacks were filtered if the suggested pixel value is within lowest and highest expression range of corresponding gene.

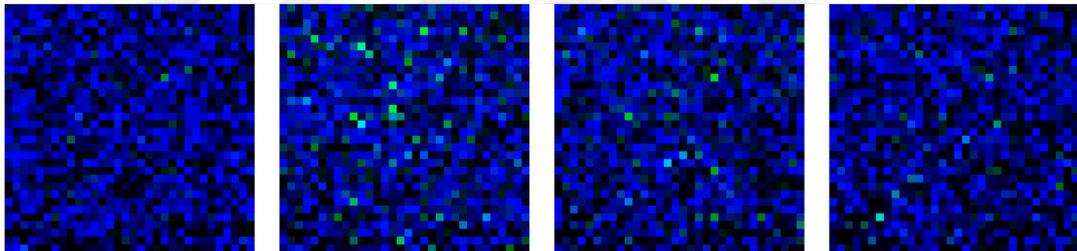


**Figure 4.4** Illustration of Differential Evolution algorithm over Ackley function.  
Taken from [178]

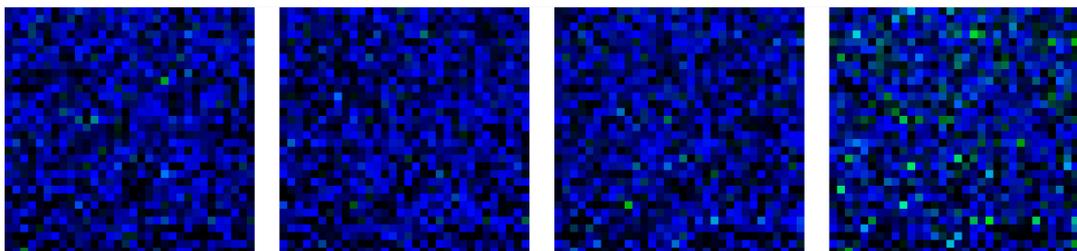
### 5.1 Gene expression data as image

After conversion of gene expression data by RGB mapping, each sample was converted into 32 x 32 2D array corresponding to 1024 genes. For each location in 2D array, three values, [R,G,B] were available thus resulting in color image representation of each sample. In Figure 5.1, sample images for Normal and Tumor samples were shown.

**a**



**b**



**Figure 5.1** Color image representation of samples for 1024 genes. (a) shows gene expression values of normal tissue samples converted to images, (b) shows gene expression values of tumor tissue samples converted to images.

### 5.2 Model Training

Training data consisted of gene expression data for 14,308 samples and 1,024 genes. training data becomes [14,308 x 32 x 32 x 3] Numpy array after RGB

mapping. Accordingly, test data becomes [3,578 x 32 x 32 x 3] Numpy array. These multidimensional arrays were used for training.

After 40 epochs the training reached %97.7 accuracy. Loss and Accuracy plots for the training process are shown in Figure 5.2. The plots indicate there's no over-fitting.

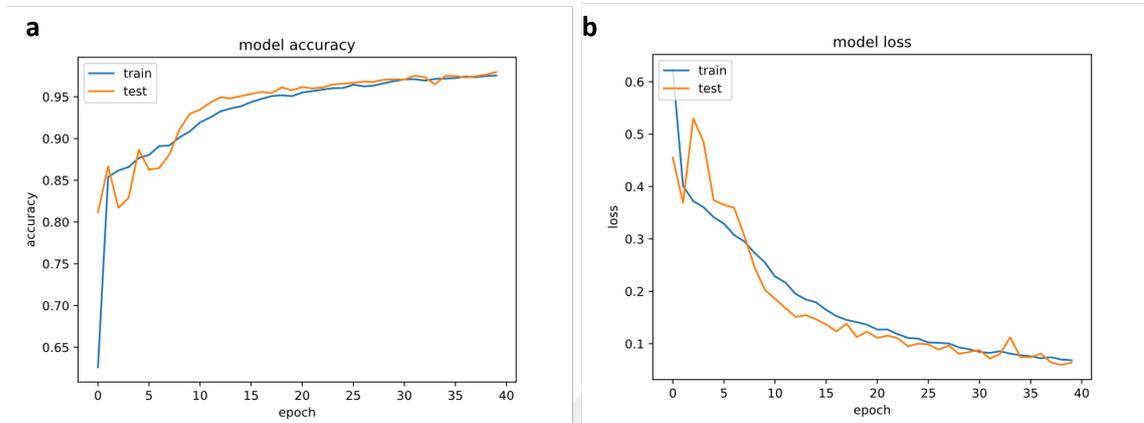


Figure 5.2 Accuracy and Loss plots of the trained model after 40 epochs.

### 5.3 Performance Measurement

Using the test dataset the performance of the model was evaluated. The evaluation of the model was summarized in Table 5.1. In Figure 5.3, the ROC curve indicates

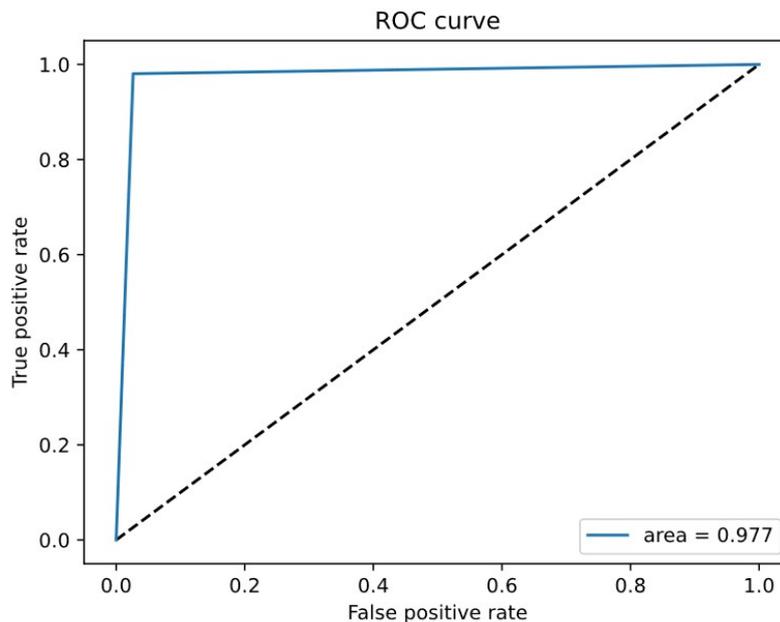


Figure 5.3 The ROC curve of designed CNN for tumor normal classification.

There are several different approaches which uses gene expression data to classify tumor and normal samples ranging from simpler machine learning approaches to

**Table 5.1** Confusion Matrix

<b>Summary of Statistics</b>				
	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
<b>Normal</b>	0.98	0.97	0.98	1628
<b>Tumor</b>	0.98	0.98	0.98	1949
<b>Accuracy</b>			0.98	3577

**Table 5.2** Comparison model with other studies. SVM; support vector machine, t-SNE; t-distributed stochastic neighbor embedding.

Expression preprocessing	Classification Method	Accuracy	Sensitivity	Specificity	Precision	F-measure	Resource
<i>RGB mapping</i>	CNN	97,73%	97,66%	97,80%	98,00%	0,975	<i>Our method</i>
Normalization	CNN	98,76%	91,43%	100,00%	100,00%	0,955	Elbashir et al. [179]
Normalization	Stacked Denoising Autoencoder (SDAE)	94,78%	94,04%	97,50%	97,20%		Danaee et al. [180]
Normalization	AlexNet	96,69%	96,89%	94,12%	99,54%	0,955	Elbashir et al [179]
t-SNE	SVM	95,87%	100,00%	51,00%	95,96%	0,97	Elbashir et al [179]

complex deep learning networks. These approaches usually start with pre-processing the gene expression data with an irreversible manipulation and even mapping data points to a different domain. Our method involves minimal and reversible change to gene expression data. The RGB mapping is reversible and does not require normalization or any dimensional reduction techniques. Table 5.2 compares our approach with several different approaches both in pre-processing and classification steps. Although Elbashir et al [179] study (Normalization + CNN) has highest accuracy, our approach has better results in overall. Please note that Elbashir et al uses smaller and unbalanced TCGA dataset. Their accuracy starts from 91% and reaches 98.7% and due to dominating number of tumor samples, their model has tendency to pick “tumor” as label, explaining their lowest sensitivity and full precision. In our case, our dataset is balanced (8156 Normal vs. 9750 Tumor) and our accuracy start from 58% and reaches 97.7%.

## 5.4 One-pixel Attack Results

As a result of the attacks, label changes took place in 240 different samples. These results are shared in Appendix A, along with sample ID, ensemble ID, expression values before and after the attack, and prediction values before and after the attack. When the label changes as a result of the attacks were analyzed, it was seen that the changes on 13 genes caused changes in the decision mechanism of the neural network. The 13 identified genes are shown in Table 5.3 along with their names and Ensemble IDs.

**Table 5.3** Identified genes resulted from one-pixel attack

<b>Gene Name</b>	<b>Ensembl ID</b>
TGFBR2 TGF-beta receptor type-2	ENSG00000163513
KIF1C Kinesin-like protein KIF1C	ENSG00000129250
DDX3X ATP-dependent RNA helicase DDX3X	ENSG00000215301
AGRN Agrin	ENSG00000188157
SLC39A8 Metal cation symporter ZIP8	ENSG00000138821
AHNAK Neuroblast differentiation-associated protein AHNAK	ENSG00000124942
ETS2 Protein C-ets-2	ENSG00000157557
CAVIN1 Caveolae-associated protein 1	ENSG00000177469
BHLHE41 Class E basic helix-loop-helix protein 41	ENSG00000123095
TSC22D3 TSC22 domain family protein 3	ENSG00000157514
CAT Catalase	ENSG00000121691
NCF2 Neutrophil cytosol factor 2	ENSG00000116701
SREBF2 Sterol regulatory element-binding protein 2	ENSG00000198911

According to the results obtained, the label change (from cancer to normal) in the attack result of 26 samples was obtained by decreasing the expression value, while

the expression value of 210 samples was obtained by increasing it. Cancer and stage data of 26 samples whose expression has decreased and whose label has changed are shown in Table 5.4.

**Table 5.4** Cancers and stage information whose labels are changed by decreasing expression values

Cancer Type	Stage	Sample Number
<i>Lung Adenocarcinoma</i>	stage ib	3
	stage iiiia	2
	stage ia	1
<i>Lung Squamous Cell Carcinoma</i>	stage ia	2
<i>Kidney Renal Clear Cell Carcinoma</i>	stage iii	1
	stage i	1
<i>Kidney Renal Papillary Cell Carcinoma</i>	stage i	4
<i>Head and Neck Squamous Cell Carcinoma</i>	stage ii	2
	stage iva	1
<i>Thyroid Carcinoma</i>	stage iii	1
<i>Bladder Urothelial Carcinoma</i>	stage iv	1
<i>Mesothelioma</i>	stage ii	1
<i>Stomach Adenocarcinoma</i>	stage ii	1
<i>Breast Invasive Carcinoma</i>	stage iib	1
<i>Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma</i>	not reported	1
<i>Esophageal Carcinoma</i>	not reported	1
<i>Ovarian Serous Cystadenocarcinoma</i>	not reported	1

This situation, which changes in the label with the decrease of the expression value in the prediction data of 26 different samples is controlled by the same gene. **Agrin (ENSG00000188157)** was found to be the gene causing the change as a result of the attack in all 26 samples. It has been observed that the first prediction of the **Agrin** gene is tumour, not only when the expression value is decreased but also when the expression value is increased, then it predicts normal.

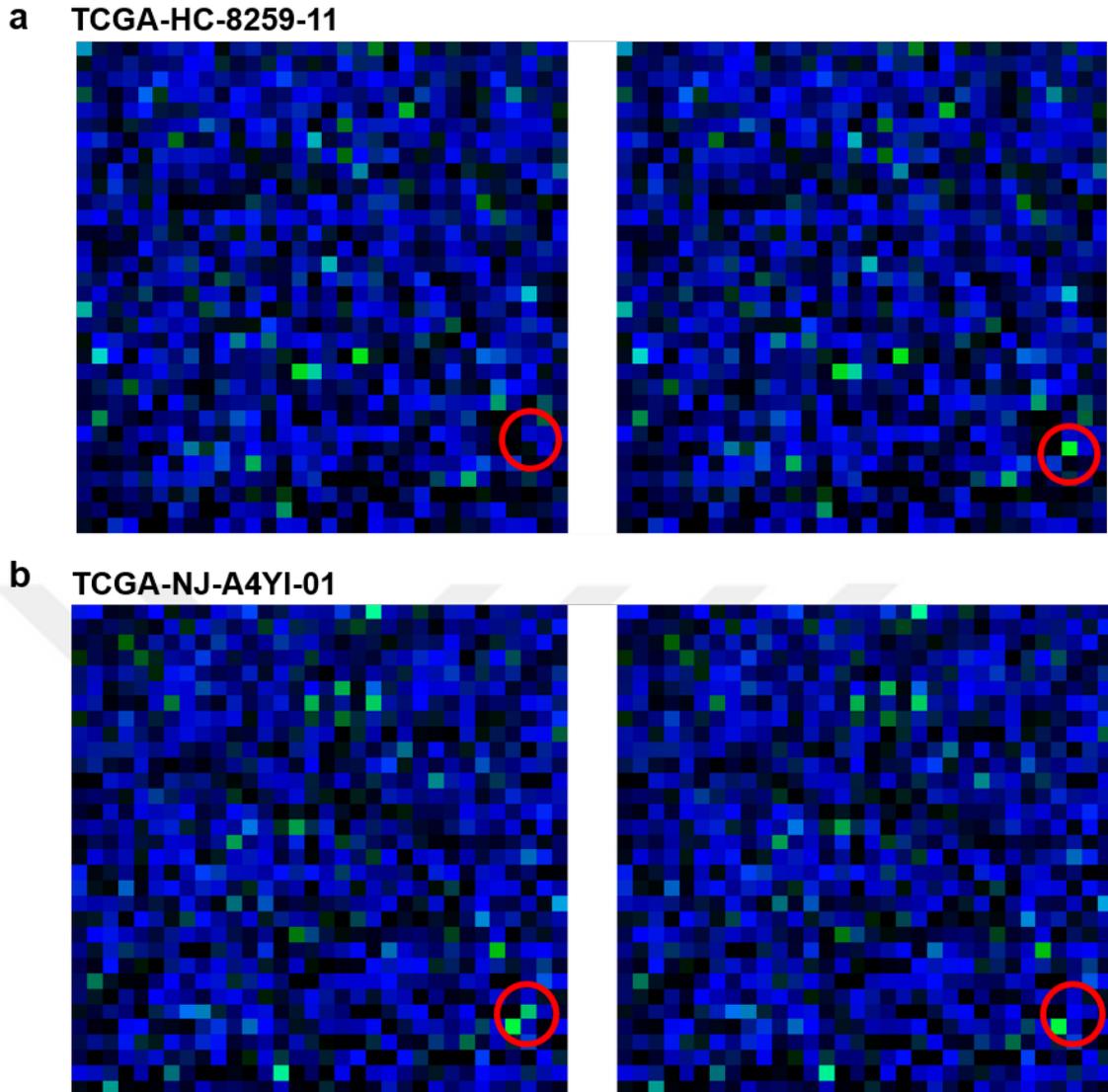
While the initial estimate of all data was generally tumour, it was labeled as normal as a result of the attack. While only 4 of the 240 results were labeled as normal, it was observed that they turned into tumour as a result of the attack, and it was observed that the gene causing this change was **Agrin**. The cancer and stage distribution of these 4 samples are shown in Table 5.5.

Images of a sample where the **Agrin** gene returned from normal to tumour by increasing its expression level in two different samples (Figure 5.4.a) and decreased gene expression level from cancer to normal (Figure 5.4.b) shown in Figure 5.4. The

**Table 5.5** Cancer type and stage distribution of samples that turned into cancer as a result of an attack when they were normal

<b>Cancer Type</b>	<b>Stage</b>	<b>Sample Number</b>
<i>Prostate Adenocarcinoma</i>	not reported	2
<i>Thyroid Carcinoma</i>	stage i	1
<i>Breast Invasive Carcinoma</i>	stage iia	1

example shown in Figure 5.4.a illustrates the gene expression data of a sample taken from the healthy tissue of a patient with has Prostate Adenocarcinome cancer type. While the gene expression level of the **Agrin** gene was 3,342 in the tissue that did not have this cancer type, the gene expression level was increased to 128,550 as a result of the attack and it was marked as tumour. The example shown in Figure 5.4.b is an image created from the gene expression data of a tissue with Lung Adenocarcinoma cancer type. **Agrin** gene expression value in tissue with this cancer type is 49,770. When this value was changed to 246 as a result of the attack, it was seen that it was labeled as normal.



**Figure 5.4** Sample images obtained as a result of the attack. The first images show the original images and the second images show the images obtained as a result of the attack. a. Original and post attack images of gene expression data with sample ID TCGA-HC-8259-11. The areas marked with a red circle show the changes in the Agrin gene as a result of the attack. The gene expression value has been increased for the TCGA-HC-8259 sample. As can be seen in the image, brighter pixels were obtained by increasing the expression value. b. Original and post attack images of gene expression data with sample ID TCGA-NJ-A4YI-01. The areas marked with a red circle show the changes in the Agrin gene as a result of the attack. The gene expression value for the TCGA-NJ-A4YI-01 sample has been reduced. As can be seen in the image, while the expression value of the pixel is brighter in the original image, it appears darker in the image resulting from the attack.

## 5.5 Annotation of Attack Results Based on Real Cancer Patient Expression Levels

According to the data obtained in Section 5.4, Agrin (ENSG00000188157) was found to be the most effective gene on attack results. Looking at the literature, it was seen that the first study on the relationship between Agrin and cancer was carried out by Rajkumar et al. [181] in 2011 to determine the genes that are effective on cervical tumors. Other studies have shown that it has an effect on prognosis due to changes in Agrin gene expression values [182]. The graph of expression data values of Agrin, which is one of the most common genes as a result of an attack, for all cancer types is shown in Figure 5.5.

As can be seen in Figure 5.5, as the expression values of the Agrin gene increase, the probability of tumor tissue increases. This explains the prediction of tumor tissue when the expression values of the Agrin gene are increased in normal tissue, and the prediction of normal tissue when it is decreased in tumor tissue (Figure 5.4). Statistical data of cancer types belonging to Agrin gene are shown in Appendix B, Table B. When that table is examined, the log<sub>2</sub> value of the tumor-normal ratio of Agrin varies. This may indicate that the probability of being found with DEG is high.

One of the genes most likely to be found with DEG as a result of an attack is the Catalase protein (ENSG00000121691) gene. When a literature search is conducted on the relationship between catalase gene and cancer, it is seen that the first study was carried out in 1950 by Appleman et al [183]. According to the study, they observed that catalase activity decreased in tumor cells. Many subsequent studies have shown that catalase is an effective gene for cancer [184, 185].

When the gene expression values of the Catalase gene are examined according to the cancer types (Figure 5.6), it is seen that there is no clear pattern in the cancer-specific change as in Agrin. Statistical data are shared in Appendix B.2.

Kinesin-like Protein (ENSG00000129250) gene is the gene with the least data on the relationship between the attack outcome and cancer in the literature. There are 2 studies on the relationship between this gene and cancer in the Pubmed database. The first of these was carried out by Shah et al. [186] in 2009. The aim of the study by Shah et al. was to sequence the lobular breast cancer genome. For this purpose, the frequency of somatic mutations found in the primary tumor tissue of the same patient 9 years ago was measured and accordingly they observed that the KIF1C gene was at low frequencies. Another study conducted by Zou et al. in 2014 is investigating the effect of Kinesin family proteins on breast cancer. According to their findings, they showed that the KIF1C gene was suppressed.

Considering the change in gene expression values of the KIF1C gene according to cancer types (Figure 5.7), it is seen that the gene expression value is higher in some normal tissues compared to tumor tissue depending on the cancer type, while the opposite is the case in some tissues. When the statistical data according to cancer types in Appendix B.3 are examined, it is seen that there is no significant change in the log<sub>2</sub> value. This shows that it is a difficult gene to find with DEG.



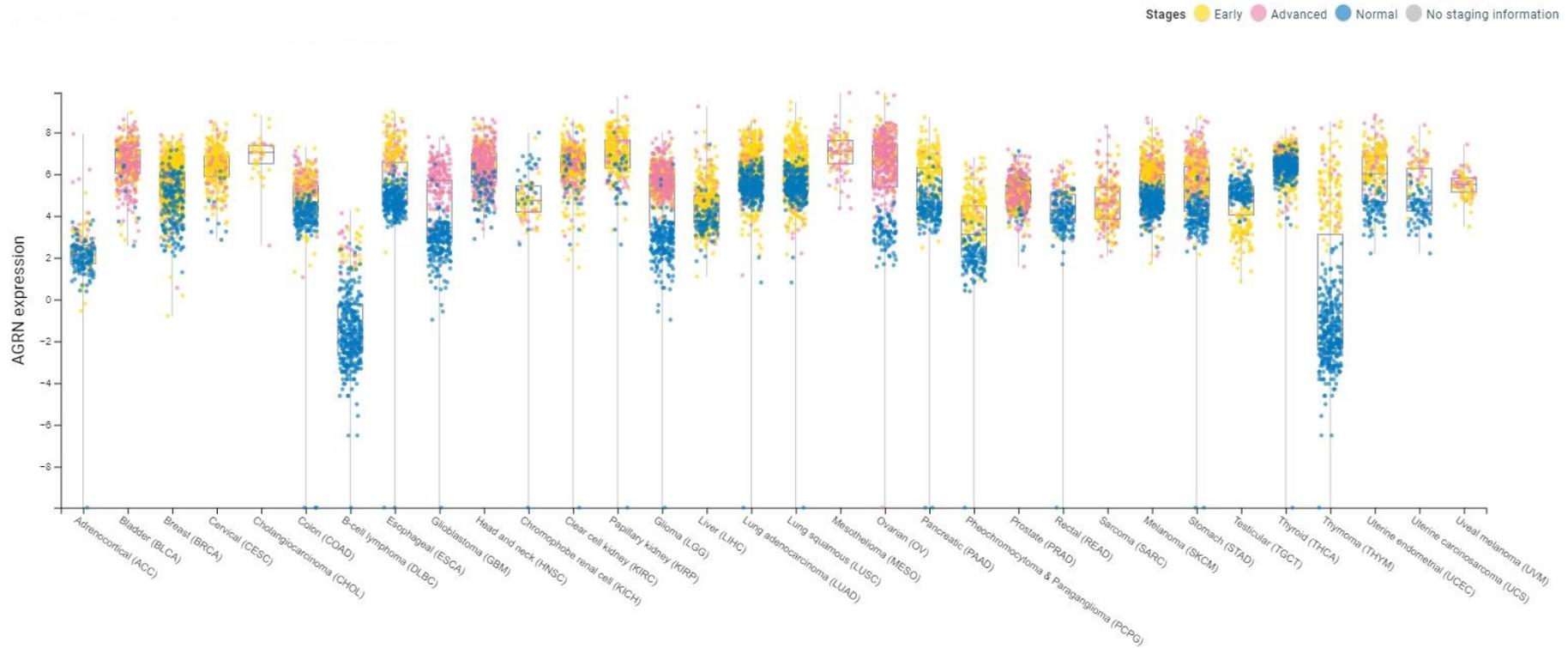
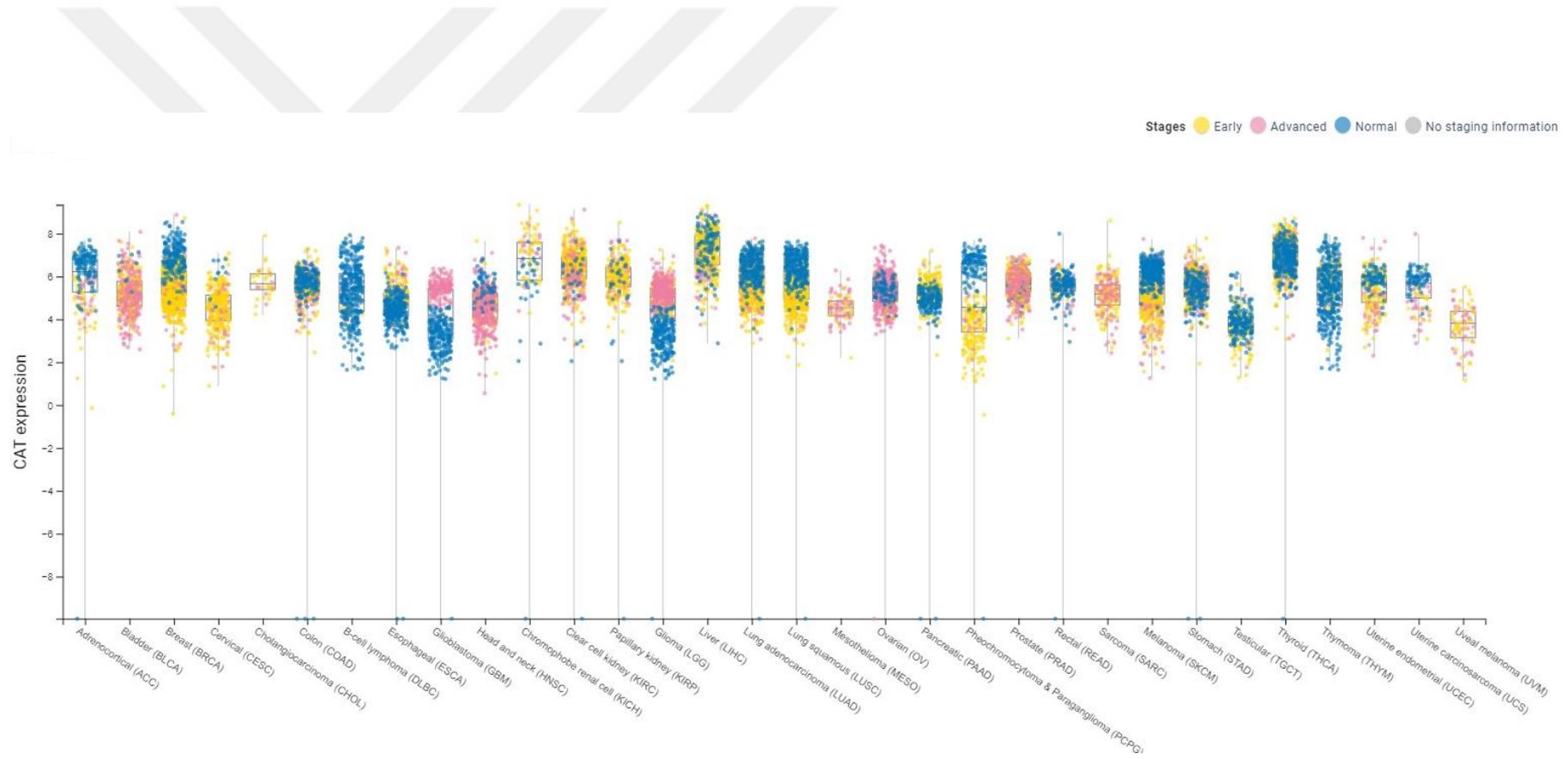
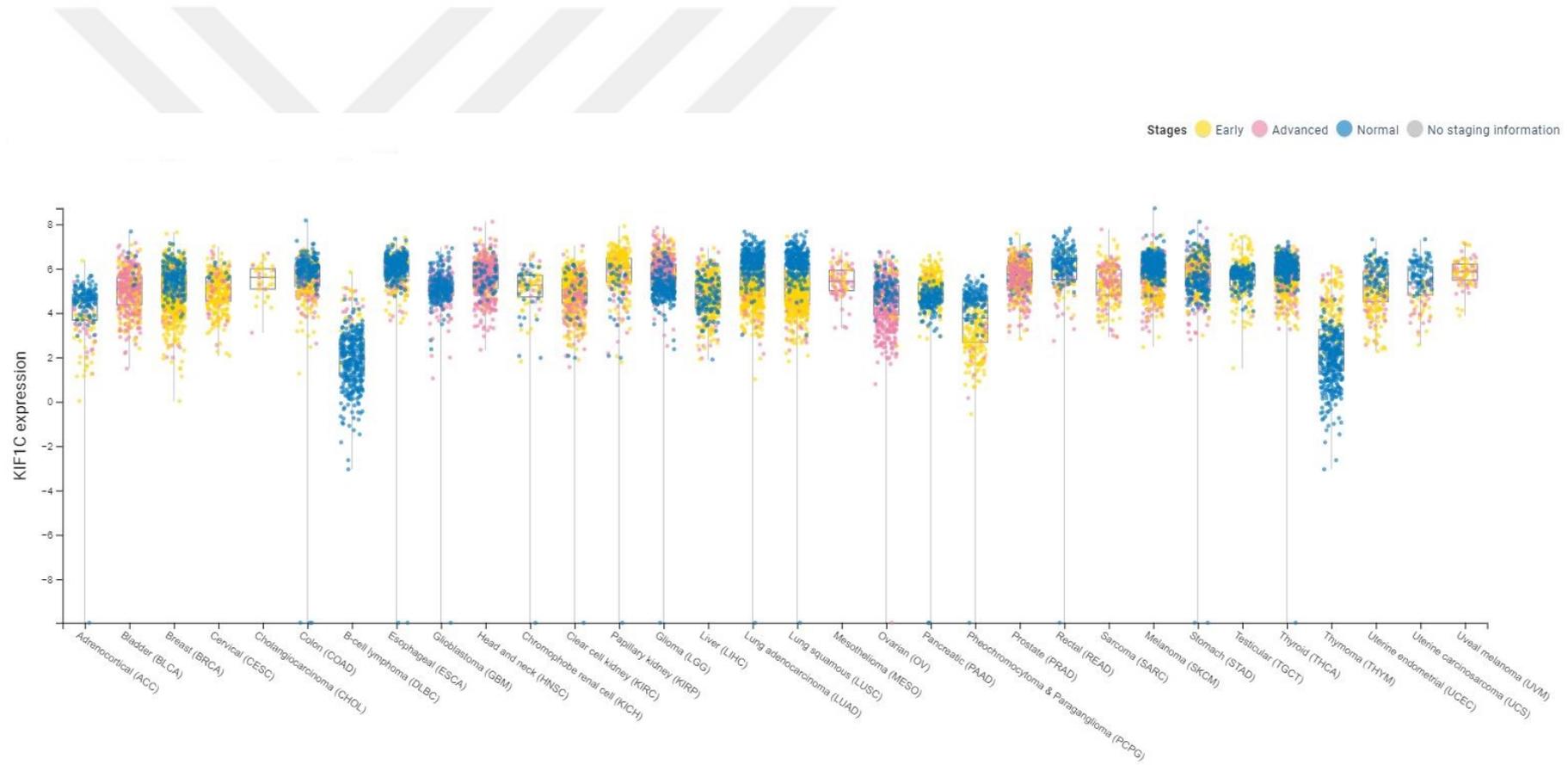


Figure 5.5 Gene expression values of Agrin gene according to cancer types [187].



**Figure 5.6** Gene expression values of Catalase gene according to cancer types [187].



**Figure 5.7** Gene expression values of Kinesin-like protein gene according to cancer types [187].

In this study, a neural network capable of distinguishing between normal and tumor tissue was developed with the convolutional neural network model. It is aimed to eliminate the difficulty of the diagnosis stage caused by the plasticity and heterogeneity of tumor cells. In order to achieve this aim, RNA sequencing method obtained directly from the genome of tumor biology was used as input for the neural network. The most important difference of the created model from the existing models is that did not apply any normalisation methods was performed on the RNA-seq data. Existing models can perform normalization or various embedding process on RNA-seq data, causing information loss on the data. In order to avoid this loss of information in the created model, RNA-seq data was preserved as it was, and converted to 24-bit RGB formats, and a 32x32 image was created for each sample in which these RGB values were arranged. These obtained images are not trained as images. The training of the model was carried out with 32x32x3 3D multidimensional arrays for each sample. As a result of the training, a high accuracy rate was obtained from the training dataset, which has a homogeneous data distribution.

The image samples of the data converted to RGB format for the input of the model are indistinguishable from the tumor or normal sample when viewed with the human eye. However, after the training, the model's ability to distinguish these data with a high accuracy rate, and the neural network model created due to the difference between these images, has inspired the idea that it can be a guide in the determination of genes that may be effective in defining the tumor.

Based on this hypothesis, the pixels that changed the prediction of the model were determined by applying the one-pixel attack method to the tumor and normal sample images obtained as input. Since each pixel value represents a gene, this information includes gene information that may be associated with cancer. All samples were attacked approximately 10 times using the one-pixel attack method. As a result of the attacks, the attack value was filtered according to the highest gene expression value of a gene, and as a result of these attacks, 240 attack results were obtained

that were effective within the specified gene expression value range (See Appendix A). By analysis of 240 attack results, 13 unique genes were found that were effective in varying the predictive value. When general screening of these genes is performed, in addition to the presence of known genes that use as biomarkers for cancer, some rarely studied genes (such as KIF1C) have been found.

All these findings have shown that it can bring a new approach to diseases such as cancer, which are complex to diagnose and treat, and which require more biomarker genes for diagnosis. With the application of this method, individual results can be obtained. Inter- and intra-tumor heterogeneity characteristics of tumor cells can be determined. It can be used as an approach that makes it possible to make individual cancer analysis by making it easier to find genes that differ from person to person. The results obtained can be strengthened with experimental data to identify new biomarkers for cancer.

## REFERENCES

---

- [1] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, J. Dean, “A guide to deep learning in healthcare,” *Nature Medicine*, vol. 25, no. 1, pp. 24–29, Jan. 2019. DOI: 10.1038/s41591-018-0316-z.
- [2] E. Persi, Y. I. Wolf, D. Horn, E. Ruppin, F. Demichelis, R. A. Gatenby, R. J. Gillies, E. V. Koonin, “Mutation–selection balance and compensatory mechanisms in tumour evolution,” *Nature Reviews Genetics*, vol. 22, no. 4, pp. 251–262, Nov. 2020. DOI: 10.1038/s41576-020-00299-4.
- [3] R. K. Sevakula, V. Singh, N. K. Verma, C. Kumar, Y. Cui, “Transfer learning for molecular cancer classification using deep neural networks,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 6, pp. 2089–2100, Nov. 2019. DOI: 10.1109/tcbb.2018.2822803. [Online]. Available: <https://doi.org/10.1109/tcbb.2018.2822803>.
- [4] F. Gao, W. Wang, M. Tan, L. Zhu, Y. Zhang, E. Fessler, L. Vermeulen, X. Wang, “DeepCC: A novel deep learning-based framework for cancer molecular subtype classification,” *Oncogenesis*, vol. 8, no. 9, Aug. 2019. DOI: 10.1038/s41389-019-0157-8. [Online]. Available: <https://doi.org/10.1038/s41389-019-0157-8>.
- [5] T. Matsubara, T. Ochiai, M. Hayashida, T. Akutsu, J. C. Nacher, “Convolutional neural network approach to lung cancer classification integrating protein interaction network and gene expression profiles,” *Journal of Bioinformatics and Computational Biology*, vol. 17, no. 03, p. 1940007, Jun. 2019. DOI: 10.1142/s0219720019400079. [Online]. Available: <https://doi.org/10.1142/s0219720019400079>.
- [6] K. G. Ardlie, D. S. Deluca, A. V. Segre, T. J. Sullivan, T. R. Young, E. T. Gelfand, C. A. Trowbridge, J. B. Maller, T. Tukiainen, M. Lek, L. D. Ward, P. Kheradpour, B. Iriarte, Y. Meng, C. D. Palmer, T. Esko, W. Winckler, J. N. Hirschhorn, M. Kellis, D. G. MacArthur, G. Getz, A. A. Shabalina, G. Li, Y.-H. Zhou, A. B. Nobel, I. Rusyn, F. A. Wright, T. Lappalainen, P. G. Ferreira, H. Ongen, M. A. Rivas, A. Battle, S. Mostafavi, J. Monlong, M. Sammeth, M. Mele, F. Reverter, J. M. Goldmann, D. Koller, R. Guigo, M. I. McCarthy, E. T. Dermitzakis, E. R. Gamazon, H. K. Im, A. Konkashbaev, D. L. Nicolae, N. J. Cox, T. Flutre, X. Wen, M. Stephens, J. K. Pritchard, Z. Tu, B. Zhang, T. Huang, Q. Long, L. Lin, J. Yang, J. Zhu, J. Liu, A. Brown, B. Mestichelli, D. Tidwell, E. Lo, M. Salvatore, S. Shad, J. A. Thomas, J. T. Lonsdale, M. T. Moser, B. M. Gillard, E. Karasik, K. Ramsey, C. Choi, B. A. Foster, J. Syron, J. Fleming, H. Magazine, R. Hasz, G. D. Walters, J. P. Bridge, M. Miklos, S. Sullivan, L. K. Barker, H. M. Traino, M. Mosavel, L. A. Siminoff, D. R. Valley, D. C. Rohrer, S. D. Jewell, P. A. Branton, L. H.

- Sobin, M. Barcus, L. Qi, J. McLean, P Hariharan, K. S. Um, S. Wu, D. Tabor, C. Shive, A. M. Smith, S. A. Buia, A. H. Undale, K. L. Robinson, N. Roche, K. M. Valentino, A. Britton, R. Burges, D. Bradbury, K. W. Hambright, J. Seleski, G. E. Korzeniewski, K. Erickson, Y. Marcus, J. Tejada, M. Taherian, C. Lu, M. Basile, D. C. Mash, S. Volpi, J. P. Struewing, G. F. Temple, J. Boyer, D. Colantuoni, R. Little, S. Koester, L. J. Carithers, H. M. Moore, P. Guan, C. Compton, S. J. Sawyer, J. P. Demchok, J. B. Vaught, C. A. Rabiner, N. C. Lockhart, K. G. Ardlie, G. Getz, F. A. Wright, M. Kellis, S. Volpi, E. T. Dermitzakis, "The genotype-tissue expression (GTEx) pilot analysis: Multitissue gene regulation in humans," *Science*, vol. 348, no. 6235, pp. 648–660, May 2015. DOI: 10.1126/science.1262110.
- [7] L. M. Coussens Z. Werb, "Inflammation and cancer," *Nature*, vol. 420, no. 6917, pp. 860–867, 2002.
- [8] D. G. DeNardo, P. Andreu, L. M. Coussens, "Interactions between lymphocytes and myeloid cells regulate pro-versus anti-tumor immunity," *Cancer and Metastasis Reviews*, vol. 29, no. 2, pp. 309–316, 2010.
- [9] S. I. Grivennikov, F. R. Greten, M. Karin, "Immunity, inflammation, and cancer," *Cell*, vol. 140, no. 6, pp. 883–899, 2010.
- [10] B.-Z. Qian J. W. Pollard, "Macrophage diversity enhances tumor progression and metastasis," *Cell*, vol. 141, no. 1, pp. 39–51, 2010.
- [11] F. Colotta, P. Allavena, A. Sica, C. Garlanda, A. Mantovani, "Cancer-related inflammation, the seventh hallmark of cancer: Links to genetic instability," *Carcinogenesis*, vol. 30, no. 7, pp. 1073–1081, 2009.
- [12] M. W. Klymkowsky P. Savagner, "Epithelial-mesenchymal transition: A cancer researcher's conceptual friend and foe," *The American journal of pathology*, vol. 174, no. 5, pp. 1588–1593, 2009.
- [13] J. P. Thiery, H. Acloque, R. Y. Huang, M. A. Nieto, "Epithelial-mesenchymal transitions in development and disease," *cell*, vol. 139, no. 5, pp. 871–890, 2009.
- [14] K. Polyak R. A. Weinberg, "Transitions between epithelial and mesenchymal states: Acquisition of malignant and stem cell traits," *Nature Reviews Cancer*, vol. 9, no. 4, pp. 265–273, 2009.
- [15] M. Yilmaz G. Christofori, "Emt, the cytoskeleton, and cancer cell invasion," *Cancer and Metastasis Reviews*, vol. 28, no. 1-2, pp. 15–33, 2009.
- [16] A. Barrallo-Gimeno M. A. Nieto, "The snail genes as inducers of cell movement and survival: Implications in development and cancer," *Development*, vol. 132, no. 14, pp. 3151–3161, 2005.
- [17] O. H. Warburg, *The metabolism of tumours: investigations from the Kaiser Wilhelm Institute for Biology, Berlin-Dahlem*. Constable & Company Limited, 1930.
- [18] O. Warburg, "On the origin of cancer cells," *Science*, vol. 123, no. 3191, pp. 309–314, 1956.
- [19] —, "On respiratory impairment in cancer cells," *Science*, vol. 124, pp. 269–270, 1956.

- [20] D. Hanahan R. A. Weinberg, “Hallmarks of cancer: The next generation,” *Cell*, vol. 144, no. 5, pp. 646–674, 2011.
- [21] N. M. Anderson M. C. Simon, “The tumor microenvironment,” *Current Biology*, vol. 30, no. 16, R921–R925, 2020.
- [22] D. C. Hinshaw L. A. Shevde, “The tumor microenvironment innately modulates cancer progression,” *Cancer Research*, vol. 79, no. 18, pp. 4557–4566, 2019.
- [23] C. A. Lyssiotis A. C. Kimmelman, “Metabolic interactions in the tumor microenvironment,” *Trends in Cell Biology*, vol. 27, no. 11, pp. 863–875, 2017.
- [24] immunovia. (2020). Current technologies for pancreatic cancer detection and diagnosis. (Visited on 2021-02-24), [Online]. Available: <https://immunovia.com/immray-pancan-d/current-technologies-for-pancreatic-cancer-detection-and-diagnosis/>.
- [25] N. McGranahan, R. A. Burrell, D. Endesfelder, M. R. Novelli, C. Swanton, “Cancer chromosomal instability: Therapeutic and diagnostic challenges: ‘exploring aneuploidy: The significance of chromosomal imbalance’ review series,” *EMBO Reports*, vol. 13, no. 6, pp. 528–538, 2012.
- [26] R. Fisher, L. Pusztai, C. Swanton, “Cancer heterogeneity: Implications for targeted therapeutics,” *British Journal of Cancer*, vol. 108, no. 3, pp. 479–485, 2013.
- [27] G. Driessens, B. Beck, A. Caauwe, B. D. Simons, C. Blanpain, “Defining the mode of tumour growth by clonal analysis,” *Nature*, vol. 488, no. 7412, pp. 527–530, 2012.
- [28] D. A. Lawson, K. Kessenbrock, R. T. Davis, N. Pervolarakis, Z. Werb, “Tumour heterogeneity and metastasis at single-cell resolution,” *Nature Cell Biology*, vol. 20, no. 12, pp. 1349–1360, 2018.
- [29] R. A. Burrell, N. McGranahan, J. Bartek, C. Swanton, “The causes and consequences of genetic heterogeneity in cancer evolution,” *Nature*, vol. 501, no. 7467, pp. 338–345, 2013.
- [30] A. H. Fischer, K. A. Jacobson, J. Rose, R. Zeller, “Hematoxylin and eosin staining of tissue and cell sections,” *Cold Spring Harbor Protocols*, vol. 2008, no. 5, pdb-prot4986, 2008.
- [31] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. N. Shih, J. Tomaszewski, F. A. González, A. Madabhushi, “Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent,” *Scientific Reports*, vol. 7, p. 46450, 2017.
- [32] Y. Liu, T. Kohlberger, M. Norouzi, G. E. Dahl, J. L. Smith, A. Mohtashamian, N. Olson, L. H. Peng, J. D. Hipp, M. C. Stumpe, “Artificial intelligence-based breast cancer nodal metastasis detection: Insights into the black box for pathologists,” *Archives of Pathology & Laboratory Medicine*, vol. 143, no. 7, pp. 859–868, 2019.

- [33] G. Litjens, P. Bandi, B. Ehteshami Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. van de Loo, R. Vogels, *et al.*, “1399 h&e-stained sentinel lymph node sections of breast cancer patients: The camelyon dataset,” *GigaScience*, vol. 7, no. 6, giy065, 2018.
- [34] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, *et al.*, “Detecting cancer metastases on gigapixel pathology images,” *arXiv preprint arXiv:1703.02442*, 2017.
- [35] M. S. Hosseini, L. Chan, G. Tse, M. Tang, J. Deng, S. Norouzi, C. Rowsell, K. N. Plataniotis, S. Damaskinos, “Atlas of digital pathology: A generalized hierarchical histological tissue type-annotated database for deep learning,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 747–11 756.
- [36] J. Folmsbee, X. Liu, M. Brandwein-Weber, S. Doyle, “Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, 2018, pp. 770–773.
- [37] M. Z. Alom, T. Aspiras, T. M. Taha, V. K. Asari, T. Bowen, D. Billiter, S. Arkell, “Advanced deep convolutional neural network approaches for digital pathology image analysis: A comprehensive evaluation with different use cases,” *arXiv preprint arXiv:1904.09075*, 2019.
- [38] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, A. Tsigirgos, “Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning,” *Nature Medicine*, vol. 24, no. 10, pp. 1559–1567, 2018.
- [39] C. Li, X. Wang, W. Liu, L. J. Latecki, “Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks,” *Medical Image Analysis*, vol. 45, pp. 121–133, 2018.
- [40] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, N. Navab, “Aggnet: Deep learning from crowds for mitosis detection in breast cancer histology images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1313–1321, 2016.
- [41] M. Saha, C. Chakraborty, D. Racoceanu, “Efficient deep learning model for mitosis detection using breast histopathology images,” *Computerized Medical Imaging and Graphics*, vol. 64, pp. 29–40, 2018.
- [42] M. G. Ertosun D. L. Rubin, “Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks,” in *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, vol. 2015, 2015, p. 1899.
- [43] K. Nagpal, D. Foote, F. Tan, Y. Liu, P.-H. C. Chen, D. F. Steiner, N. Manoj, N. Olson, J. L. Smith, A. Mohtashamian, *et al.*, “Development and validation of a deep learning algorithm for gleason grading of prostate cancer from biopsy specimens,” *JAMA Oncology*, vol. 6, no. 9, pp. 1372–1380, 2020.

- [44] A. Janowczyk, S. Doyle, H. Gilmore, A. Madabhushi, “A resolution adaptive deep hierarchical (radhical) learning scheme applied to nuclear segmentation of digital pathology images,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 3, pp. 270–276, 2018.
- [45] S. Wang, D. M. Yang, R. Rong, X. Zhan, G. Xiao, “Pathology image analysis using segmentation deep learning algorithms,” *The American Journal of Pathology*, vol. 189, no. 9, pp. 1686–1698, 2019.
- [46] A. Pimkin, G. Makarchuk, V. Kondratenko, M. Pisov, E. Krivov, M. Belyaev, “Ensembling neural networks for digital pathology images classification and segmentation,” in *International Conference Image Analysis and Recognition*, Springer, 2018, pp. 877–886.
- [47] J. T. Bushberg J. M. Boone, *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011.
- [48] N. C. Dalrymple, S. R. Prasad, F. M. El-Merhi, K. N. Chintapalli, “Price of isotropy in multidetector ct,” *Radiographics*, vol. 27, no. 1, pp. 49–62, 2007. DOI: <https://doi.org/10.1148/rg.271065037>.
- [49] V. Chan A. Perlas, “Basics of ultrasound imaging,” in *Atlas of ultrasound-guided procedures in interventional pain management*, Springer, 2011, pp. 13–19. DOI: [https://doi.org/10.1007/978-1-4419-1681-5\\_2](https://doi.org/10.1007/978-1-4419-1681-5_2).
- [50] W. E. Brant E. E. de Lange, *Essentials of body MRI*. OUP USA, 2012.
- [51] G. W. Boland, A. S. Guimaraes, P. R. Mueller, “The radiologist’s conundrum: Benefits and costs of increasing ct capacity and utilization,” *European radiology*, vol. 19, no. 1, pp. 9–11, 2009. DOI: <https://doi.org/10.1007/s00330-008-1159-7>.
- [52] R. J. McDonald, K. M. Schwartz, L. J. Eckel, F. E. Diehn, C. H. Hunt, B. J. Bartholmai, B. J. Erickson, D. F. Kallmes, “The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload,” *Academic Radiology*, vol. 22, no. 9, pp. 1191–1198, 2015, ISSN: 1076-6332. DOI: <https://doi.org/10.1016/j.acra.2015.05.007>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1076633215002457>.
- [53] R. Fitzgerald, “Error in radiology,” *Clinical Radiology*, vol. 56, no. 12, pp. 938–946, 2001, ISSN: 0009-9260. DOI: <https://doi.org/10.1053/crad.2001.0858>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S000992600190858X>.
- [54] K. J. Dreyer J. R. Geis, “When machines think: Radiology’s next frontier,” *Radiology*, vol. 285, no. 3, pp. 713–718, 2017.
- [55] A. Urushibara, T. Saida, K. Mori, T. Ishiguro, M. Sakai, S. Masuoka, T. Satoh, T. Masumoto, “Diagnosing uterine cervical cancer on a single t2-weighted image: Comparison between deep learning versus radiologists,” *European Journal of Radiology*, vol. 135, p. 109471, 2021. DOI: <https://doi.org/10.1016/j.ejrad.2020.109471>.

- [56] S. Masoudi, S. A. Harmon, S. Mehralivand, S. M. Walker, H. Raviprakash, U. Bagci, P. L. Choyke, B. Turkbey, “Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research,” *Journal of Medical Imaging*, vol. 8, no. 1, p. 010901, 2021. DOI: <https://doi.org/10.1117/1.JMI.8.1.010901>.
- [57] N. Cherian Kurian, A. Sethi, A. Reddy Konduru, A. Mahajan, S. U. Rane, “A 2021 update on cancer image analytics with deep learning,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1410, 2021. DOI: <https://doi.org/10.1002/widm.1410>.
- [58] A. A. A. Setio, A. Traverso, T. de Bel, M. S. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts, R. van der Gugten, P. A. Heng, B. Jansen, M. M. de Kaste, V. Kotov, J. Y.-H. Lin, J. T. Manders, A. Sónora-Mengana, J. C. García-Naranjo, E. Papavasileiou, M. Prokop, M. Saletta, C. M. Schaefer-Prokop, E. T. Scholten, L. Scholten, M. M. Snoeren, E. L. Torres, J. Vandemeulebroucke, N. Walasek, G. C. Zuidhof, B. van Ginneken, C. Jacobs, “Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The luna16 challenge,” *Medical Image Analysis*, vol. 42, pp. 1–13, 2017, ISSN: 1361-8415. DOI: <https://doi.org/10.1016/j.media.2017.06.015>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841517301020>.
- [59] (2021). Real noise mri, [Online]. Available: <https://realnoisemri.grand-challenge.org/Description/>.
- [60] (2021). Breast cancer wisconsin (diagnostic) data set, [Online]. Available: <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>.
- [61] (2021). Breast cancer data set analysis, [Online]. Available: <https://www.kaggle.com/bndarendeli/breast-cancer-dataset-analysis2>.
- [62] T. Minamoto, M. Mai, Z. Ronai, “K-ras mutation: Early detection in molecular diagnosis and risk assessment of colorectal, pancreas, and lung cancers—a review,” *Cancer Detection and prevention*, vol. 24, no. 1, pp. 1–12, 2000. DOI: <https://pubmed.ncbi.nlm.nih.gov/10757118/>.
- [63] M. R. Stratton, P. J. Campbell, P. A. Futreal, “The cancer genome,” *Nature*, vol. 458, no. 7239, pp. 719–724, 2009. DOI: <https://doi.org/10.1038/nature07943>.
- [64] N.-Y. Cho, J.-W. Park, X. Wen, Y.-J. Shin, J.-K. Kang, S.-H. Song, H.-P. Kim, T.-Y. Kim, J. M. Bae, G. H. Kang, “Blood-based detection of colorectal cancer using cancer-specific dna methylation markers,” *Diagnostics*, vol. 11, no. 1, p. 51, 2021. DOI: <https://doi.org/10.3390/diagnostics11010051>.
- [65] L. Y. Cheng, L. E. Haydu, P. Song, J. Nie, M. T. Tetzlaff, L. N. Kwong, J. E. Gershenwald, M. A. Davies, D. Y. Zhang, “High sensitivity sanger sequencing detection of braf mutations in metastatic melanoma ffpe tissue specimens,” *Scientific reports*, vol. 11, no. 1, pp. 1–9, 2021. DOI: <https://doi.org/10.1038/s41598-021-88391-5>.

- [66] H.-U. G. Weier, K. M. Greulich-Bode, Y. Ito, R. A. Lersch, J. Fung, “Fish in cancer diagnosis and prognostication: From cause to course of disease,” *Expert review of molecular diagnostics*, vol. 2, no. 2, pp. 109–119, 2002. DOI: 10.1586/14737159.2.2.109.
- [67] S. S. Yadav, J. Li, H. J. Lavery, K. K. Yadav, A. K. Tewari, “Next-generation sequencing technology in prostate cancer diagnosis, prognosis, and personalized treatment,” in *Urologic Oncology: Seminars and Original Investigations*, Elsevier, vol. 33, 2015, 267–e1. DOI: <https://doi.org/10.1016/j.urolonc.2015.02.009>.
- [68] R. Bu, A. K. Siraj, K. A. Al-Obaisi, S. Beg, M. Al Hazmi, D. Ajarim, A. Tulbah, F. Al-Dayel, K. S. Al-Kuraya, “Identification of novel brca founder mutations in middle eastern breast cancer patients using capture and sanger sequencing analysis,” *International journal of cancer*, vol. 139, no. 5, pp. 1091–1097, 2016. DOI: <https://doi.org/10.1002/ijc.30143>.
- [69] K. B. Mullis F. A. Faloona, “[21] specific synthesis of dna in vitro via a polymerase-catalyzed chain reaction,” *Methods in enzymology*, vol. 155, pp. 335–350, 1987. DOI: [https://doi.org/10.1016/0076-6879\(87\)55023-6](https://doi.org/10.1016/0076-6879(87)55023-6).
- [70] W. P. DOES, “Polymerase chain reaction,” *Journal of Investigative Dermatology*, vol. 133, 2013. DOI: <https://doi.org/10.1038/jid.2013.1>.
- [71] M. Kubista, J. M. Andrade, M. Bengtsson, A. Forootan, J. Jonák, K. Lind, R. Sindelka, R. Sjöback, B. Sjögren, L. Strömbom, A. Ståhlberg, N. Zoric, “The real-time polymerase chain reaction,” *Molecular Aspects of Medicine*, vol. 27, no. 2, pp. 95–125, 2006, Real-time Polymerase Chain Reaction, ISSN: 0098-2997. DOI: <https://doi.org/10.1016/j.mam.2005.12.007>.
- [72] L. Bonetta, “Prime time for real-time pcr,” *Nature Methods*, vol. 2, no. 4, pp. 305–312, 2005. DOI: <https://doi.org/10.1038/nmeth0405-305>.
- [73] G. U. Güleç Y. B. Turgut, “Real-time pcr: Advanced technologies and applications,” in *Fungal Infections of the Central Nervous System: Pathogens, Diagnosis, and Management*, M. Turgut, S. Challa, A. Akhaddar, Eds. Springer International Publishing, 2019, pp. 463–469, ISBN: 978-3-030-06088-6. DOI: 10.1007/978-3-030-06088-6\_38.
- [74] P. P. Reis, S. R. Rogatto, L. P. Kowalski, I. N. Nishimoto, J. C. Montovani, G. Corpus, J. A. Squire, S. Kamel-Reid, “Quantitative real-time pcr identifies a critical region of deletion on 22q13 related to prognosis in oral cancer,” *Oncogene*, vol. 21, no. 42, pp. 6480–6487, 2002. DOI: <https://doi.org/10.1038/sj.onc.1205864>.
- [75] S. Gal, C. Fidler, Y. Lo, M. Taylor, C. Han, J. Moore, A. Harris, J. Wainscoat, “Quantitation of circulating dna in the serum of breast cancer patients by real-time pcr,” *British journal of cancer*, vol. 90, no. 6, pp. 1211–1215, 2004. DOI: <https://doi.org/10.1038/sj.bjc.6601609>.

- [76] E. Bandrés, E. Cubedo, X. Agirre, R. Malumbres, R. Zarate, N. Ramirez, A. Abajo, A. Navarro, I. Moreno, M. Monzo, *et al.*, “Identification by real-time pcr of 13 mature micrnas differentially expressed in colorectal cancer and non-tumoral tissues,” *Molecular cancer*, vol. 5, no. 1, pp. 1–10, 2006. DOI: <https://doi.org/10.1186/1476-4598-5-29>.
- [77] D. C. Hayes, H. Secrist, C. S. Bangur, T. Wang, X. Zhang, D. Harlan, G. E. Goodman, R. L. Houghton, D. H. Persing, B. K. Zehentner, “Multigene real-time pcr detection of circulating tumor cells in peripheral blood of lung cancer patients,” *Anticancer research*, vol. 26, no. 2B, pp. 1567–1575, 2006. [Online]. Available: <https://ar.iijournals.org/content/26/2B/1567.short>.
- [78] Y. Zhang, M. Li, H. Wang, W. E. Fisher, P. H. Lin, Q. Yao, C. Chen, “Profiling of 95 micrnas in pancreatic cancer cell lines and surgical specimens by real-time pcr analysis,” *World journal of surgery*, vol. 33, no. 4, pp. 698–709, 2009. DOI: <https://doi.org/10.1007/s00268-008-9833-0>.
- [79] A. Ståhlberg, N. Zoric, P. Åman, M. Kubista, “Quantitative real-time pcr for cancer detection: The lymphoma case,” *Expert review of molecular diagnostics*, vol. 5, no. 2, pp. 221–230, 2005. DOI: <https://doi.org/10.1586/14737159.5.2.221>.
- [80] K. E. Resnick, H. Alder, J. P. Hagan, D. L. Richardson, C. M. Croce, D. E. Cohn, “The detection of differentially expressed micrnas from the serum of ovarian cancer patients using a novel real-time pcr platform,” *Gynecologic oncology*, vol. 112, no. 1, pp. 55–59, 2009. DOI: <https://doi.org/10.1016/j.ygyno.2008.08.036>.
- [81] Y. Nie, B. Wang, Z. Zhao, H. Zhou, “Allele-specific pcr and its application in forensic science,” *Fa Yi Xue Za Zhi*, vol. 30, no. 4, pp. 282–287, 2014. DOI: PMID:25434094.
- [82] Z. Yang, N. Zhao, D. Chen, K. Wei, N. Su, J.-F. Huang, H.-Q. Xu, G.-J. Duan, W.-L. Fu, Q. Huang, “Improved detection of braf v600e using allele-specific pcr coupled with external and internal controllers,” *Scientific reports*, vol. 7, no. 1, pp. 1–12, 2017. DOI: <https://doi.org/10.1038/s41598-017-14140-2>.
- [83] F. Sanger, S. Nicklen, A. R. Coulson, “Dna sequencing with chain-terminating inhibitors,” *Proceedings of the national academy of sciences*, vol. 74, no. 12, pp. 5463–5467, 1977. DOI: <https://doi.org/10.1073/pnas.74.12.5463>.
- [84] (2021). Sanger sequencing, [Online]. Available: <https://courtneyarmitagevisuals.com/sanger-sequencing>.
- [85] W. J. Martin, J. R. Warmington, B. R. Galinski, M. Gallagher, R. W. Davies, M. S. Beck, S. G. Oliver, “Automation of dna sequencing: A system to perform the sanger dideoxysequencing reactions,” *Bio/technology*, vol. 3, no. 10, pp. 911–915, 1985. DOI: <https://doi.org/10.1038/nbt1085-911>.
- [86] M. L. Metzker, “Emerging technologies in dna sequencing,” *Genome research*, vol. 15, no. 12, pp. 1767–1776, 2005. DOI: <http://www.genome.org/cgi/doi/10.1101/gr.3770505>.

- [87] J. Kulski, *Next Generation Sequencing: Advances, Applications and Challenges*. BoD–Books on Demand, 2016.
- [88] T. Tucker, M. Marra, J. M. Friedman, “Massively parallel sequencing: The next big thing in genetic medicine,” *The American Journal of Human Genetics*, vol. 85, no. 2, pp. 142–154, 2009. DOI: <https://doi.org/10.1016/j.humimm.2021.02.012>.
- [89] S. Goodwin, J. D. McPherson, W. R. McCombie, “Coming of age: Ten years of next-generation sequencing technologies,” *Nature Reviews Genetics*, vol. 17, no. 6, p. 333, 2016. DOI: <https://doi.org/10.1038/nrg.2016.49>.
- [90] D. Dressman, H. Yan, G. Traverso, K. W. Kinzler, B. Vogelstein, “Transforming single dna molecules into fluorescent magnetic particles for detection and enumeration of genetic variations,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 15, pp. 8817–8822, 2003. DOI: <https://doi.org/10.1073/pnas.1133470100>.
- [91] T. Hu, N. Chitnis, D. Monos, A. Dinh, “Next-generation sequencing technologies: An overview,” *Human Immunology*, 2021. DOI: <https://doi.org/10.1016/j.humimm.2021.02.012>.
- [92] R. England M. Pettersson, “Pyro q-cpg<sup>TM</sup>: Quantitative analysis of methylation in multiple cpg sites by pyrosequencing<sup>®</sup>,” *Nature methods*, vol. 2, no. 10, pp. i–ii, 2005. DOI: <https://doi.org/10.1038/nmeth800>.
- [93] B. Merriman, I. T. R&D Team, J. M. Rothberg, “Progress in ion torrent semiconductor chip based sequencing,” *Electrophoresis*, vol. 33, no. 23, pp. 3397–3417, 2012. DOI: <https://doi.org/10.1002/elps.201200424>.
- [94] S. E. Bohndiek K. M. Brindle, “Imaging and ‘omic’ methods for the molecular diagnosis of cancer,” *Expert review of molecular diagnostics*, vol. 10, no. 4, pp. 417–434, 2010. DOI: <https://doi.org/10.1586/erm.10.20>.
- [95] A. Singh N. Kumar, “A review on dna microarray technology,” *IJCRR*, vol. 5, no. 22, pp. 01–05, 2013. [Online]. Available: [http://ijcrr.com/article\\_html.php?did=1018](http://ijcrr.com/article_html.php?did=1018).
- [96] (2021). Dna microarray, [Online]. Available: <https://learn.genetics.utah.edu/content/labs/microarray/>.
- [97] A. Schulze J. Downward, “Navigating gene expression using microarrays—a technology review,” *Nature cell biology*, vol. 3, no. 8, E190–E195, 2001. DOI: <https://doi.org/10.1038/35087138>.
- [98] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, *et al.*, “Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling,” *Nature*, vol. 403, no. 6769, pp. 503–511, 2000. DOI: <https://doi.org/10.1038/35000501>.
- [99] M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, *et al.*, “Molecular classification of cutaneous malignant melanoma by gene expression profiling,” *Nature*, vol. 406, no. 6795, pp. 536–540, 2000. DOI: <https://doi.org/10.1038/35020115>.

- [100] T. E. Hickey, L. A. Selth, K. M. Chia, G. Laven-Law, H. H. Milioli, D. Roden, S. Jindal, M. Hui, J. Finlay-Schultz, E. Ebrahimie, *et al.*, “The androgen receptor is a tumor suppressor in estrogen receptor-positive breast cancer,” *Nature Medicine*, vol. 27, no. 2, pp. 310–320, 2021. DOI: <https://doi.org/10.1038/s41591-020-01168-7>.
- [101] Z. Wang, M. Gerstein, M. Snyder, “Rna-seq: A revolutionary tool for transcriptomics,” *Nature reviews genetics*, vol. 10, no. 1, pp. 57–63, 2009. DOI: <https://doi.org/10.1038/nrg2484>.
- [102] J. B. Wolf, “Principles of transcriptome analysis and gene expression quantification: An rna-seq tutorial,” *Molecular ecology resources*, vol. 13, no. 4, pp. 559–572, 2013. DOI: <https://doi.org/10.1111/1755-0998.12109>.
- [103] C. Wang, B. Gong, P. R. Bushel, J. Thierry-Mieg, D. Thierry-Mieg, J. Xu, H. Fang, H. Hong, J. Shen, Z. Su, *et al.*, “The concordance between rna-seq and microarray data depends on chemical treatment and transcript abundance,” *Nature biotechnology*, vol. 32, no. 9, pp. 926–932, 2014. DOI: <https://doi.org/10.1038/nbt.3001>.
- [104] S. Zhao, W.-P. Fung-Leung, A. Bittner, K. Ngo, X. Liu, “Comparison of rna-seq and microarray in transcriptome profiling of activated t cells,” *PloS one*, vol. 9, no. 1, e78644, 2014. DOI: <https://doi.org/10.1371/journal.pone.0078644>.
- [105] X. Xu, Y. Zhang, J. Williams, E. Antoniou, W. R. McCombie, S. Wu, W. Zhu, N. O. Davidson, P. Denoya, E. Li, “Parallel comparison of illumina rna-seq and affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated ht-29 colon cancer cells and simulated datasets,” *BMC bioinformatics*, vol. 14, no. 9, pp. 1–14, 2013. DOI: <https://doi.org/10.1186/1471-2105-14-S9-S1>.
- [106] S.-L. T. Normand, “Meta-analysis: Formulating, evaluating, combining, and reporting,” *Statistics in medicine*, vol. 18, no. 3, pp. 321–359, 1999. DOI: [https://doi.org/10.1002/\(SICI\)1097-0258\(19990215\)18:3<321::AID-SIM28>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0258(19990215)18:3<321::AID-SIM28>3.0.CO;2-P).
- [107] A.-B. Haidich, “Meta-analysis in medical research,” *Hippokratia*, vol. 14, no. Suppl 1, p. 29, 2010. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049418/>.
- [108] M. Wang, Y. Yang, J. Xu, W. Bai, X. Ren, H. Wu, “Circrnas as biomarkers of cancer: A meta-analysis,” *BMC cancer*, vol. 18, no. 1, pp. 1–10, 2018. DOI: <https://doi.org/10.1186/s12885-018-4213-0>.
- [109] K. M. Dale, C. I. Coleman, N. N. Henyan, J. Kluger, C. M. White, “Statins and cancer risk: A meta-analysis,” *Jama*, vol. 295, no. 1, pp. 74–80, 2006. DOI: [10.1001/jama.295.1.74](https://doi.org/10.1001/jama.295.1.74).
- [110] A. B. de Gonzalez, S. Sweetland, E. Spencer, “A meta-analysis of obesity and the risk of pancreatic cancer,” *British journal of cancer*, vol. 89, no. 3, pp. 519–523, 2003. DOI: <https://doi.org/10.1038/sj.bjc.6601140>.

- [111] B. Galili, X. Tekpli, V. N. Kristensen, Z. Yakhini, “Efficient gene expression signature for a breast cancer immuno-subtype,” *Plos one*, vol. 16, no. 1, e0245215, 2021. DOI: <https://doi.org/10.1371/journal.pone.0245215>.
- [112] S. Arora, A. Matta, N. K. Shukla, S. Deo, R. Ralhan, “Identification of differentially expressed genes in oral squamous cell carcinoma,” *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center*, vol. 42, no. 2, pp. 97–108, 2005. DOI: <https://doi.org/10.1002/mc.20048>.
- [113] L. Wang, Z. Feng, X. Wang, X. Wang, X. Zhang, “Degseq: An r package for identifying differentially expressed genes from rna-seq data,” *Bioinformatics*, vol. 26, no. 1, pp. 136–138, 2010. DOI: <https://doi.org/10.1093/bioinformatics/btp612>.
- [114] T. Park, S.-G. Yi, S. Lee, S. Y. Lee, D.-H. Yoo, J.-I. Ahn, Y.-S. Lee, “Statistical tests for identifying differentially expressed genes in time-course microarray experiments,” *Bioinformatics*, vol. 19, no. 6, pp. 694–703, 2003. DOI: [10.1093/bioinformatics/btg068](https://doi.org/10.1093/bioinformatics/btg068).
- [115] M. S. Pepe, G. Longton, G. L. Anderson, M. Schummer, “Selecting differentially expressed genes from microarray experiments,” *Biometrics*, vol. 59, no. 1, pp. 133–142, 2003. DOI: <https://doi.org/10.1186/gb-2003-4-6-r41>.
- [116] F. Hong R. Breitling, “A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments,” *Bioinformatics*, vol. 24, no. 3, pp. 374–382, 2008. DOI: <https://doi.org/10.1093/bioinformatics/btm620>.
- [117] H. Hozhabri, A. Lashkari, S.-M. Razavi, A. Mohammadian, “Integration of gene expression data identifies key genes and pathways in colorectal cancer,” *Medical Oncology*, vol. 38, no. 1, pp. 1–14, 2021. DOI: <https://doi.org/10.1007/s12032-020-01448-9>.
- [118] Z. Du, X. Zhang, W. Gao, J. Yang, “Differentially expressed genes pcca, echs1, and hadh are potential prognostic biomarkers for gastric cancer,” *Science Progress*, vol. 104, no. 2, p. 00368504211011344, 2021. DOI: <https://doi.org/10.1177/00368504211011344>.
- [119] Z.-B. Ke, Y.-P. Wu, P. Huang, J. Hou, Y.-H. Chen, R.-N. Dong, F. Lin, Y. Wei, X.-Y. Xue, C.-F. Ng, *et al.*, “Identification of novel genes in testicular cancer microenvironment based on ESTIMATE algorithm-derived immune scores,” *Journal of cellular physiology*, vol. 236, no. 1, pp. 706–713, 2021. DOI: [10.1002/jcp.29898](https://doi.org/10.1002/jcp.29898).
- [120] S. Dasgupta, S. Mondal, A. Khan, R. K. Pal, G. Saha, “Identification of differentially expressed genes using deep learning in bioinformatics,” in *Proceedings of International Conference on Frontiers in Computing and Systems*, Springer, 2021, pp. 521–532. DOI: [https://doi.org/10.1007/978-981-15-7834-2\\_49](https://doi.org/10.1007/978-981-15-7834-2_49).

- [121] B. Zhang S. Horvath, “A general framework for weighted gene co-expression network analysis,” *Statistical applications in genetics and molecular biology*, vol. 4, no. 1, 2005. DOI: 10.2202/1544-6115.1128.
- [122] M. Hamed, C. Spaniol, A. Zapp, V. Helms, “Integrative network-based approach identifies key genetic elements in breast invasive carcinoma,” *BMC Genomics*, vol. 16, no. 5, pp. 1–14, 2015. DOI: <https://doi.org/10.1186/1471-2164-16-S5-S2>.
- [123] M. Shi, R. D. Beauchamp, B. Zhang, “A network-based gene expression signature informs prognosis and treatment for colorectal cancer patients,” *PLoS ONE*, vol. 7, no. 7, e41292, Jul. 2012. DOI: 10.1371/journal.pone.0041292.
- [124] M. Ameri, M. Taheri-Anganeh, A. Movahedpour, A. Savardashtaki, “A network-based approach to identify key genes between follicular thyroid cancer and follicular thyroid adenoma,” *Gene Reports*, vol. 23, p. 101075, 2021. DOI: <https://doi.org/10.1016/j.genrep.2021.101075>.
- [125] M. A. Hossain, T. A. Asa, M. R. Rahman, M. A. Moni, “Network-based approach to identify key candidate genes and pathways shared by thyroid cancer and chronic kidney disease,” *Informatics in Medicine Unlocked*, vol. 16, p. 100240, 2019. DOI: <https://doi.org/10.1016/j.imu.2019.100240>.
- [126] R. Liu, C.-X. Guo, H.-H. Zhou, “Network-based approach to identify prognostic biomarkers for estrogen receptor-positive breast cancer treatment with tamoxifen,” *Cancer Biology & Therapy*, vol. 16, no. 2, pp. 317–324, 2015. DOI: <https://doi.org/10.1080/15384047.2014.1002360>.
- [127] M. J. Hossain, U. N. Chowdhury, M. B. Islam, S. Uddin, M. B. Ahmed, J. M. Quinn, M. A. Moni, “Machine learning and network-based models to identify genetic risk factors to the progression and survival of colorectal cancer,” *Computers in Biology and Medicine*, p. 104539, 2021. DOI: <https://doi.org/10.1016/j.combiomed.2021.104539>.
- [128] A. M. Turing, “Computing machinery and intelligence,” in *Parsing the Turing Test*, Springer, 2009, pp. 23–65.
- [129] M. Flasiński, “History of artificial intelligence,” in *Introduction to Artificial Intelligence*, Springer, 2016, pp. 3–13.
- [130] J. McCarthy, “Artificial intelligence, logic and formalizing common sense,” in *Philosophical logic and artificial intelligence*, Springer, 1989, pp. 161–190.
- [131] American History Collection. (2020). Perceptron, mark 1. (Visited on 2021-1-8), [Online]. Available: [https://americanhistory.si.edu/collections/search/object/nmah\\_334414](https://americanhistory.si.edu/collections/search/object/nmah_334414).
- [132] M. Minsky S. A. Papert, *Perceptrons: An introduction to computational geometry*. MIT Press, 2017.
- [133] R. Carpenter. (2011). Jabberwacky - live chat bot - ai artificial intelligence chatbot, [Online]. Available: <http://www.jabberwacky.com/>.
- [134] M. Campbell, A. J. Hoane Jr, F.-H. Hsu, “Deep blue,” *Artificial Intelligence*, vol. 134, no. 1-2, pp. 57–83, 2002.

- [135] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, *et al.*, “Stanley: The robot that won the darpa grand challenge,” *Journal of Field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.
- [136] J. Aron, “How innovative is apple’s new voice assistant, siri?” *New Scientist*, vol. 212, no. 2836, p. 24, 2011. DOI: [https://doi.org/10.1016/S0262-4079\(11\)62647-X](https://doi.org/10.1016/S0262-4079(11)62647-X).
- [137] J. X. Chen, “The evolution of computing: Alphago,” *Computing in Science & Engineering*, vol. 18, no. 4, pp. 4–7, 2016.
- [138] E. Musk, “An integrated brain-machine interface platform with thousands of channels,” *J Med Internet Res*, vol. 21, no. 10, e16194, 2019. DOI: <https://doi.org/10.2196/16194>.
- [139] E. Callaway, “it will change everything’: Deepmind’s ai makes gigantic leap in solving protein structures.,” *Nature*, 2020.
- [140] N. Buduma N. Locascio, *Fundamentals of deep learning: Designing next-generation machine intelligence algorithms*. O’Reilly Media, Inc., 2017.
- [141] V. B. Mountcastle, “Modality and topographic properties of single neurons of cat’s somatic sensory cortex,” *Journal of Neurophysiology*, vol. 20, no. 4, pp. 408–434, 1957.
- [142] D. E. Rumelhart, G. E. Hinton, R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [143] J. Patterson A. Gibson, *Deep Learning: A practitioner’s Approach*. O’Reilly Media, Inc., 2017.
- [144] Y. LeCun, Y. Bengio, G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [145] N. McClure, *TensorFlow Machine Learning Cookbook*. Packt Publishing Ltd, 2017.
- [146] S. Saha. (2020). A comprehensive guide to convolutional neural networks - the eli5 way, [Online]. Available: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.
- [147] Y. Shi, Y. E. Sagduyu, K. Davaslioglu, R. Levy, “Vulnerability detection and analysis in adversarial deep learning,” in *Computer Communications and Networks*, Springer International Publishing, 2018, pp. 211–234. DOI: 10.1007/978-3-319-92624-7\_9.
- [148] I. J. Goodfellow, J. Shlens, C. Szegedy, *Explaining and harnessing adversarial examples*, 2015. arXiv: 1412.6572 [stat.ML]. [Online]. Available: <https://arxiv.org/abs/1412.6572>.
- [149] J. Su, D. V. Vargas, K. Sakurai, “One pixel attack for fooling deep neural networks,” *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 5, pp. 828–841, Oct. 2019. DOI: 10.1109/tevc.2019.2890858.

- [150] J. Zuluaga-Gomez, Z. Al Masry, K. Benaggoune, S. Meraghni, N. Zerhouni, "A cnn-based methodology for breast cancer diagnosis using thermal images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–15, 2020.
- [151] M. Gour, S. Jain, T. S. Kumar, "Residual learning based CNN for breast cancer histopathological image classification," *International Journal of Imaging Systems and Technology*, vol. 30, no. 3, pp. 621–635, Feb. 2020. DOI: 10.1002/ima.22403.
- [152] B. E. Bejnordi, M. Mullooly, R. M. Pfeiffer, S. Fan, P. M. Vacek, D. L. Weaver, S. Herschorn, L. A. Brinton, B. van Ginneken, N. Karssemeijer, A. H. Beck, G. L. Gierach, J. A. W. M. van der Laak, M. E. Sherman, "Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies," *Modern Pathology*, vol. 31, no. 10, pp. 1502–1512, Jun. 2018. DOI: 10.1038/s41379-018-0073-z.
- [153] H. D. Couture, L. A. Williams, J. Geradts, S. J. Nyante, E. N. Butler, J. S. Marron, C. M. Perou, M. A. Troester, M. Niethammer, "Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype," *npj Breast Cancer*, vol. 4, no. 1, Sep. 2018. DOI: 10.1038/s41523-018-0079-1.
- [154] L. Duran-Lopez, J. P. Dominguez-Morales, A. F. Conde-Martin, S. Vicente-Diaz, A. Linares-Barranco, "Prometeo: A cnn-based computer-aided diagnosis system for wsi prostate cancer detection," *IEEE Access*, vol. 8, pp. 128 613–128 628, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9139241>.
- [155] A. Hartenstein, F. Lübbe, A. D. J. Baur, M. M. Rudolph, C. Furth, W. Brenner, H. Amthauer, B. Hamm, M. Makowski, T. Penzkofer, "Prostate cancer nodal staging: Using deep learning to predict 68ga-PSMA-positivity from CT imaging alone," *Scientific Reports*, vol. 10, no. 1, Feb. 2020. DOI: 10.1038/s41598-020-60311-z.
- [156] E. Arvaniti, K. S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P. J. Wild, J. H. Rüschoff, M. Claassen, "Automated gleason grading of prostate cancer tissue microarrays via deep learning," *Scientific Reports*, vol. 8, no. 1, Aug. 2018. DOI: 10.1038/s41598-018-30535-1.
- [157] K. Nagpal, D. Foote, Y. Liu, P-H. C. Chen, E. Wulczyn, F. Tan, N. Olson, J. L. Smith, A. Mohtashamian, J. H. Wren, G. S. Corrado, R. MacDonald, L. H. Peng, M. B. Amin, A. J. Evans, A. R. Sangoi, C. H. Mermel, J. D. Hipp, M. C. Stumpe, "Development and validation of a deep learning algorithm for improving gleason scoring of prostate cancer," *npj Digital Medicine*, vol. 2, no. 1, Jun. 2019. DOI: 10.1038/s41746-019-0112-2.
- [158] S. Yoo, I. Gujrathi, M. A. Haider, F. Khalvati, "Prostate cancer detection using deep convolutional neural networks," *Scientific Reports*, vol. 9, no. 1, Dec. 2019. DOI: 10.1038/s41598-019-55972-4.
- [159] P. Afshar, A. Mohammadi, P. N. Tyrrell, P. Cheung, A. Sigiuk, K. N. Plataniotis, E. T. Nguyen, A. Oikonomou, "\text {DRTOP}: Deep learning-based radiomics for the time-to-event outcome prediction in lung cancer," *Scientific Reports*, vol. 10, no. 1, Jul. 2020. DOI: 10.1038/s41598-020-69106-8.

- [160] F. Kanavati, G. Toyokawa, S. Momosaki, M. Rambeau, Y. Kozuma, F. Shoji, K. Yamazaki, S. Takeo, O. Iizuka, M. Tsuneki, “Weakly-supervised learning for lung carcinoma classification using deep learning,” *Scientific Reports*, vol. 10, no. 1, Jun. 2020. DOI: 10.1038/s41598-020-66333-x.
- [161] Y.-H. Lai, W.-N. Chen, T.-C. Hsu, C. Lin, Y. Tsao, S. Wu, “Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning,” *Scientific Reports*, vol. 10, no. 1, Mar. 2020. DOI: 10.1038/s41598-020-61588-w.
- [162] P. Fontaine, O. Acosta, J. Castelli, R. D. Crevoisier, H. Müller, A. Depeursinge, “The importance of feature aggregation in radiomics: A head and neck cancer study,” *Scientific Reports*, vol. 10, no. 1, Nov. 2020. DOI: 10.1038/s41598-020-76310-z. [Online]. Available: <https://doi.org/10.1038/s41598-020-76310-z>.
- [163] P. Tschandl, C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, M. Janda, A. Lallas, C. Longo, J. Malvehy, J. Paoli, S. Puig, C. Rosendahl, H. P. Soyer, I. Zalaudek, H. Kittler, “Human–computer collaboration for skin cancer recognition,” *Nature Medicine*, vol. 26, no. 8, pp. 1229–1234, Jun. 2020. DOI: 10.1038/s41591-020-0942-0.
- [164] J. M. Dolezal, A. Trzcinska, C.-Y. Liao, S. Kochanny, E. Blair, N. Agrawal, X. M. Keutgen, P. Angelos, N. A. Cipriani, A. T. Pearson, “Deep learning prediction of BRAF-RAS gene expression signature identifies noninvasive follicular thyroid neoplasms with papillary-like nuclear features,” *Modern Pathology*, vol. 34, no. 5, pp. 862–874, Dec. 2020. DOI: 10.1038/s41379-020-00724-3.
- [165] N. H. Tran, R. Qiao, L. Xin, X. Chen, B. Shan, M. Li, “Personalized deep learning of individual immunopeptidomes to identify neoantigens for cancer vaccines,” *Nature Machine Intelligence*, vol. 2, no. 12, pp. 764–771, Nov. 2020. DOI: 10.1038/s42256-020-00260-4.
- [166] A. Mencattini, D. D. Giuseppe, M. C. Comes, P. Casti, F. Corsi, F. R. Bertani, L. Ghibelli, L. Businaro, C. D. Natale, M. C. Parrini, E. Martinelli, “Discovering the hidden messages within cell trajectories using a deep learning approach for in vitro evaluation of cancer drug treatments,” *Scientific Reports*, vol. 10, no. 1, May 2020. DOI: 10.1038/s41598-020-64246-3.
- [167] R. Ramirez, Y.-C. Chiu, S. Zhang, J. Ramirez, Y. Chen, Y. Huang, Y.-F. Jin, “Prediction and interpretation of cancer survival using graph convolution neural networks,” *Methods*, Jan. 2021. DOI: 10.1016/j.ymeth.2021.01.004.
- [168] Y. Xie, W.-Y. Meng, R.-Z. Li, Y.-W. Wang, X. Qian, C. Chan, Z.-F. Yu, X.-X. Fan, H.-D. Pan, C. Xie, Q.-B. Wu, P.-Y. Yan, L. Liu, Y.-J. Tang, X.-J. Yao, M.-F. Wang, E. L.-H. Leung, “Early lung cancer diagnostic biomarker discovery by machine learning methods,” *Translational Oncology*, vol. 14, no. 1, p. 100907, Jan. 2021. DOI: 10.1016/j.tranon.2020.100907.
- [169] B. Zeng, B. S. Glicksberg, P. Newbury, E. Chekalin, J. Xing, K. Liu, A. Wen, C. Chow, B. Chen, “OCTAD: An open workspace for virtually screening therapeutics targeting precise cancer patient groups using gene expression features,” *Nature Protocols*, vol. 16, no. 2, pp. 728–753, Dec. 2020. DOI: 10.1038/s41596-020-00430-z.

- [170] T. Ahn, T. Goo, C.-h. Lee, S. Kim, K. Han, S. Park, T. Park, “Deep learning-based identification of cancer or normal tissue using gene expression data,” in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, Dec. 2018. DOI: 10.1109/bibm.2018.8621108.
- [171] A. Binder, M. Bockmayr, M. Hägele, S. Wienert, D. Heim, K. Hellweg, M. Ishii, A. Stenzinger, A. Hocke, C. Denkert, K.-R. Müller, F. Klauschen, “Morphological and molecular breast cancer profiling through explainable machine learning,” *Nature Machine Intelligence*, vol. 3, no. 4, pp. 355–366, Mar. 2021. DOI: 10.1038/s42256-021-00303-4.
- [172] B. Galili, X. Tekpli, V. N. Kristensen, Z. Yakhini, “Efficient gene expression signature for a breast cancer immuno-subtype,” *PLOS ONE*, vol. 16, no. 1, J. J. Li, Ed., e0245215, Jan. 2021. DOI: 10.1371/journal.pone.0245215.
- [173] J. Vivian, A. A. Rao, F. A. Nothhaft, C. Ketchum, J. Armstrong, A. Novak, J. Pfeil, J. Narkizian, A. D. Deran, A. Musselman-Brown, H. Schmidt, P. Amstutz, B. Craft, M. Goldman, K. Rosenbloom, M. Cline, B. O’Connor, M. Hanna, C. Birger, W. J. Kent, D. A. Patterson, A. D. Joseph, J. Zhu, S. Zaranek, G. Getz, D. Haussler, B. Paten, “Toil enables reproducible, open source, big biomedical data analyses,” *Nature Biotechnology*, vol. 35, no. 4, pp. 314–316, Apr. 2017. DOI: 10.1038/nbt.3772.
- [174] A. D. Rouillard, G. W. Gundersen, N. F. Fernandez, Z. Wang, C. D. Monteiro, M. G. McDermott, A. Ma’ayan, “The harmonizome: A collection of processed datasets gathered to serve and mine knowledge about genes and proteins,” *Database*, vol. 2016, baw100, 2016. DOI: 10.1093/database/baw100.
- [175] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. DOI: 10.1038/s41586-020-2649-2.
- [176] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, Xiaoqiang Zheng, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: <https://www.tensorflow.org/>.
- [177] J. Su, D. V. Vargas, K. Sakurai, “One pixel attack for fooling deep neural networks,” *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 5, pp. 828–841, 2019. DOI: 10.1109/TEVC.2019.2890858.

- [178] O. Dulgerci. (2019). Minimizing with differential evolution. (Visited on 2021-6-18), [Online]. Available: <https://mathematica.stackexchange.com/questions/193009/minimizing-with-differential-evolution>.
- [179] M. K. Elbashir, M. Ezz, M. Mohammed, S. S. Saloum, "Lightweight convolutional neural network for breast cancer classification using RNA-seq gene expression data," *IEEE Access*, vol. 7, pp. 185 338–185 348, 2019. DOI: 10.1109/access.2019.2960722.
- [180] P. DANAEE, R. GHAEINI, D. A. HENDRIX, "A DEEP LEARNING APPROACH FOR CANCER DETECTION AND RELEVANT GENE IDENTIFICATION," in *Biocomputing 2017*, WORLD SCIENTIFIC, Nov. 2016. DOI: 10.1142/9789813207813\_0022.
- [181] T. Rajkumar, K. Sabitha, N. Vijayalakshmi, S. Shirley, M. V. Bose, G. Gopal, G. Selvaluxmy, "Identification and validation of genes involved in cervical tumorigenesis," *BMC Cancer*, vol. 11, no. 1, Feb. 2011. DOI: 10.1186/1471-2407-11-80. [Online]. Available: <https://doi.org/10.1186/1471-2407-11-80>.
- [182] X. Mao, X. Zhang, X. Zheng, Y. Chen, Z. Xuan, P. Huang, "Curcumin suppresses LGR5( )colorectal cancer stem cells by inducing autophagy and via repressing TFAP2a – mediated," *Journal of Natural Medicines*, vol. 75, no. 3, pp. 590–601, Mar. 2021. DOI: 10.1007/s11418-021-01505-1. [Online]. Available: <https://doi.org/10.1007/s11418-021-01505-1>.
- [183] D. Appleman, E. R. Skavinski, A. M. Stein, "Catalase studies on normal and cancerous rats," *Cancer Research*, vol. 10, no. 8, pp. 498–505, 1950. [Online]. Available: <https://cancerres.aacrjournals.org/content/10/8/498.long>.
- [184] O. A. Adebayo, O. Akinloye, O. A. Adaramoye, "Cerium oxide nanoparticles elicit antitumorigenic effect in experimental breast cancer induced by n-methyl- n-nitrosourea and benzo( a )pyrene in female wistar rats," *Journal of Biochemical and Molecular Toxicology*, vol. 35, no. 4, Dec. 2020. DOI: 10.1002/jbt.22687.
- [185] A. C. Famurewa, C. A. Ekeleme-Egedigwe, E. E. David, C. O. Eleazu, A. M. Folawiyo, N. A. Obasi, "Zinc abrogates anticancer drug tamoxifen-induced hepatotoxicity by suppressing redox imbalance, NO/iNOS/NF- $\kappa$  signaling, and caspase-3-dependent apoptosis in female rats," *Toxicology Mechanisms and Methods*, vol. 30, no. 2, pp. 115–123, Oct. 2019. DOI: 10.1080/15376516.2019.1669243.
- [186] S. P. Shah, R. D. Morin, J. Khattra, L. Prentice, T. Pugh, A. Burleigh, A. Delaney, K. Gelmon, R. Guliany, J. Senz, C. Steidl, R. A. Holt, S. Jones, M. Sun, G. Leung, R. Moore, T. Severson, G. A. Taylor, A. E. Teschendorff, K. Tse, G. Turashvili, R. Varhol, R. L. Warren, P. Watson, Y. Zhao, C. Caldas, D. Huntsman, M. Hirst, M. A. Marra, S. Aparicio, "Mutational evolution in a lobular breast

tumour profiled at single nucleotide resolution,” *Nature*, vol. 461, no. 7265, pp. 809–813, Oct. 2009. DOI: 10.1038/nature08489.

- [187] C. Mitsopoulos, P. Di Micco, E. V. Fernandez, D. Dolciami, E. Holt, I. L. Mica, E. A. Coker, J. E. Tym, J. Campbell, K. H. Che, B. Ozer, C. Kannas, A. A. Antolin, P. Workman, B. Al-Lazikani, “canSAR: Update to the cancer translational research and drug discovery knowledgebase,” *Nucleic Acids Research*, vol. 49, no. D1, pp. D1074–D1082, Nov. 2020. DOI: 10.1093/nar/gkaa1059. [Online]. Available: <https://doi.org/10.1093/nar/gkaa1059>.



## ONE-PIXEL ATTACK RESULTS<sup>A</sup>

---

The gene expression values of the tissues in the TCGA and GTEx dataset were converted to images as mentioned in *Methods 4.3* and the training was carried out. After the training, the one-pixel attack method was applied to these images, which were obtained by estimating the neural network with an accuracy value of 97.7% (See *4.Methods*). The results of the applied one-pixel attack method on the basis of all samples are shared in Table A.

**Table A.1** One-pixel attack results applied to each sample

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-06-0646-01	ENSG00000163513	2381	65393	1	0
TCGA-13-0720-01	ENSG00000129250	3453	131040	1	0
TCGA-13-0906-01	ENSG00000129250	3459	131016	1	0
TCGA-2J-AABA-01	ENSG00000129250	12221	126643	1	0
TCGA-33-A4WN-01	ENSG00000129250	6974	130816	1	0
TCGA-3U-A98D-01	ENSG00000215301	14570	65378	1	0
TCGA-44-A47B-01	ENSG00000188157	50975	149	1	0
TCGA-50-5932-01	ENSG00000129250	8389	130816	1	0
TCGA-5U-AB0E-01	ENSG00000129250	3086	131024	1	0
TCGA-61-2110-01	ENSG00000129250	1862	130816	1	0
TCGA-67-6217-01	ENSG00000129250	10535	130816	1	0
TCGA-78-7540-01	ENSG00000188157	51134	4085	1	0
TCGA-78-7540-01	ENSG00000188157	51134	66455	1	0
TCGA-85-A513-01	ENSG00000188157	52754	255	1	0
TCGA-95-A4VK-01	ENSG00000129250	6224	130816	1	0
TCGA-96-8170-01	ENSG00000188157	35721	66976	1	0
TCGA-96-8170-01	ENSG00000188157	35721	67004	1	0
TCGA-97-7941-01	ENSG00000129250	9437	130820	1	0
TCGA-A1-A0SF-01	ENSG00000129250	4400	130816	1	0
TCGA-A2-A0D3-01	ENSG00000129250	4126	130816	1	0
TCGA-A2-A0EU-01	ENSG00000129250	3199	42457	1	0
TCGA-A2-A0EX-01	ENSG00000129250	6185	130816	1	0
TCGA-A2-A1FW-01	ENSG00000129250	1415	130816	1	0
TCGA-A4-8630-01	ENSG00000129250	14447	130816	1	0

Table A.1 continued from previous page

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-A5-A2K7-01	ENSG00000138821	1220	131017	1	0
TCGA-A7-A425-01	ENSG00000138821	805	131071	1	0
TCGA-A8-A06T-01	ENSG00000129250	3295	130816	1	0
TCGA-A8-A07W-01	ENSG00000129250	5259	130816	1	0
TCGA-A8-A094-01	ENSG00000129250	6230	130816	1	0
TCGA-AB-2815-03	ENSG00000129250	961	52614	1	0
TCGA-AB-2956-03	ENSG00000129250	3655	125628	1	0
TCGA-AB-2979-03	ENSG00000129250	247	130933	1	0
TCGA-AB-2991-03	ENSG00000129250	930	130816	1	0
TCGA-AB-3005-03	ENSG00000129250	1413	130816	1	0
TCGA-AN-A0AR-01	ENSG00000124942	8027	131015	1	0
TCGA-AO-A0J8-01	ENSG00000129250	4195	130816	1	0
TCGA-AO-A12C-01	ENSG00000129250	2822	119809	1	0
TCGA-AQ-A04L-01	ENSG00000129250	2953	130816	1	0
TCGA-AR-A1AH-01	ENSG00000129250	5996	130893	1	0
TCGA-B0-5705-01	ENSG00000129250	6262	60613	1	0
TCGA-B6-A3ZX-01	ENSG00000129250	8750	130816	1	0
TCGA-B8-A8YJ-01	ENSG00000188157	36283	255	1	0
TCGA-B9-5155-01	ENSG00000129250	12296	130816	1	0
TCGA-BA-A4IG-01	ENSG00000129250	7327	130816	1	0
TCGA-BC-A3KG-01	ENSG00000129250	5396	130816	1	0
TCGA-BH-A0GY-01	ENSG00000129250	3815	130816	1	0
TCGA-BH-A0W3-01	ENSG00000129250	4587	130816	1	0
TCGA-BH-A18F-01	ENSG00000129250	3435	54397	1	0
TCGA-BH-AB28-01	ENSG00000129250	8074	57861	1	0

Table A.1 continued from previous page

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-BH-AB28-01	ENSG00000138821	708	128689	1	0
TCGA-BK-A0CC-01	ENSG00000129250	5437	131030	1	0
TCGA-BL-A13I-01	ENSG00000215301	10894	65280	1	0
TCGA-BQ-5879-01	ENSG00000129250	5948	130574	1	0
TCGA-BQ-7060-01	ENSG00000188157	24585	962	1	0
TCGA-BT-A20T-01	ENSG00000129250	6856	130816	1	0
TCGA-C5-A1BL-01	ENSG00000129250	3718	119727	1	0
TCGA-C5-A3HF-01	ENSG00000129250	11144	130816	1	0
TCGA-CH-5762-01	ENSG00000129250	5959	130816	1	0
TCGA-CJ-4903-01	ENSG00000129250	5399	130816	1	0
TCGA-CN-4734-01	ENSG00000188157	33272	255	1	0
TCGA-CQ-A4CE-01	ENSG00000215301	22471	126049	1	0
TCGA-CV-7101-01	ENSG00000188157	27497	255	1	0
TCGA-CV-7409-01	ENSG00000129250	5941	130816	1	0
TCGA-CZ-4866-01	ENSG00000129250	6559	130905	1	0
TCGA-D3-A1QB-06	ENSG00000157557	1074	131026	1	0
TCGA-DB-5270-01	ENSG00000177469	1930	129660	1	0
TCGA-DB-5270-01	ENSG00000177469	1930	64636	1	0
TCGA-DB-A64V-01	ENSG00000215301	6446	59252	1	0
TCGA-DB-A64V-01	ENSG00000215301	6446	62565	1	0
TCGA-DD-AAE2-01	ENSG00000215301	12507	55729	1	0
TCGA-DU-7012-01	ENSG00000129250	5638	130816	1	0
TCGA-DX-AB2O-01	ENSG00000129250	6928	130816	1	0
TCGA-E2-A107-01	ENSG00000138821	1030	64504	1	0
TCGA-E2-A156-01	ENSG00000129250	2892	130904	1	0

Table A.1 continued from previous page

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-E2-A15D-01	ENSG00000129250	4767	59214	1	0
TCGA-E2-A15P-01	ENSG00000129250	5233	131005	1	0
TCGA-E2-A1L9-01	ENSG00000129250	3397	130869	1	0
TCGA-EA-A3HQ-01	ENSG00000129250	4337	130816	1	0
TCGA-EB-A41A-01	ENSG00000129250	5024	130892	1	0
TCGA-EB-A42Z-01	ENSG00000129250	4377	130905	1	0
TCGA-EE-A29A-06	ENSG00000129250	3781	130816	1	0
TCGA-EE-A2M8-06	ENSG00000129250	7583	130816	1	0
TCGA-EJ-5494-01	ENSG00000129250	6308	130816	1	0
TCGA-EJ-5532-01	ENSG00000129250	7120	60950	1	0
TCGA-EJ-5532-01	ENSG00000129250	7120	61401	1	0
TCGA-EJ-7125-01	ENSG00000129250	15709	130816	1	0
TCGA-EJ-7218-01	ENSG00000129250	15510	130816	1	0
TCGA-EJ-7793-01	ENSG00000129250	10841	60788	1	0
TCGA-EJ-7793-01	ENSG00000129250	10841	63642	1	0
TCGA-EJ-A46B-01	ENSG00000129250	12339	110892	1	0
TCGA-EK-A2RC-01	ENSG00000123095	1162	131071	1	0
TCGA-EK-A2RE-01	ENSG00000129250	10806	130816	1	0
TCGA-EM-A2CU-01	ENSG00000188157	49681	255	1	0
TCGA-EV-5903-01	ENSG00000188157	48571	255	1	0
TCGA-EW-A1OX-01	ENSG00000163513	846	65535	1	0
TCGA-F4-6856-01	ENSG00000129250	12625	130816	1	0
TCGA-FD-A43U-01	ENSG00000188157	33708	255	1	0
TCGA-FJ-A871-01	ENSG00000129250	9592	131065	1	0
TCGA-FU-A3HY-01	ENSG00000129250	5352	130816	1	0

Table A.1 continued from previous page

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-FU-A3WB-01	ENSG00000129250	7102	130816	1	0
TCGA-FX-A2QS-01	ENSG00000129250	6684	130816	1	0
TCGA-HC-7211-01	ENSG00000129250	6075	106745	1	0
TCGA-HC-7211-01	ENSG00000177469	4106	117632	1	0
TCGA-HC-7821-01	ENSG00000157557	2276	130858	1	0
TCGA-HC-8259-11	ENSG00000188157	3342	128550	0	1
TCGA-HC-8262-01	ENSG00000129250	7803	64173	1	0
TCGA-HT-7474-01	ENSG00000177469	1657	129424	1	0
TCGA-HT-7693-01	ENSG00000177469	1869	131071	1	0
TCGA-HW-7489-01	ENSG00000177469	1110	126849	1	0
TCGA-J7-6720-01	ENSG00000129250	8173	130816	1	0
TCGA-JY-A6FB-01	ENSG00000129250	16257	130816	1	0
TCGA-KK-A8IA-01	ENSG00000129250	3410	130816	1	0
TCGA-KL-8323-01	ENSG00000129250	5726	130648	1	0
TCGA-KL-8341-01	ENSG00000129250	6768	130816	1	0
TCGA-KO-8417-01	ENSG00000157557	2367	128082	1	0
TCGA-KS-A41I-11	ENSG00000188157	9357	63880	0	1
TCGA-L5-A43M-01	ENSG00000188157	46968	255	1	0
TCGA-L5-A4OM-01	ENSG00000129250	8826	130816	1	0
TCGA-NJ-A4YI-01	ENSG00000188157	49770	246	1	0
TCGA-NQ-A638-01	ENSG00000188157	46174	7332	1	0
TCGA-O1-A52J-01	ENSG00000188157	44171	1266	1	0
TCGA-OR-A5L5-01	ENSG00000215301	11024	65366	1	0
TCGA-P4-A5EA-01	ENSG00000129250	6731	130816	1	0
TCGA-Q1-A5R2-01	ENSG00000129250	6674	131018	1	0

Table A.1 continued from previous page

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-QQ-A5VA-01	ENSG00000129250	5217	130816	1	0
TCGA-R5-A7ZI-01	ENSG00000123095	182	130816	1	0
TCGA-S7-A7X2-01	ENSG00000163513	1259	130828	1	0
TCGA-S8-A6BV-01	ENSG00000129250	7389	130874	1	0
TCGA-S9-A6TY-01	ENSG00000163513	592	130816	1	0
TCGA-TT-A6YO-01	ENSG00000215301	20327	61963	1	0
TCGA-UB-AA0U-01	ENSG00000163597	4115	61999	1	0
TCGA-UZ-A9PJ-01	ENSG00000188157	43394	255	1	0
TCGA-UZ-A9PP-01	ENSG00000188157	40637	65564	1	0
TCGA-V1-A9O5-01	ENSG00000123095	367	131056	1	0
TCGA-V1-A9ZR-01	ENSG00000138821	521	116813	1	0
TCGA-VQ-A91V-01	ENSG00000129250	5404	131055	1	0
TCGA-VS-A8EG-01	ENSG00000129250	10647	130054	1	0
TCGA-W2-A7HD-01	ENSG00000129250	605	130737	1	0
TCGA-W2-A7UY-01	ENSG00000129250	1495	130937	1	0
TCGA-XV-A9VZ-01	ENSG00000129250	5591	130816	1	0
TCGA-Z4-AAPG-01	ENSG00000129250	4441	130816	1	0
TCGA-ZH-A8Y2-01	ENSG00000129250	8672	130229	1	0
TCGA-ZH-A8Y2-01	ENSG00000129250	8672	63399	1	0
TCGA-10-0926-01	ENSG00000188157	53176	255	1	0
TCGA-2L-AAQM-01	ENSG00000157557	4465	62184	1	0
TCGA-50-5932-01	ENSG00000215301	10715	126312	1	0
TCGA-AB-2955-03	ENSG00000157514	3845	129101	1	0
TCGA-AN-A0FS-01	ENSG00000121691	4569	60547	1	0
TCGA-AO-A0JL-01	ENSG00000215301	9397	65280	1	0

Table A.1 continued from previous page

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-AQ-A04L-01	ENSG00000129250	2953	130816	1	0
TCGA-BH-A0DE-01	ENSG00000116701	613	41600	1	0
TCGA-BR-6707-01	ENSG00000157514	608	63675	1	0
TCGA-CD-5798-01	ENSG00000188157	36067	255	1	0
TCGA-CV-5966-01	ENSG00000188157	51590	255	1	0
TCGA-DJ-A2Q0-01	ENSG00000157557	1844	124125	1	0
TCGA-DX-A6BE-01	ENSG00000121691	3852	130816	1	0
TCGA-ET-A40S-01	ENSG00000215301	12494	65280	1	0
TCGA-FG-A87N-01	ENSG00000157514	2120	131071	1	0
TCGA-G9-6351-01	ENSG00000215301	8439	124557	1	0
TCGA-J4-A83L-01	ENSG00000215301	11183	112019	1	0
TCGA-KC-A4BL-01	ENSG00000215301	14256	53561	1	0
TCGA-KL-8336-01	ENSG00000215301	8658	122131	1	0
TCGA-KO-8417-01	ENSG00000215301	9231	62181	1	0
TCGA-V1-A9OA-01	ENSG00000138821	533	129315	1	0
TCGA-WB-A81G-01	ENSG00000157557	4375	130816	1	0
TCGA-ZH-A8Y2-01	ENSG00000138821	213	130816	1	0
TCGA-2G-AAGZ-01	ENSG00000215301	7385	56885	1	0
TCGA-2L-AAQM-01	ENSG00000177469	5404	120549	1	0
TCGA-4A-A93W-01	ENSG00000188157	40826	255	1	0
TCGA-78-7540-01	ENSG00000188157	51134	255	1	0
TCGA-A7-A0CH-11	ENSG00000188157	8074	120732	0	1
TCGA-B0-4827-01	ENSG00000188157	39326	148	1	0
TCGA-BH-A18F-01	ENSG00000215301	10719	112517	1	0
TCGA-DB-5270-01	ENSG00000177469	1930	130811	1	0

Table A.1 continued from previous page

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-DS-A5RQ-01	ENSG00000215301	19258	123418	1	0
TCGA-EM-A2CN-01	ENSG00000188157	36828	66176	1	0
TCGA-G9-6351-01	ENSG00000215301	8439	120452	1	0
TCGA-HU-A4H0-01	ENSG00000188157	37986	255	1	0
TCGA-09-0367-01	ENSG00000198911	10408	65359	1	0
TCGA-09-1662-01	ENSG00000129250	5514	130816	1	0
TCGA-2Z-A9JT-01	ENSG00000215301	2091	105495	1	0
TCGA-A2-A1FW-01	ENSG00000129250	1415	130816	1	0
TCGA-A2-A3XW-01	ENSG00000188157	36255	255	1	0
TCGA-A7-A425-01	ENSG00000138821	805	131071	1	0
TCGA-A8-A07W-01	ENSG00000129250	5259	130816	1	0
TCGA-AB-2889-03	ENSG00000129250	558	123910	1	0
TCGA-B3-3925-01	ENSG00000129250	9464	126426	1	0
TCGA-BA-A4IG-01	ENSG00000129250	7327	130816	1	0
TCGA-BH-A1FB-01	ENSG00000129250	3977	64565	1	0
TCGA-BH-AB28-01	ENSG00000138821	708	129362	1	0
TCGA-BP-5192-01	ENSG00000138821	3127	129504	1	0
TCGA-C5-A7CG-01	ENSG00000215301	11934	59902	1	0
TCGA-CV-7409-01	ENSG00000129250	5941	130816	1	0
TCGA-D8-A27I-01	ENSG00000129250	5949	128407	1	0
TCGA-DX-A3UD-01	ENSG00000215301	8891	126774	1	0
TCGA-EE-A3AD-06	ENSG00000129250	3712	130816	1	0
TCGA-EJ-7125-11	ENSG00000188157	1287	130285	0	1
TCGA-EK-A2RE-01	ENSG00000129250	10806	130907	1	0
TCGA-EM-A2CU-01	ENSG00000188157	49681	66722	1	0

Table A.1 continued from previous page

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-EX-A1H5-01	ENSG00000138821	343	127576	1	0
TCGA-HM-A4S6-01	ENSG00000188157	39038	255	1	0
TCGA-K4-A4AB-01	ENSG00000215301	17126	125925	1	0
TCGA-KL-8341-01	ENSG00000215301	8609	64178	1	0
TCGA-LI-A9QH-01	ENSG00000215301	7111	118693	1	0
TCGA-QR-A6H0-01	ENSG00000129250	7058	65104	1	0
TCGA-RL-AAAS-01	ENSG00000215301	12649	124783	1	0
TCGA-RN-AAAQ-01	ENSG00000138821	244	130919	1	0
TCGA-TM-A84G-01	ENSG00000215301	8974	124761	1	0
TCGA-X7-A8M6-01	ENSG00000215301	2778	65056	1	0
TCGA-ZP-A9CY-01	ENSG00000215301	8080	64613	1	0
TCGA-25-2404-01	ENSG00000129250	5181	130816	1	0
TCGA-2G-AAGZ-01	ENSG00000215301	7385	61365	1	0
TCGA-85-A513-01	ENSG00000188157	52754	255	1	0
TCGA-A4-7584-01	ENSG00000215301	9715	52718	1	0
TCGA-B6-A0I9-01	ENSG00000138821	665	130996	1	0
TCGA-BC-A3KG-01	ENSG00000215301	11642	63574	1	0
TCGA-BH-A0AZ-01	ENSG00000129250	4581	130816	1	0
TCGA-BH-A0DV-01	ENSG00000121691	7409	129279	1	0
TCGA-BH-A1FL-01	ENSG00000215301	13228	126204	1	0
TCGA-BH-A204-01	ENSG00000129250	716	130816	1	0
TCGA-BK-A0CC-01	ENSG00000138821	483	131071	1	0
TCGA-DK-A3X2-01	ENSG00000129250	4498	130816	1	0
TCGA-DK-A6B0-01	ENSG00000215301	9967	59006	1	0
TCGA-E2-A156-01	ENSG00000215301	10569	65400	1	0

Table A.1 continued from previous page

Sample ID	Ensemble ID	Expression Value	Expression Value After Attack	Prediction	Prediction After Attack
TCGA-EM-A4FH-01	ENSG00000157557	2931	130750	1	0
TCGA-EX-A1H5-01	ENSG00000138821	343	125364	1	0
TCGA-FG-7641-01	ENSG00000215301	12547	122417	1	0
TCGA-G4-6294-01	ENSG00000215301	18214	123355	1	0
TCGA-HC-A6AQ-01	ENSG00000215301	4434	123856	1	0
TCGA-HT-8105-01	ENSG00000215301	11003	64510	1	0
TCGA-HW-7489-01	ENSG00000177469	1110	130947	1	0
TCGA-KL-8339-01	ENSG00000215301	9972	120787	1	0
TCGA-L5-A8NG-01	ENSG00000138821	793	63201	1	0
TCGA-MQ-A6BS-01	ENSG00000215301	11981	62466	1	0
TCGA-NJ-A4YI-01	ENSG00000188157	49770	255	1	0
TCGA-OR-A5L9-01	ENSG00000215301	9404	125923	1	0
TCGA-PE-A5DC-01	ENSG00000129250	3086	130816	1	0
TCGA-VP-A87B-01	ENSG00000215301	5596	50151	1	0
TCGA-WB-A817-01	ENSG00000215301	8520	121014	1	0
TCGA-ZH-A8Y2-01	ENSG00000138821	213	130911	1	0

## **B** **STATISTICAL DATA OF GENE EXPRESSION VALUES OF** **AGRIN, CATALASE AND KINESIN-LIKE PROTEIN GENES** **ACCORDING TO CANCER TYPES**

---

As a result of the applied one-pixel attack, 13 critical genes that are crucial in predicting whether the neural network is tumor or normal were determined. Among these 13 genes determined, the gene with the highest availability (CAT) with the DEG method, the gene with the most effect as a result of the one-pixel attack (AGRIN), and the analysis of the expression data of the DEG method and the rarely encountered genes (KIF1C) in the literature are Shared in Table B.2, Table B.1, Table B.3.

**Table B.1** canSAR Black expression analysis of AGRN gene. For various cancer stuides, distribution of KIF1C expression in Normal and Tumor samples is summarized. Distribution of expression is summarized with maximum, minimum and median values per sample.

Study	Tumor Samples	Max	Median	Min	Normal Samples	Max	Median	Min	Log2 (Tvs N)
Thymoma (THYM)	117	8.52	5.32	2.33	337	2.68	-2.24	-0.04	7.56
B-cell lymphoma (DLBC)	41	4.28	2.41	1.21	337	2.68	-2.24	-0.04	4.66
Ovarian (OV)	419	9.90	6.90	-9.97	88	4.64	3.22	1.58	3.68
Glioblastoma (GBM)	159	7.78	5.90	1.99	216	6.47	2.83	-0.27	3.07
Glioma (LGG)	509	8.02	5.80	3.80	216	6.47	2.83	-0.27	2.97
Pheochromocytoma & Paraganglioma (PCPG)	167	6.80	4.37	0.42	128	3.79	2.00	-9.97	2.38
Esophageal (ESCA)	158	8.99	6.98	2.26	271	6.40	4.66	-9.97	2.32
Uterine endometrial (UCEC)	180	8.82	6.61	3.40	78	5.96	4.34	2.21	2.27
Uterine carcinosarcoma (UCS)	57	8.35	6.49	3.90	78	5.96	4.34	2.21	2.16
Stomach (STAD)	389	8.71	5.94	2.16	174	5.88	3.89	-9.97	2.05
Pancreatic (PAAD)	175	8.72	6.29	2.48	167	6.78	4.36	-9.97	1.93
Cervical (CESC)	297	8.51	6.42	2.87	10	6.15	4.67	2.85	1.74
Papillary kidney (KIRP)	258	9.70	7.08	3.33	28	7.99	5.68	-9.97	1.40
Head and neck (HNSC)	443	8.66	6.47	3.37	55	6.47	5.13	2.90	1.34
Melanoma (SKCM)	414	8.68	5.90	1.74	323	5.65	4.66	2.49	1.24
Colon (COAD)	276	7.31	5.26	1.07	167	5.17	4.04	-9.97	1.22
Rectal (READ)	82	6.62	5.13	2.83	141	5.32	4.01	-9.97	1.12
Lung adenocarcinoma (LUAD)	505	8.58	6.33	1.16	288	6.79	5.41	-9.97	0.92
Clear cell kidney (KIRC)	527	8.65	6.55	1.54	28	7.99	5.68	-9.97	0.87
Liver (LIHC)	345	9.24	4.56	1.10	110	7.42	3.70	1.84	0.85
Breast (BRCA)	1067	7.88	5.65	-0.78	179	7.16	4.84	1.31	0.81
Bladder (BLCA)	405	8.95	6.57	2.78	9	7.17	5.88	2.57	0.69
Lung squamous (LUSC)	494	9.44	6.01	2.17	288	6.79	5.41	-9.97	0.61
Adrenocortical (ACC)	75	7.93	2.51	-0.20	128	3.79	2.00	-9.97	0.51
Thyroid (THCA)	502	8.20	6.66	3.47	279	7.50	6.35	-9.97	0.30
Prostate (PRAD)	495	7.18	5.07	1.58	100	7.11	5.04	2.36	0.03
Cholangiocarcinoma (CHOL)	36	8.82	7.06	2.58	0				
Mesothelioma (MESO)	87	9.90	7.11	4.36	0				
Sarcoma (SARC)	205	8.27	4.62	2.07	0				
Uveal melanoma (UVM)	78	7.42	5.52	3.48	0				
Chromophobe renal cell (KICH)	66	7.74	4.56	2.75	28	7.99	5.68	-9.97	-1.12
Testicular (TGCT)	125	7.20	3.85	0.86	165	6.63	5.12	3.71	-1.27

**Table B.2** canSAR Black expression analysis of CAT gene. For various cancer stuides, distribution of KIF1C expression in Normal and Tumor samples is summarized. Distribution of expression is summarized with maximum, minimum and median values per sample.

Study	Tumor Samples	Max	Median	Min	Normal Samples	Max	Median	Min	Log2 (T vs N)
Glioblastoma (GBM)	159	6.39	5.51	2.81	216	4.60	3.24	-9.97	2.27
Glioma (LGG)	509	7.25	5.28	2.08	216	4.60	3.24	-9.97	2.03
Chromophobe renal cell (KICH)	66	9.35	7.27	4.29	28	7.63	6.06	-9.97	1.21
Esophageal (ESCA)	158	7.34	5.21	3.59	271	6.17	4.46	-9.97	0.75
Pancreatic (PAAD)	175	7.21	5.36	2.33	167	5.87	4.91	-9.97	0.45
Clear cell kidney (KIRC)	527	9.12	6.51	2.74	28	7.63	6.06	-9.97	0.44
Thymoma (THYM)	117	7.53	5.85	2.56	337	7.93	5.44	1.65	0.42
Stomach (STAD)	389	7.78	5.60	1.94	174	6.56	5.40	-9.97	0.20
Prostate (PRAD)	495	7.00	5.80	3.10	100	6.73	5.63	4.31	0.17
Thyroid (THCA)	502	8.72	7.00	3.08	279	8.55	6.91	-9.97	0.08
Rectal (READ)	82	6.71	5.65	3.55	141	8.00	5.64	-9.97	0.01
Cholangiocarcinoma (CHOL)	36	7.91	5.72	4.23	0				
Mesothelioma (MESO)	87	6.28	4.54	2.21	0				
Sarcoma (SARC)	205	8.61	5.19	2.44	0				
Uveal melanoma (UVM)	78	5.51	3.82	1.15	0				
Papillary kidney (KIRP)	258	8.53	5.98	3.43	28	7.63	6.06	-9.97	-0.09
Ovarian (OV)	419	7.45	5.33	-9.97	88	6.28	5.44	4.14	-0.11
Colon (COAD)	276	7.29	5.61	2.45	167	6.71	5.80	-9.97	-0.19
Testicular (TGCT)	125	5.64	3.62	1.28	165	6.15	3.86	2.46	-0.24
Liver (LIHC)	345	9.31	7.25	3.75	110	8.93	7.50	2.88	-0.25
B-cell lymphoma (DLBC)	41	6.07	4.72	3.55	337	7.93	5.44	1.65	-0.71
Uterine endometrial (UCEC)	180	7.80	5.17	2.30	78	6.56	5.93	4.26	-0.76
Lung adenocarcinoma (LUAD)	505	7.78	5.53	2.89	288	7.64	6.45	-9.97	-0.92
Uterine carcinosarcoma (UCS)	57	7.98	4.95	2.88	78	6.56	5.93	4.26	-0.98
Head and neck (HNSC)	443	7.65	4.57	0.55	55	6.84	5.58	4.34	-1.01
Lung squamous (LUSC)	494	7.27	5.14	1.88	288	7.64	6.45	-9.97	-1.31
Melanoma (SKCM)	414	7.75	4.88	1.26	323	7.24	6.23	3.19	-1.34
Breast (BRCA)	1067	8.87	5.55	-0.41	179	8.54	6.97	4.79	-1.42
Bladder (BLCA)	405	8.08	5.17	2.61	9	7.14	6.65	5.27	-1.49
Cervical (CESC)	297	7.08	4.51	0.90	10	6.81	6.18	4.58	-1.68
Adrenocortical (ACC)	75	7.19	5.03	-0.13	128	7.69	6.70	-9.97	-1.68
Pheochromocytoma & Paranglioma (PCPG)	167	6.60	3.61	-0.45	128	7.69	6.70	-9.97	-3.09

**Table B.3** canSAR Black expression analysis of KIF1C gene. For various cancer stuides, distribution of KIF1C expression in Normal and Tumor samples is summarized. Distribution of expression is summarized with maximum, minimum and median values per sample.

Study	Tumor Samples	Max	Median	Min	Normal Samples	Max	Median	Min	Log2 (T vs N)
Thymoma (THYM)	117	6.11	4.31	0.28	337	4.61	1.89	-0.04	2.42
B-cell lymphoma (DLBC)	41	5.84	3.97	1.63	337	4.61	1.89	-0.04	2.07
Glioma (LGG)	509	7.85	5.96	2.51	216	6.92	5.10	-9.97	0.86
Papillary kidney (KIRP)	258	7.91	6.08	1.78	28	6.17	5.28	-9.97	0.81
Pancreatic (PAAD)	175	6.68	5.30	2.84	167	5.52	4.64	-9.97	0.66
Stomach (STAD)	389	7.82	5.83	2.80	174	8.12	5.39	-9.97	0.44
Liver (LIHC)	345	6.92	4.89	1.91	110	6.37	4.67	1.90	0.22
Head and neck (HNSC)	443	8.11	5.71	2.34	55	7.03	5.61	4.49	0.10
Glioblastoma (GBM)	159	6.98	5.16	1.05	216	6.92	5.10	-9.97	0.05
Cholangiocarcinoma (CHOL)	36	6.70	5.62	3.11	0				
Mesothelioma (MESO)	87	6.83	5.43	3.32	0				
Sarcoma (SARC)	205	7.77	5.35	2.91	0				
Uveal melanoma (UVM)	78	7.18	5.87	3.87	0				
Chromophobe renal cell (KICH)	66	6.68	5.27	1.81	28	6.17	5.28	-9.97	-0.00
Prostate (PRAD)	495	7.57	5.75	2.83	100	7.32	5.75	4.56	-0.01
Testicular (TGCT)	125	7.52	5.56	1.51	165	6.25	5.71	4.09	-0.16
Clear cell kidney (KIRC)	527	7.03	5.04	1.56	28	6.17	5.28	-9.97	-0.23
Thyroid (THCA)	502	7.04	5.74	3.26	279	7.21	6.07	-9.97	-0.33
Breast (BRCA)	1067	7.64	5.20	0.03	179	7.14	5.58	3.82	-0.38
Esophageal (ESCA)	158	7.40	5.76	3.56	271	7.34	6.20	-9.97	-0.44
Ovarian (OV)	419	6.76	4.58	-9.97	88	6.75	5.12	3.65	-0.55
Colon (COAD)	276	6.87	5.47	1.27	167	8.17	6.02	-9.97	-0.55
Melanoma (SKCM)	414	7.52	5.53	2.46	323	8.72	6.16	4.78	-0.64
Cervical (CESC)	297	7.00	5.08	2.07	10	6.27	5.78	5.43	-0.69
Rectal (READ)	82	6.85	5.37	2.74	141	7.81	6.29	-9.97	-0.93
Uterine endometrial (UCEC)	180	6.70	4.80	2.27	78	7.32	5.90	4.75	-1.10
Uterine carcinosarcoma (UCS)	57	6.37	4.75	2.56	78	7.32	5.90	4.75	-1.15
Adrenocortical (ACC)	75	6.35	3.51	0.03	128	5.67	4.69	-9.97	-1.17
Lung adenocarcinoma (LUAD)	505	6.87	4.98	1.02	288	7.67	6.37	-9.97	-1.39
Bladder (BLCA)	405	7.15	5.05	1.49	9	7.67	6.50	5.24	-1.44
Lung squamous (LUSC)	494	7.45	4.79	1.82	288	7.67	6.37	-9.97	-1.58
Pheochromocytoma & Paraganglioma (PCPG)	167	4.86	2.83	-0.55	128	5.67	4.69	-9.97	-1.86

## PUBLICATIONS FROM THESIS

---

### Papers

1. Convolutional Neural Network Approach to Predict Tumor Samples Using Gene Expression Data

\*Darendeli, B , Yılmaz, A . (2021). Convolutional Neural Network Approach to Predict Tumor Samples Using Gene Expression Data . Journal of Intelligent Systems: Theory and Applications , 4 (2) , 136-141 . DOI: 10.38016/jista.946954

### Conference Papers

1. 1.Temel Onkoloji Sempozyumu  
May 09-11, 2018, Izmir, Turkey  
Kanser Teşhisinde Derin Öğrenme ile Sınıflandırma
2. The International Conference on Advanced Engineering, Technology and Applications (ICAETA-2021)  
July 09-10, 2021, Istanbul, Turkey  
Convolutional Neural Network Approach to Distinguish and Characterize Tumor Samples Using Gene Expression Data

### Awards

1. Best Poster Award 3rd Place  
1.Temel Onkoloji Sempozyumu  
May 09-11, 2018, Izmir, Turkey  
Kanser Teşhisinde Derin Öğrenme ile Sınıflandırma