**NEAR EAST UNIVERSITY**
**INSTITUTE OF GRADUATE STUDIES**
**DEPARTMENT OF BIOMEDICAL ENGINEERING**

# *BRCA* VARIATIONS RISK ASSESSMENT IN BREAST CANCERS USING DIFFERENT ARTIFICIAL INTELLIGENCE MODELS

**Ph.D. THESIS**

**Niyazi ŞENTÜRK**

**Supervisor**
**Assoc. Prof. Mahmut Çerkez ERGÖREN**

**Nicosia**
**December, 2021**

# Approval

We certify that we have read the thesis submitted by Niyazi ŞENTÜRK titled "***BRCA* Variations Risk Assessment in Breast Cancers using Different Artificial Intelligence Models**" and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy in Biomedical Engineering.

| Examining Committee | Name-Surname | Signature |
|---|---|---|
| Head of the Committee: | Prof. Dr. Tulin Bodamyalı | |
| Committee Member: | Prof. Dr. Terin Adalı | |
| Committee Member: | Assoc. Prof. Pınar Tulay | |
| Committee Member: | Assist. Prof. Dr. Özel Yürüker | |
| Supervisor: | Assoc. Prof. Mahmut Çerkez Ergören | |

Approved by the Head of the Department

…../…../20…

Prof. Dr. Ayşe Günay Kibarer

Head of Department

Approved by the Institute of Graduate Studies

…../…../20…

Prof. Dr. Kemal Hüsnü Can Başer

Head of the Institute

## Declaration

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

Niyazi Şentürk

16/12/2021

## Acknowledgments

First of all, I would like to thank my supervisor Assoc. Prof. Mahmut Çerkez Ergören for his supervision, support, and sharing his knowlege with me during my thesis work.

I would like to thank Prof. Munis Dündar (Erciyes University) and Assoc. Prof. Şehime G. Temel (Bursa Uludağ Unversity) for their support and help. I also would like to thank Prof. Dr. Terin Adalı (Near East University) for her support and help.

For the last, I would like to thank my beloved family for their trust, support, help and unconditional love. They are biggest moral support for me in this thesis.

**Niyazi Şentürk**

**To my dear mom…**

## Abstract

### *BRCA* Variations Risk Assessment in Breast Cancers using Different Artificial Intelligence Models

**Şentürk, Niyazi**

**PhD, Department of Biomedical Engineering**

**December, 2021, 82 pages**

Artificial intelligence provides modeling on machines by simulating the human brain using learning and decision-making abilities. Early diagnosis is highly effective to reduce mortality in cancer. This study aimed to combine cancer associated risk factors including genetic variations and design an artificial intelligence system for risk assessment.

A total 268 breast cancer patients' data have been analyzed for sixteen different risk factors including genetic variation classifications. A total of 61 *BRCA1*, 128 *BRCA2* and 11 both *BRCA1* and *BRCA2* genes associated breast cancer patients' data were used to train the system. Mamdani's Fuzzy Inference Method and Feedforward Neural Network Method were used as model software on MATLAB.

Six different tests were performed with twelve different subjects who had not been introduced to the system before. We obtained 99.9% train success, 99.6% validation success 99.7% test success and overall 99.9% whole system. The software showed 95.5% accuracy when new patients were tested.

To conclude, the developed artificial intelligence models gave successful results for risk assessment of breast cancer. Moreover, *BRCA1* and *BRCA2*-associated breast cancer was firstly studied in these models. On the other hand, the system also classifies pathogenicity of *BRCA1* and *BRCA2* variants using patients' clinical data. Thus, the generated Mamdani's fuzzy logic and feed-forward neural network systems will become a good source of identification of variant's status for breast cancer diagnosis. Thus, these systems may provide a better understanding of the effects of genetic factors on breast cancer and a unique tool for personalized medicine software. Finally, the innovations it adds to the current literature are that it mainly focuses on the genetic data associated with breast cancer and the developed

model makes prediction in a wider perspective using more risk factors than similar studies in the literature.

***Keywords:*** Breast cancer, *BRCA1, BRCA2,* variation, artificial intelligence, translational fuzzy logic

# Özet


## Farklı Yapay Zekâ Modelleri Kullanarak Meme Kanserinde *BRCA* Varyasyonlarının Risk değerlendirilmesi

**Şentürk, Niyazi**

**PhD, Biyomedikal Mühendisliği Bilim Dalı**

**Aralık, 2021, 82 sayfa**

Yapay zekâ, öğrenme ve karar verme yeteneklerini kullanarak insan beynini simüle ederek makineler üzerinde modelleme sağlar. Erken teşhis kanserde mortaliteyi azaltmada oldukça etkilidir. Bu çalışma, genetik varyasyonlar dahil olmak üzere kanserle ilişkili risk faktörlerini birleştirmeyi ve risk değerlendirmesi için yapay zekâ sistemi tasarlamayı amaçladı.

Toplam 268 meme kanseri hastasının verileri, genetik varyasyon sınıflandırmaları da dahil olmak üzere on altı farklı risk faktörü ile analiz edildi. Sistemi eğitmek için meme kanseri hastalarının verileriyle ilişkili toplam 61 *BRCA1,* 128 *BRCA2* ve 11 hem *BRCA1* hem de *BRCA2* geni kullanıldı. MATLAB üzerinde model yazılım olarak Mamdani'nin Bulanık Çıkarım Yöntemi ve İleri Beslemeli Sinir Ağı Yöntemi kullanılmıştır.

Daha önce sisteme tanıtılmamış on iki farklı denek ile altı farklı test yapılmıştır. %99,9 eğitim başarısı, %99,6 doğrulama başarısı %99,7 test başarısı ve genel olarak %99,9 tüm sistem elde ettik. Yazılım, yeni hastalar test edildiğinde %95,5 doğruluk gösterdi.

Sonuç olarak, geliştirilen Mamdani'nin Bulanık Mantık ve İleri Beslemeli Sinir Ağı modelleri meme kanserinin risk değerlendirmesi için başarılı sonuçlar verdi. Ayrıca *BRCA1* ve *BRCA2* ile ilişkili meme kanseri ilk olarak bu modellerde incelenmiştir. Öte yandan sistem, hastaların klinik verilerini kullanarak *BRCA1* ve *BRCA2* varyantlarının patojenitesini de sınıflandırır. Böylece, oluşturulan bulanık mantık ve sinir ağı sistemleri, meme kanseri teşhisi için varyantın durumunun tanımlanması için iyi bir kaynak olacaktır. Böylece bu sistemler, genetik faktörlerin meme kanseri üzerindeki etkilerinin daha iyi anlaşılmasını ve kişiselleştirilmiş ilaç yazılımı için benzersiz bir araç sağlayabilir. Son olarak mevcut literatüre kattığı

yenilikler ise ağırlıklı olarak meme kanseri ile ilişkili genetik verilere odaklanması ve geliştirilen modelin literatürdeki benzer çalışmalara göre daha fazla risk faktörü kullanarak daha geniş bir perspektifte tahminde bulunmasıdır.

***Anahtar Kelimeler:*** Meme kanseri, *BRCA1*, *BRCA2*, varyasyon, yapay zekâ, translasyonel bulanık mantık

# Table of Contents

CHAPTER II



CHAPTER III

## List of Tables

## List of Figures

## List of Abbreviations

| | |
|---|---|
| **AI:** | Artificial Intelligence |
| **ANN:** | Artificial Neural Networks |
| **APC:** | Adenomatous Polyposis Coli |
| **BARD1:** | BRCA1 Associated RING Domain 1 |
| **BMI:** | Body Mass Index |
| **BRCA1:** | Breast Cancer Type 1 Susceptibility Protein |
| **BRCA2** | Breast Cancer Type 2 Susceptibility Protein |
| **BRIP1:** | BRCA1 Interacting Helicase 1 |
| **CD95:** | Custer of Differentiation 95 |
| **CDH1:** | Cadherin 1 |
| **CHEK2:** | Checkpoint Kinase 2 |
| **DNA:** | Deoxyribonucleic Acid |
| **FNA:** | Fine Needle Aspirate |
| **HER2/neu:** | Human Epidermal Growth Factor Receptor 2 |
| **MF:** | Membership Function |
| **NBN:** | Nibrin |
| **PALB2:** | Partner and Localizer of *BRCA2* |
| **PARP:** | Poly (ADP-Ribose) Polymerase |
| **PRS:** | Polygenic Risk Scores |
| **Rb:** | Retinoblastoma |
| **RNA:** | Ribonucleic Acid |
| **SNPs:** | Single Nucleotide Polymorphisms |
| **STK11:** | Serine/Threonine Kinase 11 |
| **TP53:** | Tumor Protein p53 |
| **VUS:** | Variant of Unknown Significance |
| **WHO:** | World Health Organization |

# CHAPTER I
# Introduction

## Introduction

Artificial intelligence (AI) has been used extensively in the field of health in recent years. In the field of oncology, emerging AI technologies can detect tumors, diagnose cancers, and even generate chemotherapy treatment recommendations that are adjusted in real time based on patient responses. Integration of AI approaches such as machine learning and fuzzy logic transforms big data into clinically actionable knowledge and becomes the foundation of precision medicine. Especially, AI has been continuing to improve characterizations in genetic and molecular medicine since the first day by providing machine learning and knowledge management. This has given rise to evidence-based computerized diagnostic tools, intended to aid the physician in making primary medical decisions and hence an early diagnosis which helps reduce the treatment time or increase survival rate.

Cancer can develop anywhere in the body. The disease begins with the uncontrolled multiplication of cells. There are more than 200 types of cancer. That can begin in the lungs, breast, large intestine and even blood. Although cancers are similar in many ways, they differ in growth and spread (Wasif, et al., 2019).

All the cells in our bodies have certain tasks. Normal cells divide regularly. When they get old or damaged, they die and new cells take their place. Cancer occurs when cells start to multiplication by uncontrolled. Cancer cells continue to grow and form new cells. They do not leave place for normal cells. This causes problems in the body part where the cancer begins. Cancer cells can also spread to other parts of the body. For example, cancer cells in the lung can spread to bones and grow there. The spread of cancer cells is called metastasis. Lung cancer is also called lung cancer when it spreads to the bones. According to doctors, cancer cells in the bones look just like those in the lungs. However, the disease is not called bone cancer unless it starts in the bones (Wasif, et al., 2019).

One of the heterogeneous disease breast cancer is the cause of the most common cancer death in women globally and affects one in eight women (Wolfe, et al., 2015).

Molecular, pathological and clinical characteristics rise breast cancer a challenging diseases to handle (Ellsworth, et al., 2010). However, early detection of the breast cancer is an effective method to reduce mortality (Rue, et al., 2009). Despite its complex aetiology, breast cancer is affected by both environmental and genetic factor (Ellsworth, et al., 2010). On the other hand, it is well known that cancer is the most common genetic diseases resulting from the accumulation of genetic variations known as somatic and germline.

Billions of genetic defects as mutations, occur in genomic material of cells, and most of them are repaired or destroyed by protective genes. Two genes, *BRCA1* and *BRCA2*, are known as crucial to recognize mutations and DNA molecules that are responsible for repaie. These help to stabilize DNA in normal cells and prevent uncontrolled cell growt named as tumor suppressive properties. However, pathogenic mutations within these two genes abolish the suppressive function, increasing the susceptibility to familial autosomal dominant breast and ovarian cancers. Therefore, identification of genetics variations within the *BRCA* gene family is crucial for early diagnoses (De Silva, et al., 2019). Thus, the aim this current study to develop a variant selection method based on a artificial intelligence system and classification algorithms, and to validate the variant signatures obtained on *BRCA1* and *BRCA2* positive breast cancer patient cohorts.

To best in our knowledge, these are the first artificial intelligence approaches, Mamdani's fuzzy logic and feed-forward artificial neural network, that include *BRCA1* and *BRCA2*-associated breast cancers in model systems. On the other hand, systems are trained to assess the pathogenetic risk of the variants, especially variants with unknown significance, using patients' other risk factors. This system may promisingly provide a better understanding of the effects of genetic factors as well as the classification of the genetic variants in breast cancer and will be a good candidate for unique personalized medicine software.

**Cancer: Abnormal Cell Growth**

Cancer cells can be divided as benign tumors or malignant tumors. Benign tumors are not cancerous. Operated benign tumors generally do not grow back, whereas malignant tumors sometimes do. Cells with benign tumors does not metastasis, most importantly, they are rarely life-threatening. Cancerous malignant

tumors are abnormal in shaped, showing uncontrolled and irregularly prolifiration (Wu, et al., 2016).

### *The Incidance of the World's Second Leading Cause of Death*

Cancer is the second leading cause of death, globally. According to the World Health Organization (WHO) in 2020, cancer caused 10 million deaths. The WHO report indicated that the most common causes of cancer death were lung (1.80 million), colorectal (935.000), liver (830.000), stomach (769.000), and breast (685.000) (https://www.who.int/). Cancer deaths vary according to development clasification of countries; for example, 70% of cancer deaths occur in low and middle-income countries. Approximately, one-third of cancer deaths are mainly caused by five behavioral and nutritional risks: high level of body mass index (BMI), low fruit and vegetable intake, lack of physical activity, tobacco use and alcohol use. Tobacco use is the most important risk factor for cancer pathogenesis which included approximately 22% of cancer deaths in 2015 (GBD 2015 Risk Factors Collaborators, 2016). The economic impact of cancer is important and increasing. In 2010, the total annual economic cost of cancer is estimated to be approximately US $ 1.16 trillion. Therefore, only one in five low and middle income countries have the essential data to drive cancer policy. (Stewart & Wild, 2014).

### *Breast Cancer Development*

Breast cancer develops from cells that cover the mammary glands or the channel that carry breast milk. Genetic predispositions and mutations may increase the risk of developing the breast cancer. There are many types of breast cancer: invasive ductal carcinoma, invasive lobular carcinoma, Paget's disease of the nipple, inflammatory breast cancer, Phyllodes tumors of the breast, locally advanced breast cancer, metastic breast cancer.

Metastases to lymph nodes are the most common in breast cancers. However, other tissue metastatic can be observed. Breast cancer is 100 times more seen in women than men. Despite having more breast tissue puts every women in a risk. Every woman is at risk of developing breast cancer. If breast cancer is detected early before it spreads, the patient has a 96% chance of survival. Every year, approximately 44.000 women die from breast cancer (Kang, et al., 2019).

Although breast cancer is seen in men, having more breast tissue in women increases cases 100 times more than men. There has been an increase in the incidence of breast cancer since the 1970s due to moving modern Western lifestyle. According to studies, Northern European women have the most risk for developing breast cancer as with the highest mortality rate (22.6 per 100 thousand) among the other geographical regions. On the other hand, Chinese and Japanese women have the lowest mortality rate (5.6 per 100 thousand in China and 8.3 per 100 thousand in Japan, respectively). One in every 8 women in the United States throughout her life, (Bray, et al., 2018).

***Breast Cancer Aetiology and Risk Factors***

There are many different factors that are responsible for breast cancer pathogenesis, for instance; gender, viral pathogens, enviromental, hormonal and genetics. A study showed that advanced age in women is the important risk factor (Açıkgöz & Yıldız, 2017). Other important cause is genetic predisposition or family history. The presence of a cancer family history increases the cancer risk 1.9 fold more than a patient without a family history. However, this ratio rises up to 2.1 fold in the first degree relatives. 5-10% of breast cancer germline mutation can be seen in *BRCA1* and *BRCA2* genes. Women with *BRCA1* and *BRCA2* mutations have a 40-85% lifetime risk of breast cancer (Açıkgöz & Yıldız, 2017). Thirdly, factors such as early menarche, late menopause and hormones used after menopause increase the risk of breast cancer. Fourthly, the effects of obesity on the risk of breast cancer are affected by the menopausal condition. High BMI has a significant protective effect while pre-menopausal but, it positively correlates with post-menopausal breast cancer risk. Furthermore, high levels of alcohol and tobacco consumption also increase the risk of breast cancer. Lastly, breast cancer incidence and mortality risk also differ in geographical structure, culture, race, ethnicity and socio-economic status. Compared to other races, white women have a higher incidence of breast cancer (Açıkgöz & Yıldız, 2017).

There are many factors that cause cancer. These factors are smoking, alcohol, exposure to radiation/ X-ray, some viruses, nutritions, long term exposure to sunlight, some chemicals (tar, gasoline, dye stuff, asbestos, etc.), air pollution, and genetic. Also, being a woman, being 55 years or older, if one of the close relatives

has breast cancer, and having started menstruation before the age of 12 increases the risk of cancer. It also increase the risk of breast cancer if they have never been pregnant or if they have continued their menstruation after the age of 50 (Smith, et al., 2015).

In 2018, Rudolph and her indicated that polygenic risk scores (PRS) for breast cancer can be used to differentiate the population into different risk levels. They combined PRS and environmental risk factors to work out a more accurate risk estimate (Rudolph, et al., 2018). In 2018, Lilyquist and her colleagues worked on one study about genetic factor of breast cancer. Breast cancer is the most common cancer among women in the United States. Also, 30% of the cases in which they worked were diagnosed with breast cancer in the family. Therefore, 18% of familial breast cancer risk is associate with SNPs. So, they focused on SNPs. As a result of their study, they have reached findings of general breast cancer, pathological subtypes and mutation carriers (*BRCA1*, *BRCA2* and *CHEK2*) (Lilyquist, et al., 2018).

***Pathology of Breast Cancer***

Cytology is the process of taking a cell from a suspicious lesion in the breast for examination and pathological diagnosis. And, tissue removal is called biopsy. As a result of these procedures, the obtained cell or tissue is examined by the pathology specialist under the microscope, the diagnosis is made and a pathology report is prepared. According to the result of the pathology report, the treatment of the patient is determined. There are different methods. The method depends on the type of the suspected lesion (Cheung, et al., 2015).

First method is fine-needle aspiration cytology. In this method, removing fluid or cells from a suspicious lesion with a syringe. It is a very safe small surgical method. Mostly, it is preferred for evacuation of breast cysts and biopsy of armpit lymph nodes. In addition, biopsies can be performed for very small masses or deep-seated masses in the breast. The second method is Tru-cut biopsy. A small tissue is removed from the suspected area of breast cancer using a thick needle. This procedure may be performed by the surgeon with local anesthesia or by the radiologist under ultrasound guidance. Lastly, the removal of some or all of the suspected breast tissue is called a surgical biopsy. Very small tumors that can be detected by radiological methods such as mammography are marked with wire by the

radiologist and the marked area is removed by the surgeon in the operating (Cheung, et al., 2015).

***Molecular Mechanisms of Breast Cancer***

Cancer is considered a genetic disease, mainly because it concerns the genetic material of the cell, whether it is hereditary or environmental factors. The core mechanisms of carcinogenesis are non-lethal genetic damage in the cell. Although malignant tumors generally differ in terms of macroscopic, microscopic and genetic material, they share some characteristics that determine their growth and biological behavior among themselves. DNA damage can be caused by environmental factors such as chemicals, radiation, viruses. Also, it can be caused by germline cells (Ayhan, 2013). Breast cancer is a complex disease and breast tumors are heterogeneous tumors with different characteristics, making it difficult to treat.

Cancer is a genetic disorder caused by a defect in genes that encode molecules responsible for cell communication processes. If a gene defect causes cancer in one organ, it can cause cancer in other organs, regardless of histological and anatomical status. Cancer cells that do not take into account the regulatory processes in the body begin to grow uncontrolled. Genes called oncogen play a fundamental role in tumor formation. Oncogenes are a defective gene responsible for the production of abnormal protein that can convert a normal cell into a cancerous cell. There are also genes that prevent tumor formation. These are tumor inhibitory or suppressor genes (Feng, et al., 2019).

Anti-oncogen or a tumor suppressor gene is a gene that regulates cells during cell division and replication Uncontrolled growing cells can cause cancer, a mutation in these suppressor genes may cause abnormal cell growth in the body. Loss of function of these genes can be more important in the development of cancer than activation of oncogenes. Some important tumor suppressor genes are *TP53*, *Rb*, *APC*, *CD95* (McGrath, et al., 2019).

Recent studies have shown that microRNAs have very important functions in cells and these molecules play a role in the development of breast cancer. miRNA chips' have begun use to detect target miRNAs that may cause breast cancer, but there are limited studies in this area (Öztemur, et al., 2015).

**Role of Human Genetic Variations in the Breast Cancer Development.** Genetic factors may play a role in the formation of certain types of cancer. Those with a family history of breast or ovarian cancer need to be more careful about scans. This risk can be prevented by regular checks from a certain age. Some types of cancer directly asssociate with genetic disorders. Nowadays, scientists have a lot of information about the relationship between cancer and genes. Ultraviolet, radiation, viruses and various chemicals cause damage to the genes in the human body. If some specific genes (oncogenes and tumor suppressor genes) are affected, then cancer develops (Arthur, et al., 2019).

The cancer causing mutation are inherited from the parent and can be passed on to subsequent generation is known as hereditary cancer. Approximately 10% of all cancers are hereditary. It is vital to know the genetic mutations that cause hereditary cancers prevent cancer in the family (Arthur, et al., 2019).

There are three main gene types that can be associated with cancer. The first one is tumor suppressor genes. These are protective genes. These genes normally limit the rate of cell division, repair damaged DNA and control cell death. When these genes are mutated, the cells continue to grow and eventually the tumor forms. Approimately, 30 tumor suppressor genes have been identified, including *BRCA1*, *BRCA2*, and *TP53* genes. In about 50% of all cancer cases, the *TP53* gene is either absent or damaged. The second one is oncogenes. This type of genes convert healthy cells into cancerous cells. *RAS* is the two most common oncogene. The last one is DNA repair genes. It provie repairs errors that occur during DNA replication. Unrepaired errors lead to mutation and eventually develop cancer. For example, *BRCA1* and *BRCA2* genes are associate with breast and ovary cancers (Phillips, et al., 2018).

In 2018, Aggarwal and his colleagues conducted a study to test germline mutations in breast cancer patients. In their study, they analyzed many genes including *BRCA1* and *BRCA2* (Aggarwal, et al., 2018).

**Breast Cancer Gene *(BRCA1* and *BRCA2)* Mutations: Cancer Risk and Genetic Testing.** *BRCA1* is a gene that generally limits the growth of cells in the breast, but when it has mutated in mammals, it may occur to breast cancer. The *BRCA1* gene belongs to a class of genes known as tumor suppressor genes. Like

other tumor suppressor genes, *BRCA1* regulates the cycle of cell division by allowing the cells to multiplication and separate very rapidly or remain uncontrolled. In particular, it prevents the growth of cells that align the milk channels in the breast (Phillips, et al., 2018). The protein produced by the *BRCA1* gene is directly effective in repairing damaged DNA. In the nucleus of many normal cells, the *BRCA1* protein interacts with the protein produced by the *RAD51* gene to correct DNA breaks. These breaks can be caused by environmental agents such as radiation, but chromosomes can also occur naturally in the exchange of genetic material. *BRCA2*, which has a similar function to *BRCA1*, also interacts with *RAD51* and repairs DNA. These three proteins play an important role in maintaining the stability of the human genome (Table 1) (Copson, et al., 2018).

Each human have two copies of *BRCA* genes that inheriting from a parent. Individuals, who carry one mutated copy of *BRCA* genes, have a higher risk of developing breast cancer, ovarian cancer, and prostate cancer in their lifetime. Many genetics alterations in *BRCA* genes can not cause cancer development (Copson, et al., 2018). However, inherited *BRCA* gene mutations are associate with an autosomal dominant familial breast and ovarian cancer. The characteristic properties of the *BRCA1* and *BRCA2* genes are shown in Table 1.

Table 1.

*Characteristic Properties of BRCA1 and BRCA2 Genes* (Copson *et al.*, 2018)

| | ***BRCA1*** | ***BRCA2*** |
|---|---|---|
| **Genetic Transition** | Autosomal Dominant | Autosomal Dominant |
| **Rate of passing to child** | 50% | 50% |
| **Gene Location** | 17q | 13q |
| **Gene Type** | Tumor suppressor gene | Tumor suppressor gene |
| **Breast Cancer Development Risk** | 90% | 85% |
| **Male Breast Cancer Risk** | 1% | 6% |
| **Other cancers types** | Ovarium, Colon and Prostate | Ovarium, Colon, Prostate, Pancreas and Gastric |

More than 15000 variations have been identified in *BRCA1* and *BRCA2* genes. High-resolution molecular techniques such as next-generation sequencing enaled to identify more variants rather than the low-resolution sequencing methods Sanger sequencing (Kostsopoulos, et al., 2019).

**The Role of Genetic Databases and American College of Molecular Genetics and American College of Pathology (AMG-ACP) Guidelines.** Bioinformatics processes biological information using computer support to develop, organize, and analyze the data. Bioinformatics has became an essential tool especially in health fields such as medicine, biology, and bioengineering (Hu, et al., 2019).

American College of Molecular Genetics (ACMG) is an organization that aimed to develop clinical practice guidelines, laboratory services directories and to establish uniform laboratory standards as well as quality assurance and proficiency testing (Richards, et al., 2015).

American College of Pathology is a member-centered physician organization, established in 1946 and composed of approximately 18,000 board-certified pathologists. This guideline is used for quality-centered clinical decision-making in pathology and laboratory. The aim of this guideline is to improve the practice of laboratory medicine. It helps pathologists and laboratory experts conduct more effective tests with consistent, high-quality results and expert comments. It is evidence-based guidelines to help pathologists and other clinicians as mentioned above. Thus, it supports to make more informed decisions about diagnosis and optimal treatment (Maxwell, et al., 2016).

The ACMG has recommended five different variant classifications: pathogenic, likely pathogenic, variant with unknown significance (VUS), likely benign and benign (Richards, et al., 2015). The pathogenic variants contribute to the development of diseases (Tavtigian, et al., 2018). However, a single pathogenic variant may not be sufficient to cause disease. Likely pathogenic variants have a high likelihood (greater than 90% certainty) of causing disease; however, further evidence will be needed to confirm this assertion of pathogenicity (Nykamp, et al., 2017). VUS variants are crucial as the potential effect of the variant in the protein structure is either unknown or rare in the population or has not been registered before

(Richards, et al., 2015). Thus, the identification of VUS variants is important for precise treatment and targeted therapies.

## The Simulation of Human Intelligence Process by Machines: Artificial Intelligence

Artificial intelligence is the ability of a computer or a computer-controlled robot to perform various activities similar to intelligent living things. In other words, artificial intelligence is a set of software and hardware systems that have many capabilities such as digital logic execution, motion, speech, and sound detection. For example, when the X question is administered to artificial intelligence, it selects the most rational one among the X question answers previously defined in the system. Thus, artificial intelligence filters the answers of the X question and presents the most rational one. The most four important application areas of artificial intelligence are as follows; voice recognition, image processing, natural language processing, and reasoning.

Nowadays, voice recognition technology is an advanced level. Artificial intelligence not only recognizes the voice also reaches the level of understanding and responding to the voice. This technology is used as virtual assistants such as Siri, Cortana. Secondly, image processing is used at many points in our lives. Driverless vehicles need to understand their surroundings in order to move more safely. Image processing turns pixels into code for support this technology. As in other artificial intelligence applications, there are many subtitles and uses in the information retrieval systems. In the simplest example, search algorithms like Google have been developed for access easily to data. Lastly, reasoning is the most fundamental application of artificial intelligence. It is very important that artificial intelligence will filter a few events and make the most logical and rational decision.

### *Feed-Forward Neural Network System*

Artificial neural networks (ANNs) are computer systems developed to automatically perform the abilities of the human brain, such as deriving new information, creating and discovering new information through learning. Artificial neural networks have emerged as a result of mathematical modeling of the learning process by taking the human brain as an example. It mimics the structure of biological neural networks in the brain and their ability to learn, remember, and

generalize. The learning process in artificial neural networks is carried out using examples. During learning, input and output information is given and rules are set (Vijayakumar, et al., 2021).

In a feed-forward neural network, cells are arranged in layers, and the outputs of cells in one layer are input to the next layer weights. The input layer transmits the information it receives from the external environment to the cells in the hidden layer without making any changes. The network output is determined by processing the information in the hidden and output layers. With this structure, feedforward networks perform a non-linear static function. It has been shown that a feedforward three-layer neural network can approximate any continuous function with the desired accuracy, provided that there are sufficient numbers of cells in its hidden layer (Vijayakumar, et al., 2021). The network is given both inputs and the output that should be obtained from the inputs (expected output). The network generates a solution space representing the problem space by making generalizations from its examples. This solution space can produce results and solutions for similar examples shown later (Bakr, et al., 2020).

In the feed-forward phase, neurons in the input layer transmit the data values directly to the hidden layer. Each neuron in the hidden layer calculates the total value by weighting its input values. It processes it with an activation function and transmits it to the next layer or directly to the output layer. The weights between the layers are initially chosen from randomly small numbers (Sultana, et al., 2021).

After calculating the weighted value of each neuron in the output layer, this value is again compared with the activation function to try to minimize the current error. The iteration process is continued until the error value reaches a certain level and thus the training phase of the network is completed. The weight values in the connections between the layers are taken from the network that has completed the training and is stored (Bakr, et al., 2020).

In feed-forward networks, the processor elements are usually divided into layers. Signals are transmitted from the input layer to the output layer by one-way connections. In feedforward ANN, cells are arranged in layers and the outputs of cells in one layer are input to the next layer weights. The input layer transmits the information it receives from the external environment to the cells in the hidden layer

without making any changes. The network output is determined by processing the information in the hidden and output layer (Sultana, et al., 2021).

### *Mamdani's Fuzzy Logic Inference System*

In the computer world, the transfer, storage, and processing of data are based on binary systems. Fuzzy logic principles do not work like classical logic principles. It is not like the dual working system seen in computers. There are no completely true or completely false values in fuzzy logic principles. Instead, uncertain options such as false or close to true, partially true, partially false, are also included and graded within the framework of certain rules (HomaieFasih, et al., 2020).

With fuzzy logic applications, alternatives that are in classical logic principles and that express certainty are stretched. Thus, alternatives are multiplied through models that cover more uncertain situations, and variables are graded in the light of certain rules. It is aimed to solve complex and multidimensional problems related to variables that are uncertain, contradictory, and ambiguous and to make them manageable (Tuan, et al., 2020).

Mamdani inference is the most common fuzzy logic inference since it appeals more to human perception. In this inference, the inputs and outputs clusters are fuzzy values. Membership values are calculated according to the rules triggered by the input values (Shoumi & Syulistyo, 2021). Then the calculated values are given to the max or min operator according to the rules and/or their logical connectors. If the facts in the rule are connected with 'and', the calculated membership values are given to the min operator; If it is connected with 'or', it is given to the max operator. These operators, as their name suggests, return the smallest or largest of the multiple values they take (Tuan, et al., 2020).

### *Artificial Intelligence in Medical Diagnosis*

Systems based on artificial intelligence (AI) are increasingly used to assist physicians in their diagnostic decisions. Nowadays, many scientific studies aim to develop AI technologies and discuss their social effects. In 2021, Jussupow and her colleagues experimented to understand how AI advice affects physicians' decision-making. In this experiment, they used a total of 68 novices and 12 experienced physicians to diagnose patient cases with an AI-based system. The findings from this experiment provide a first insight into the metacognitive mechanisms that decision-

makers use to evaluate system recommendations (Jussupow, et al., 2021). Furthermore, artificial intelligence systems are used in various fields of healthcare, such as the invention of new medical treatments and the treatment of various diseases. AI systems provide significant contributions in the diagnosis of tumors or hereditary diseases that are difficult to detect in their early stages (Shaheen, 2021).

Breast cancer screening has been shown to significantly reduce mortality in women. Nowadays, artificial intelligence and deep learning-based methods are showing promising results in breast cancer diagnosis accuracy. Recent studies leverage artificial intelligence-based systems to efficiently interpret digital mammograms and breast tomosynthesis imaging. However, AI potentially provides more efficient and standardized processes, and models with high diagnostic accuracy make diagnoses for breast cancer patients (Tran, et al., 2021).

In 2020, Adachi *et al.* aimed to develop artificial intelligence (AI) system that can detect and diagnose maximum intensity projection lesions using dynamic contrast breast magnetic resonance images. The system was analyzed using test sets of 13 normal, 20 benign, and 52 malignant cases. All four human readers scored this test data with and without the assistance of the AI system. Sensitivity and specificity were 0.926 and 0.828 for the AI system; 0.847 and 0.841 for non-AI human readers. The AI system showed better diagnostic performance compared to human readers. Thus, they found that the system improved diagnostic performance (Adachi, et al., 2020).

### *Next-Generation in Artificial Intelligence*

Nowadays, healthcare is among the areas where artificial intelligence applications are used the most. Artificial intelligence in the field of health; automated data collection, molecular simulations, literature mining, drug discovery, biomarker discovery, toxicity prediction, disease diagnosis, treatment selection, automated surgery, patient monitoring, etc (Piccialli, et al., 2021). Thus, artificial intelligence might be used in clinical application areas. However, it is certain that the areas where artificial intelligence is used will expand in the future. Humans who can prevent nuclear disasters and traffic accidents, save people's lives after natural disasters, and have lost any of their limbs, will make their lives easier with mechanical limbs with artificial intelligence (Kumar, et al., 2021).

Furthermore, receiving data in digitized form, such as images, enables artificial intelligence applications to become widespread more rapidly and in a variety of ways in specialties such as radiology, pathology, and dermatology. Nowadays, it has been shown that radiology and pathology images are evaluated with a higher level of reliability than human evaluation by artificial intelligence applications using machine learning and image processing methods. Some applications predict significantly the probability of cancer development from radiological images within five years. Likewise, algorithms provide early diagnosis of some diseases from ECG data by using clues that are invisible to the human eye (Ghazal, 2021).

Moreover, artificial intelligence technologies have begun to affect human life more in recent years. Generative Pre-Training Transformer 3 (GPT-3) released by OpenAI company took this a step further (Dale, 2021). This model is the most advanced language model ever released. This artificial intelligence can translate text, make dialogue and even write poetry. This artificial intelligence uses all the texts found on the internet to give a reasonable response based on the input text it receives. Their answers tend to be pretty accurate as well since they have a lot of data (Yang, et al., 2021).

### *Artificial Intelligence in Breast Cancer: Application to Decision Making in the Diagnosis*

Nowadays, there are many different artificial intelligence applications related to the pre-diagnosis of breast cancer. Nowadays, breast cancer is a vital problem affecting human health, artificial intelligence is very important to provide early diagnosis of this disease and to start treatment as soon as possible.

In a study conducted in 2021, they classified breast cancer using the artificial neural networks method. They achieved an accuracy rate of 98.8% in this study. Many different artificial intelligence models have been used in the literature for breast cancer classification (Rahman, et al., 2021). Jijitha and Amudha (2020) used the machine learning techniques K-Nearest Neighbor (K-NN) and Logistic Regression (LR). With these techniques, the classification accuracy of 96.5% and 97.02%, respectively, were achieved (Jijitha & Amudha, 2021). In another study,

they achieved 98.24% breast cancer classification accuracy using the artificial neural network-based Adaptive Particle Swarm Optimization algorithm (Quy, et al., 2020).

Furthermore, in another study, the authors used four different machine learning algorithms to detect breast cancer in women. The accuracy rates obtained from the applied Convolutional Neural Network, Recurrent Neural Network, Fuzzy logic, and Genetic algorithms are 96.49%, 63.15%, 88.81%, and 80.399%, respectively (Sandeep & Bethel, 2021).

Moreover, Adalı and Şekeroğlu (2012) developed a neural network system with a back-propagation learning algorithm for the early prediction of microRNAs responsible for cancer pathogenesis. They used a total of 677 microarray data in the study. The accuracy rate was set to 80% and 90%. The best performance was obtained as 0.00099. This study shows that it is possible to predict microRNAs responsible for cancer pathogenesis at earlier stages using this algorithm.

# CHAPTER II
# Materials And Methods

## Materials

### *Study Design and Cohorts*

A retrospective integrated analysis was performed from two independent breast cancer cohorts from Bursa Uludag University, Department of Genetics, and Erciyes University, Department of Medical Genetics, respectively. In total 932 different patients' data were collected and 268 patients associated with *BRCA1* and *BRCA2* were selected as suitable for the study. Additionally, 12 different individuals were selected for the test group apart from the main cohort. Cancer-causing risk factors such as age, gender, consanguinity of parents, family history with breast cancer, affected family member number, tumor size, lymph node, stage of malignancy, tumor position, estrogen receptor hormone, progesterone hormone, *BRCA1* gene, *BRCA2* gene, other genes alterations, diagnosis and variant classification. In hereditary breast cancer panel *BLM*, *BARD1*, *RAD50*, *PALB2*, *MSH2*, *ATM*, *MLH1*, *MRE11A*, *PMS2*, *MUTHY*, *XRCC2*, *ATN*, *CDH1*, *BARD*, *FAM175A*, *EPCAM*, *PKD1*, *STK11*, *NBN*, *MSH2*, *CHEK2*, *MSH6*, *CDH2*, *BRIP1*, *PTEN*, *PIK3CA*, *MEN1*, *TP53,* and *RAD51D* were also analyzed and included to the study. These risk factors were used as input clusters in the fuzzy logic and neural network systems. Membership functions are added to each input cluster. Each membership function expresses different membership degrees within the cluster.

### *Variant Analysis*

The raw sequence data (FASTQ) was processed into the variant analysis program (Sophia Genetics, Sophia DDM V5.3.8, Saint Sulpice, Switzerland). Genetic variants within breast cancer-related genes were analyzed by community databases NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP/, accessed on 13 September 2019), 1000 Genomes Project **(**http://www.1000genomes.org, accessed on 13 September 2019), Exome Aggregation Consortium (ExAC) (http://exac.broadinstitude.org/, accessed on 13 September 2019) and NHLBI Exome Sequencing Project (ESP) Exome Variant Server

(http://evs.gs.washington.edu/EVS/, accessed on 13 September 2019), and those with a frequency of more than 0.5% were eliminated. The effect of the determined variants at the level of protein structure was evaluated with the MutationTaster, Polyphen-2, PolyPhen2, and Sorting Intolerant From Tolerant (SIFT) in silico detection programs. Genomic Evolutionary Rate Profiling (GERP) was used when considering evolutionary conservation across species. Variant analysis and interpretation were performed with ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/, accessed on 13 September 2019, Varsome (https://www.varsome.com/, accessed on 13 September 2019) and HGMD Professional 2020.2 (https://portal.biobase-international.com/cgi-bin/portal/login.cgi?redirecturl=/hgmd/pro/start.php?, accessed on 13 September 2019) databases.

***Computer***

The main material of this thesis was computer and patients' data from several hospitals. The computer was used to create the artificial intelligence for risk assessment of breast cancer on MATLAB. The computer operating system that use in this study was Windows 10 Pro. The system processor was Intel (R) Core (TM) i5 CPU and 2.53 GHz. Random Access Memory (RAM) was 3.00GB and the memory of the computer was 444GB. The system type was a 32-bit operating system and X64-based processor. The software that created the database was created using the fuzzy logic system via the computer. The software was written on MATLAB 2018a edition. Firstly, MATLAB was loaded on a PC and it required 13 GB of free memory for load MATLAB.

***MATLAB***

MATLAB (matrix laboratory) is a multi-paradigm numerical computing software and a fourth-generation programming language. MATLAB is a proprietary programming language developed by MathWorks. MATLAB allows the user to perform matrix processing, function and data drawing, algorithm implementation, user interface creation, and interfacing with programs written in other languages such as C, C ++, Java, and Fortran (Potter, et al., 2019). Nowadays, MATLAB is frequently used in various fields of Artificial Intelligence, Signal Processing, and Image Processing.

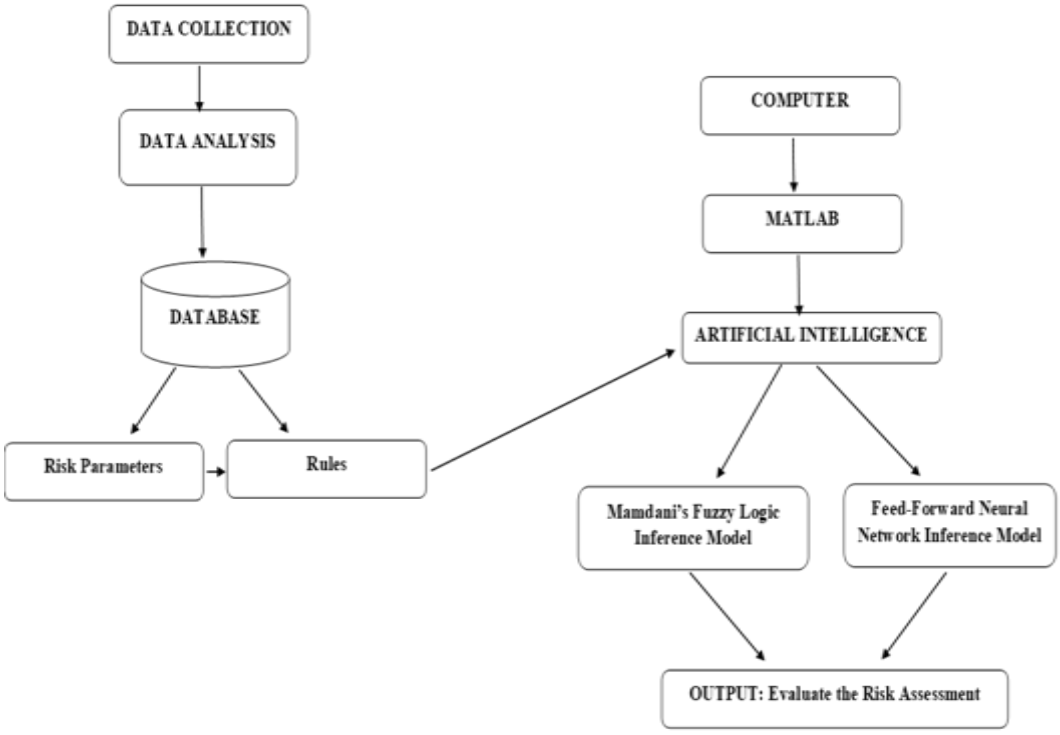
Typical applications of MATLAB are mathematical calculation operations, algorithm development and code writing ie programming, linear algebra, statistics, Fourier analysis, filtering, optimization, numerical integration. Also, drawing 2D and 3D graphics, modeling, and simulation, creating a graphical user interface, data analysis, and control, real-world application development can be summarized as (Potter, et al., 2019). In this study, the MATLAB r2018a version was used.

## Methods

At Figure 1, the workflow of the study was given. Firstly, we collected data from two independent breast cancer cohorts. Then, we analyzed this data and collected the suitable patient data and risk parameters for use in the study into a database. After that, Mamdani's fuzzy logic and feedforward neural network models were developed on MATLAB. We trained our models using patient data selected as a rule and the risk parameters we determined. Finally, these trained artificial intelligence models succeeded in making variant risk assessments as output.

Figure 1

*Demonstrates the Workflow of This Study*



### *Introducing Patients' Data on the System*

Before training our models, we introduced 16 different risk parameters to the system as input clusters. After that, we defined a total of 43 membership functions within these input clusters. Each parameter used enables the models to make a more reliable and consistent risk assessment by evaluating with a different perspective and multiple risk factors. In addition, the patient data introduced to the system as a rule, increased the learning and evaluation accuracy of the systems and contributed to a more reliable and precise output result.

### *Mamdani's Fuzzy Inference Method*

Fuzzy logic is the principle of creating an artificial intelligence application. In other words, it makes computers think like humans. Mamdani's inference is a widely used rule-based method, it is a fuzzy logic method that requires expert knowledge and can be applied to solving all kinds of problems. Mamdani type fuzzy model is very easy to construct and is suitable for human behavior. Therefore, it is widely used and forms the basis of other fuzzy logic models (Pourjavad & Mayorga, 2019).

The five well-known main steps were used: (i) The fuzzification of inputs, (ii) Rule values were determined by using fuzzy logic operations, (iii) The implementation of fuzzy cluster logical processors as "and", "or", (iv) Collection of results; the combination of fuzzy clusters was represented as output of each rule, (v) Defuzzification, where the system clarified the total fuzzy cluster results and converted them into a single number.

The accuracy rate is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we calculated the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as: $\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$

The sensitivity of a test is its ability to determine the patient cases correctly. Mathematically, this can be stated as: $\text{Sensitivity} = \frac{TP}{TP + FN}$

The specificity of a test is its ability to determine the healthy cases correctly. Mathematically, this can be stated as: $\text{Specificity} = \frac{TN}{TN + FP}$

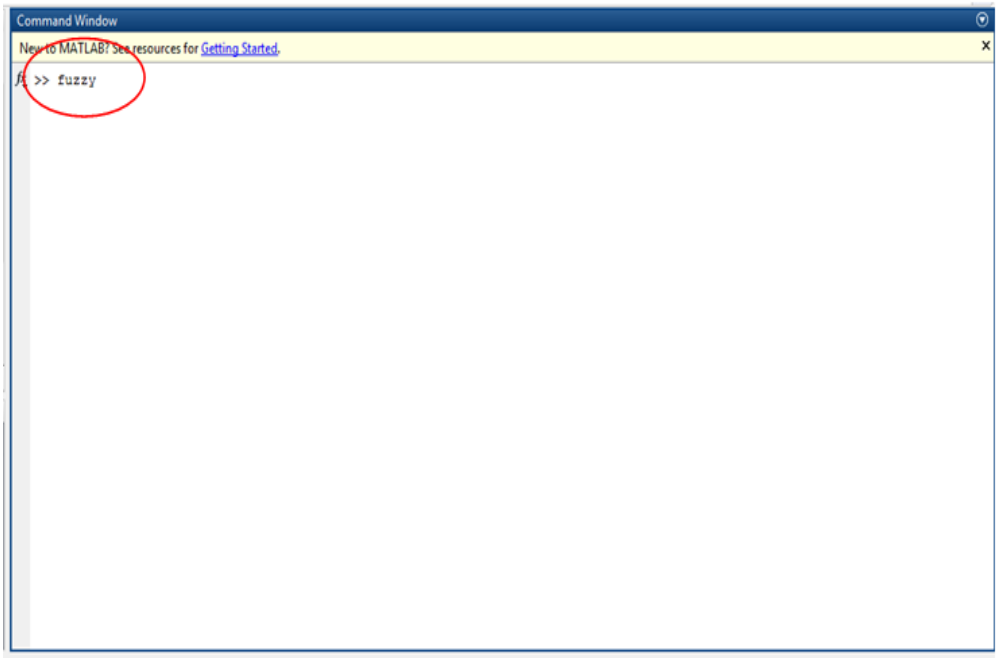
The definitions of the abbreviations in the mathematical formula we used are as follows: True positive (TP) ( the number of cases correctly identified as the patient), False positive (FP) (the number of cases incorrectly identified the as the patient), True negative (TN) (the number of cases correctly identified as healthy), and False negative (FN) (the number of cases incorrectly identified as healthy) (Baratloo, et al., 2015).

### *Mamdani's Fuzzy Logic System Operation in MATLAB*

MATLAB's main interface has five windows: command window, command history, current directory, workspace, and array editor. We used the command window part to switch to the fuzzy logic interface. After typing and confirming the "fuzzy" command in the command window (Figure 2), MATLAB provided access to the fuzzy interface and tool.

Figure 2

*Command Window of MATLAB for Enter Fuzzy Logic Interface.*




After accessing the fuzzy system interface, Fuzzy Logic Designer properties, the first of the five windows used for fuzzy modeling or system creation, are opened in the fuzzy logic toolbox. Other windows are Membership Function Editor, Rule

editor, Rule viewer and surface viewer. In Fuzzy Logic Designer window, we added and removed variables as needed using the Edit menu (Figure 4).

Figure 3

*Interface of Fuzzy Logic System on MATLAB*

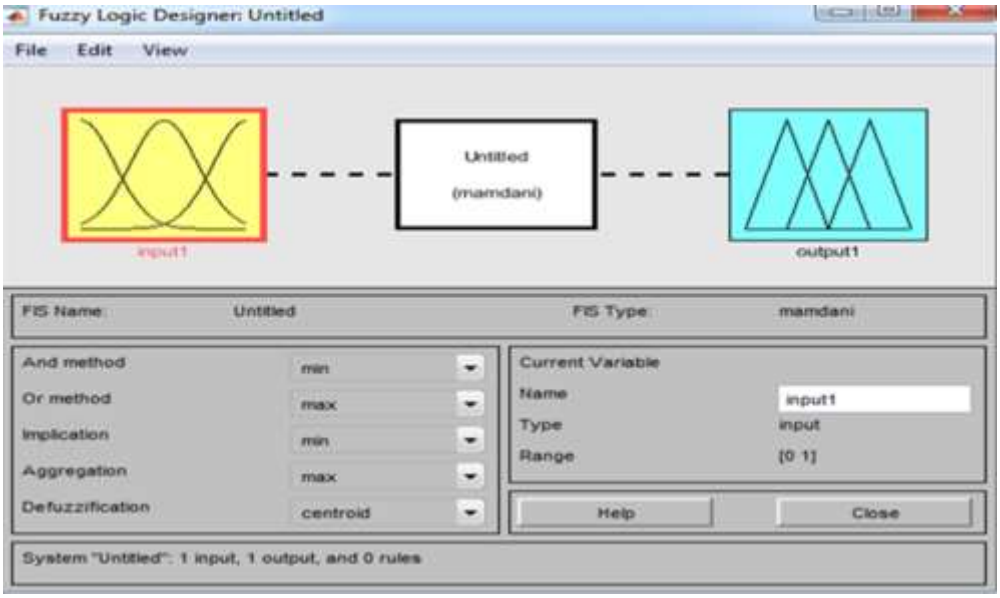
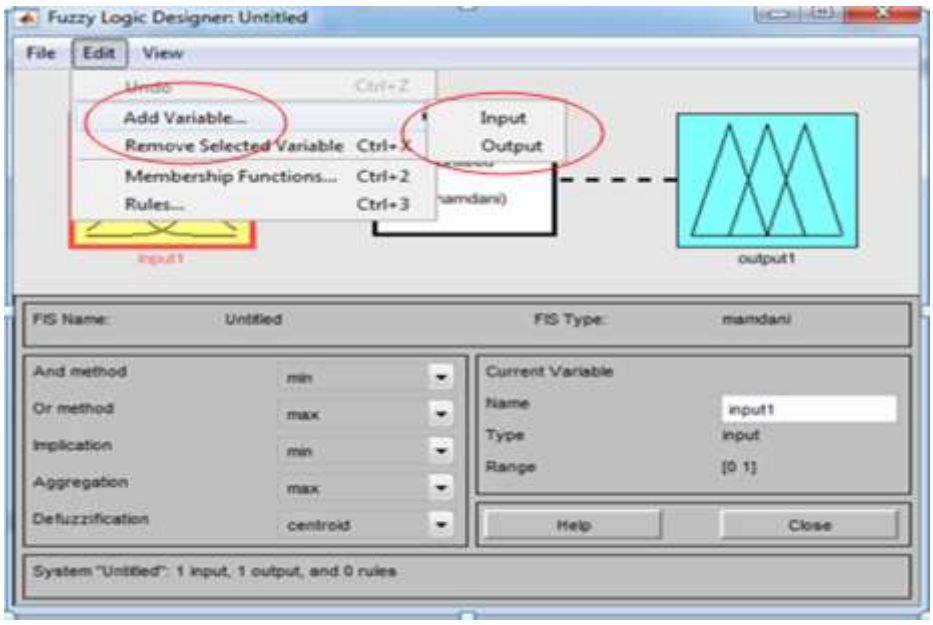



Figure 4

*Adding and Removing Variables in Fuzzy Inference Systems (FIS) Editor*




Next step was Membership Function Editor. In this Editor, type and name of the membership functions of input and output variables, parameters of these

functions and the width of the change of data in the variables were determined. MATLAB supports 11 membership functions under the pop-up menu in the Membership Function Editor. These include dsigmf (d sigmoidal membership function), gauss2mf (Gaussian 2 membership function), gaussmf (Gaussian membership function), gbellmf (Generalized bell membership function), sigmf (sigmoidal membership function), pimf (Pi membership function), psigmf (P sigmoidal membership function), smf (S membership function), trapmf (trapezoid membership function), trimf (triangle membership function), zmf (Z membership function) stop. We used the triangle membership function for input variables and used the trapezoid membership function for output variables (Figure 5). Figure 6 illustrates how membership functions were added or removed from input and output sets.

Figure 5

*Standard Membership Function Editor*

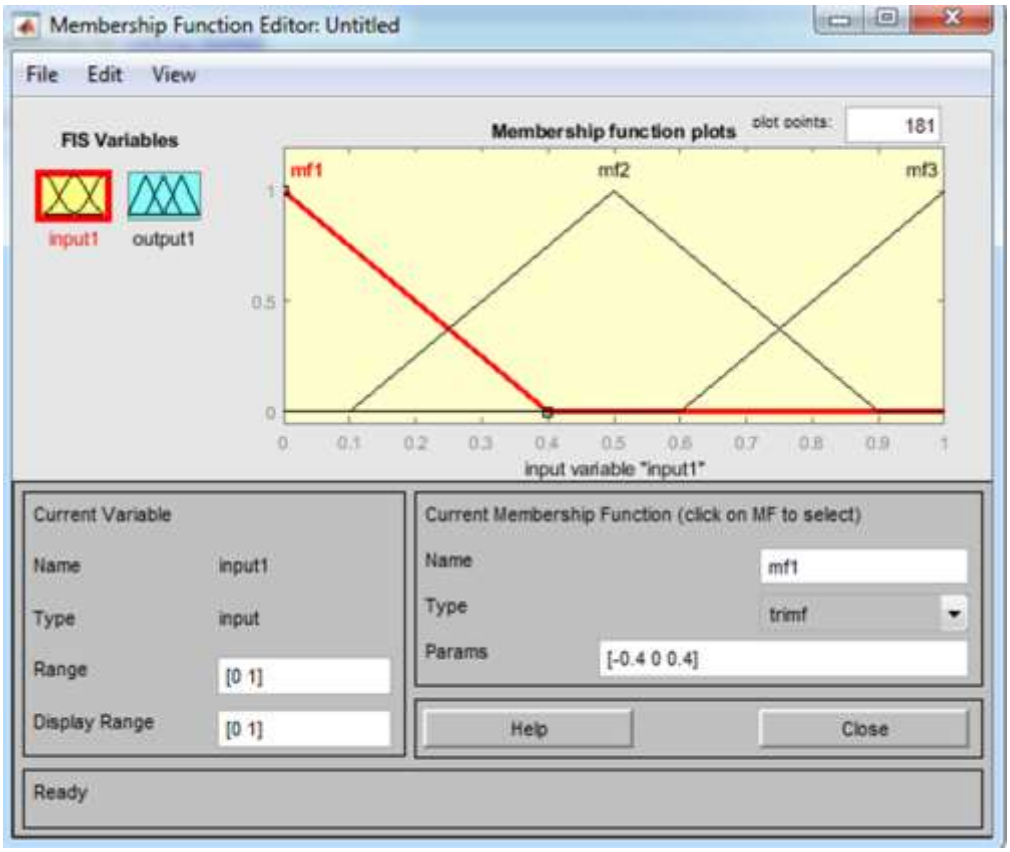
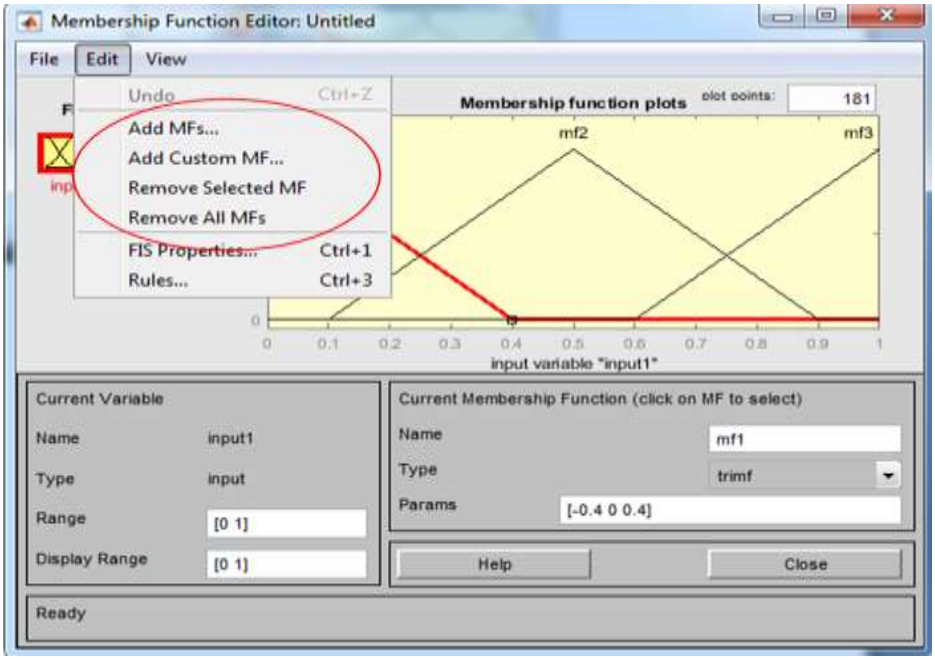


Figure 6

*Add or Delete a New Membership Function*




By the data obtained, the rules of the system were written. The rules are written between the IF-THEN conditions with the AND-OR connectors. All created rules can be observed in the rule window. Then, the researcher can add or remove these rules with the help of the Delete, Add and Change buttons. After creating the related rules, the Rule viewer window was accessed by using View> Rules in the Rule Editor window. When different values were entered in the input variables, the output value estimated by the fuzzy system created can be observed in this window. Researchers can also use the View> Surface path in the Rule Viewer window to obtain a 3D graph of the output value corresponding to the input value they entered.

### *Feed-Forward Neural Network Inference Method*

A neural network is a computational model developed inspired by the information processing technique of the human brain. A neural network can learn from data. Thus, it can be trained to recognize patterns, classify data, and predict future events (Poznyak, et al., 2018).

Different types of neural networks use different principles in determining their own rules. There are many types of artificial neural networks, each with its unique strengths. In Feed-Forward neural network method, data moves in only one

direction from the first stage onwards until it reaches the output node. Unlike in more complex neural networks' types, there is no backpropagation, and data moves only in one direction (Belciug, 2020).

In this study, the neural network-based artificial intelligence model was developed on MATLAB using the feed-forward method. The three well-known main steps were used: (i) The Initialization of network, (ii) Feed-Forward; input values were set and hidden layer values were calculated by using neural network operations, (iii) Backpropagation, where the system clarified the total neural network cluster results and converted them into a single number

***Feed-Forward Neural Network Operation in MATLAB***

In the study, after results were tested in Mamdani's fuzzy logic system, feed-forward neural networks were also developed to verify the results obtained and to confirm with two different systems. Thus, it is aimed to evaluate the data with two different artificial intelligence methods and compare the results obtained. Feed-forward neural networks were also developed on MATLAB. We ran the "nntool" command to access the toolbox, which we will work with neural networks in the MATLAB command window (Figure 7). We have accessed the neural network interface illustrated in Figure 8 with this command.

Figure 7

*Command Window of MATLAB for Neural Network Interface*

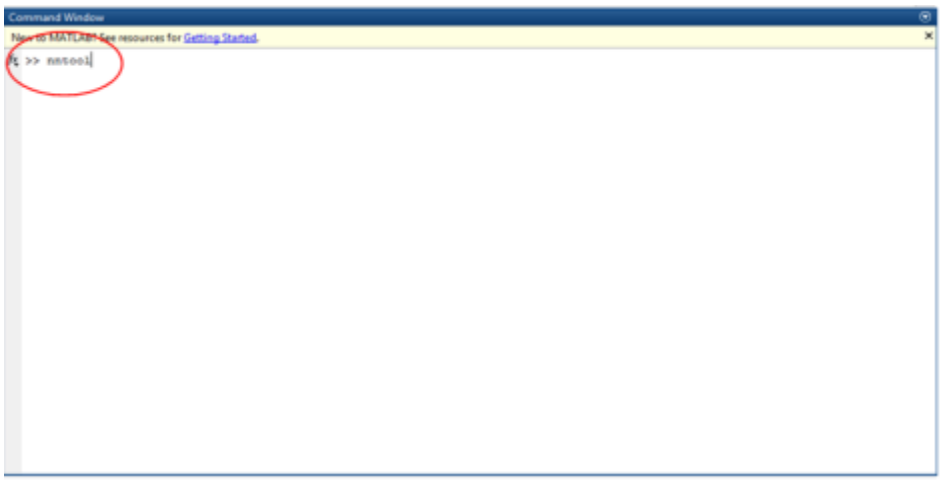
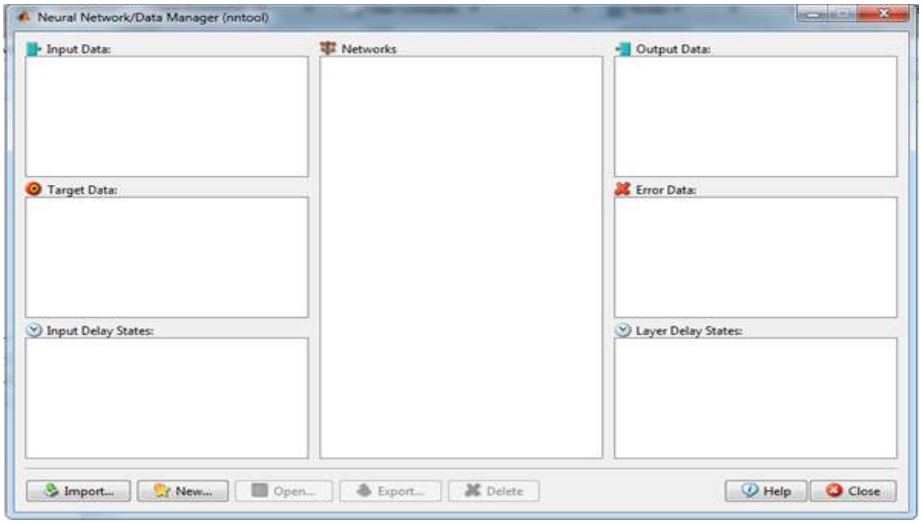


Figure 8

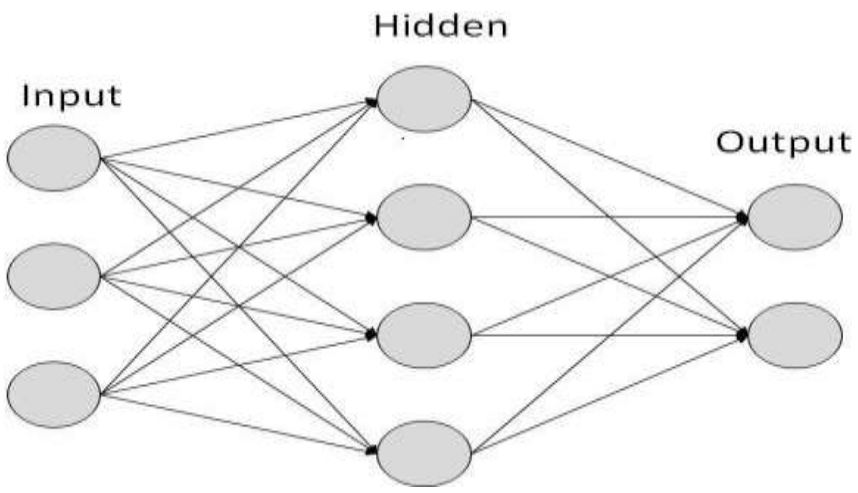
*Interface of Neural Network System on MATLAB*




The next step is to introduce layers for start to create system. The Feed-Forward Neural Network was constructed with three layers as input, hidden, and output. In the input layer, risk parameters and rules were introduced to the system, then in the hidden layer, the system was trained, tested, and validated, and the signals obtained after all calculations were sent to the output layer. The output layer, which is the last stage, converted the obtained signals into single numeric numbers and variant risk assessment results were evaluated.

Figure 9

*Feed-Forward Neural Network Structure* (adapted from https://www.tutorialspoint.com)



After the system had trained, the regression results of the system were checked. There are a total of four different regression results, these are training, validation, testing, and overall. Thus, we calculated the success and error rates of the system, which allowed us to calculate how accurately the system performs to evaluate variant risk assessment.

# CHAPTER III
# Results of Fuzzy Logic System

## Introduction

Cancer begins with the uncontrolled proliferation of cells and can be originated from many different organs such as the lung, breast, intestine, blood, etc. Although cancers are similar in many ways, they differ in growth (Gresner, et al., 2020).

Normal body cells divide, get older, able to repair the damage within the cells and can produce new cells. On the other hand, cancer cells cannot be controlled and they overproliferation damage the normal healthy cells around. Some types grow and spread rapidly, others not. Each cancer type responds differently to a treatment. Even different mutations causing the same type of cancer might be different in treatment (An, et al., 2020).

Breast cancer is cancer cells that occur within the milk channels in the breast tissues. 80% of breast cancers are invasive ductal carcinoma. Invasive ductal carcinoma indicates that breast cancer occurs in the milk ducts. Also, 20 percent of breast cancer is invasive lobular carcinoma that develops in the milk glands, not in the milk channels. Early diagnosis is crucial in every cancer type including breast cancer (Gresner, et al., 2020).

In the literature, many studies have used artificial intelligence models and created risk assessment software. To the best of our knowledge, this is the first study to assess breast cancer risk using BRCA1 and BRCA2 genetic variants using MATLAB for both fuzzy logic and neural network.

## Data Collection and Study Design

A total 932 breast cancer patients information collected from Erciyes University and Uludağ University. 280 of 932 were *BRCA1* and/or *BRCA2* associated breast cancers. 268 patients of out 280 were used to train the systems and 12 patients were not introduced to systems. These 12 patients were used to test the systems. 652

patients were not associated with either *BRCA1* nor *BRCA2* genes. 462 in 652 patients could not be associated with the gene that was analyzed.

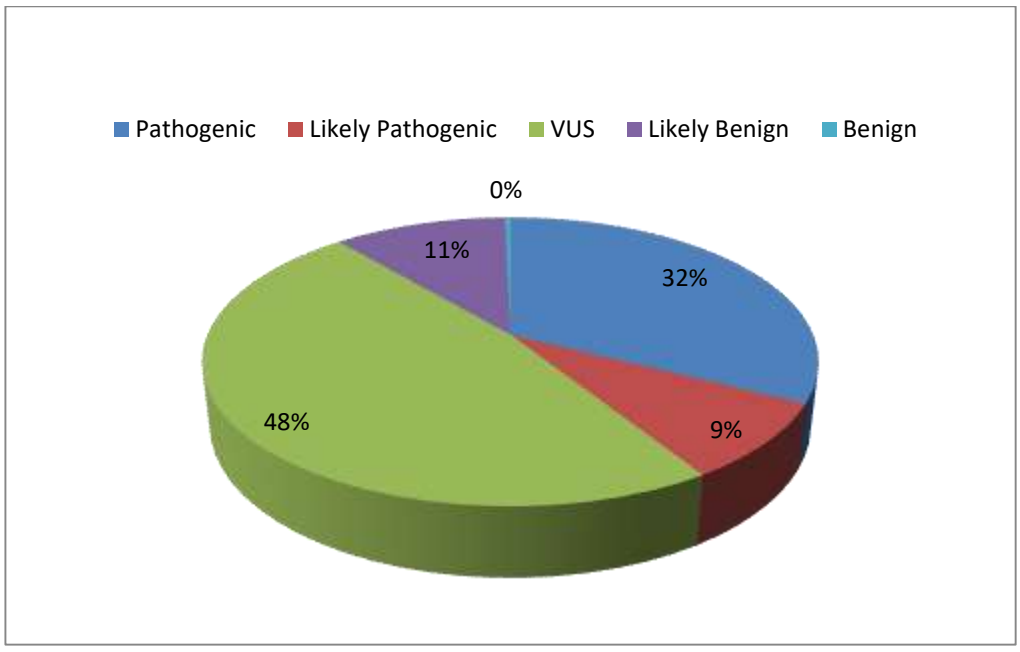
16 risk factors as 16 input parameters included: age, sex, consanguinity of parents, cancer-related family history, number of an affected family member, tumor size, lymph node, degree of malignancy, tumor position, estrogen receptor hormone, progesterone hormone, *BRCA1* gene, *BRCA2* gene, other genes that analyzed, diagnosis and variant classification.

61 *BRCA-1*, 128 *BRCA-2* and 11 both *BRCA1* and *BRCA2*-associated breast cancer patients clinical, histopathological, and genetic information were used to train the system. Additionally, 68 other genes related breast cancers also were trained to the system. As expected only 8% of them were male breast cancer.

The variants were classified in five different subgroups according to ACMG guidelines such as, pathogenic, likely pathogenic, variant with unknown significance (VUS), likely benign and benign. In our cohort, 87 patients were classified as pathogenic, 23 patients were in likely pathogenic group, 128 patients were indicated VUS, 29 were associated wth likely benign and just one patient was benign. The distribution of this classification is given in Figure 10.

Figure 10

*Distribution of the Variant and Their ACMG Guidelines.*



**Determining Risk Factors of Breast Cancer**

Cancers associated with physical characteristics such as age, sex and race. Some types of cancer are sex-related such as prostate cancer in men. Breast cancer can occur in both men and women, but women have a higher risk of developing breast cancer. In our study, gene alteration was the main focus. As gene variants impact cancer development and it can be inherited. The determination of the variant is important for the family members as a preventive medical approach.

As mentioned before, 16 determined different risk factors were divided into groups in itself. These are called membership functions. These functions allowed us to rate risk factors (Table 2).

Table 2.

*Determined Risk Factors and Membership Functions*

| Risk Factors | Membership Functions |
|---|---|
| Age | <15 |
| | 15-19 |
| | 20-39 |
| | 40-59 |
| | >=60 |
| Sex | Male |
| | Female |
| Consanguinity | No |
| | Yes |
| Family History | No |
| | Yes |
| Number of Family Member | 0 |
| | 1 & 2 |
| | >=3 |
| Tumor Size | 0-19cm |
| | 20-39cm |
| | >=40cm |
| Lymph Node | Negative |
| | Positive |
| Degree of Malignancy | Grade 1 |
| | Grade 2 |
| | Grade 3 |

Table 2 (Continued).

| | |
|---|---|
| Position | Right Breast |
| | Left Breast |
| | Both Breast |
| | Other |
| Estrogen Receptor | Negative |
| | Positive |
| Progesterone | Negative |
| | Positive |
| *BRCA1* Genes | Negative |
| | Positive |
| *BRCA2* Genes | Negative |
| | Positive |
| Other Genes | Negative |
| | Positive |
| Diagnosis | No |
| | Yes |
| Classification | Pathogenic |
| | Likely Pathogenic |
| | VUS |
| | Likely Benign |
| | Benign |

The input clusters of the system were developed using the risk factors and membership functions described in Table 2 above.

**Merging the Big Data**

268 patients' information which are illustrated and 16 major risk factors and their membership functions were merged to train the system. A total of 43 different membership functions were identified to system for 16 risk factors. This huge stack of merged data was used to train the fuzzy logic system for variant risk assessment.

The main and last list was created from 268 carefully selected and previously identified breast cancer patients. These data were registered to our developed fuzzy logic system. For each patient, risk factors and membership functions were created as input parameters.

Table 3-5 showed to illustrate the patient' data used to train the models. Five out of 268 breast cancer patients used in the database were randomly selected and showed as examples.

Table 3.

*First Example of Patient Data for Train the System*

| **Input Parameters** | **Sample Patient Informations** |
|---|---|
| Age | 35 |
| Sex | Female |
| Consanguinity | No |
| Family History | Yes |
| Number of Family Member | 1 |
| Tumor Size | 21cm |
| Lymph Node | Unknown |
| Degree of Malignancy | Unknown |
| Position | Right Breast |
| Estrogen Receptor | Negative |
| Progesterone | Negative |
| *BRCA1* Gene | Positive |
| *BRCA2* Gene | Negative |
| Other Genes | Positive |
| Diagnosis | Positive |
| Classification | Pathogenic |

Table 4.

*Second Example of Patient Data for Train the System*

| **Input Parameters** | **Sample Patient Informations** |
|---|---|
| Age | 43 |
| Sex | Female |
| Consanguinity | Unknown |
| Family History | Yes |

Table 4 (Continued).

| | |
|---|---|
| Number of Family Member | Unknown |
| Tumor Size | 26cm |
| Lymph Node | Negative |
| Degree of Malignancy | Grade 2 |
| Position | Both Breasts |
| Estrogen Receptor | Positive |
| Progesterone | Positive |
| *BRCA1* Gene | Negative |
| *BRCA2* Gene | Positive |
| Other Genes | Negative |
| Diagnosis | Positive |
| Classification | Pathogenic |

Table 5.

*Third Example of Patient Data for Train the System*

| **Input Parameters** | **Sample Patient Informations** |
|---|---|
| Age | 32 |
| Sex | Female |
| Consanguinity | Yes |
| Family History | No |
| Number of Family Member | 0 |
| Tumor Size | Unknown |
| Lymph Node | Positive |
| Degree of Malignancy | Unknown |
| Position | Right Breast |
| Estrogen Receptor | Positive |
| Progesterone | Positive |
| *BRCA1* Gene | Negative |
| *BRCA2* Gene | Negative |

Table 5 (Continued).

| | |
|---|---|
| Other Genes | Positive (*BARD1*, *RAD50*) |
| Diagnosis | Positive |
| Classification | VUS |

**Fuzzy Logic System is Generated on MATLAB r2018a Edition**

In our study, the fuzzy logic system was developed to variant risk assessment of breast cancer. MATLAB r2018a version was used in this study. Many functional processes were performed in MATLAB. Firstly, fuzzy logic interface was opened via MATLAB. This interface was accessed by the fuzzy command from the MATLAB command window (Figure 11).

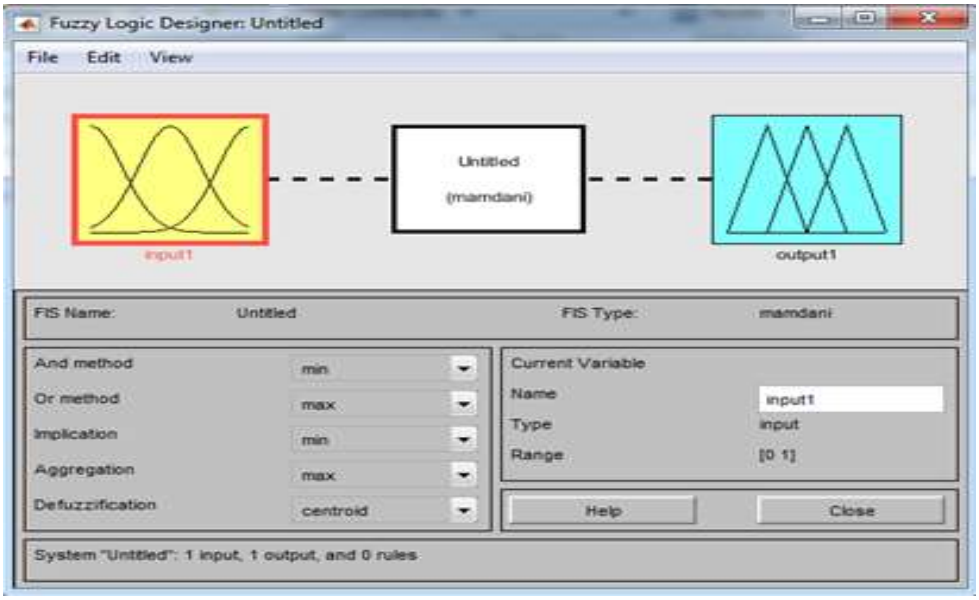
Figure 11

*Command Window of MATLAB*


The standard fuzzy logic interface has three different stages. First, fuzzy sets and membership functions were created and determined in the input stage, then we introduced the rules to the system and combined them with the input sets and membership functions we created in the input stage. Then the system was trained using all this data in the rules stage. In the last stage, the system made the decision-making process, namely the risk assessment, in the output stage.

Figure 12

*Standard Fuzzy Logic Interface in MATLAB*




An unlimited number of input and output parameters can be added to the system. In addition, an unlimited number of data can be used as a rule. After clicking edit section from the main page of the interface, all input and output clusters were added to the system via the add variable section (Figure 13). The 16 different input clusters and an output cluster that we added to the model are shown in Figure 14.

Figure 13

*Adding Input and Output Parameters To the System*

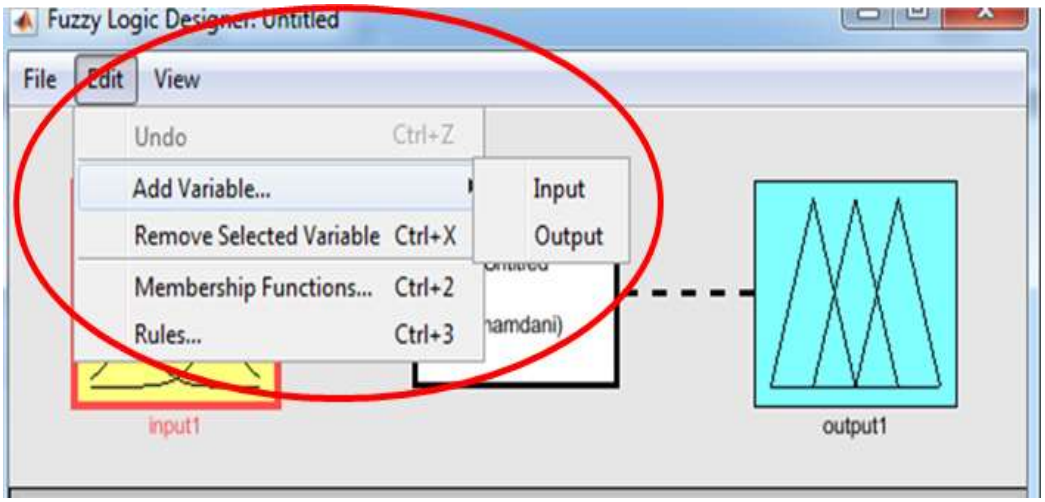
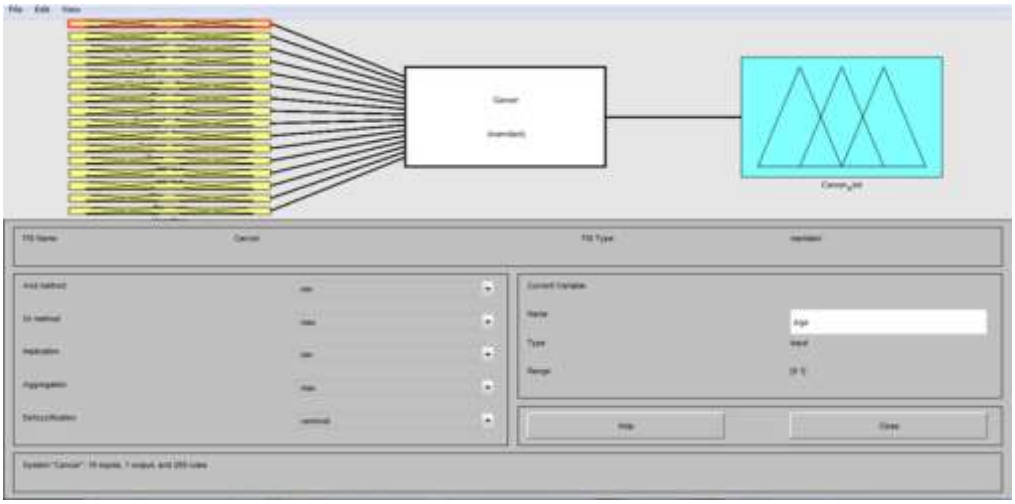


Figure 14

*Input and Output Parameters of the Developed System*



Different membership functions were created within each fuzzy set. These functions determined members with different degrees within each cluster. Each element of the membership function was mapped to a value between 0 and 1. This value, called membership degree. In our study, triangular functions were used in the input section as a membership function and trapezoidal functions were used in the output section as a membership function. Membership functions allowed us to graphically represented a fuzzy set. The x-axis represented the universe of discoursed ($\mu A:X \rightarrow [0,1]$), whereas the y axis represented the degrees of membership in the [0,1] interval.

Figure 15

*Triangular Membership Function*

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{m-a}, & a < x \le m \\ \frac{b-x}{b-m}, & m < x < b \\ 0, & x \ge b \end{cases}$$

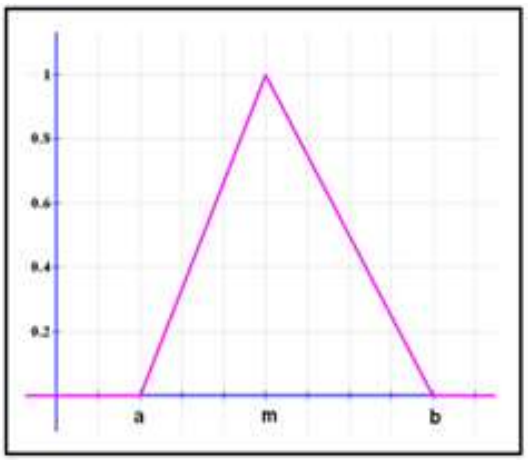
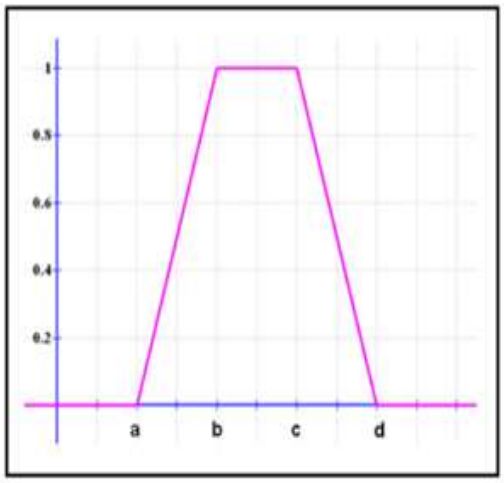


Figure 16

*Trapezoidal Membership Function*

$$\mu_A(x) = \begin{cases} 0, & (x < a) \text{ or } (x > d) \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \end{cases}$$




Triangular function as defined by a lower limit a, an upper limit b, and a value m, where a < m < b (Figure 15). Trapezoidal function as defined by a lower limit a, an upper limit d, a lower support limit b, and an upper support limit c, where a < b < c < d (Figure 16).

In this study, the fuzzy logic-based artificial intelligence model was developed on this platform using Mamdani's fuzzy inference method. The five well-known main steps were used: (i) The fuzzification of inputs, (ii) Rule values were determined by using fuzzy logic operations, (iii) The implementation of fuzzy cluster logical processors as "and", "or", (iv) Collection of results; the combination of fuzzy clusters was represented as output of each rule, (v) Defuzzification, where the system clarified the total fuzzy cluster results and converted them into a single number.

**Gathering Patient Data to the Generated Fuzzy Logic System**

Patient data are very important to training the system. Unsuitable patient data prevents the system from operating correctly. Therefore, it is essential that accurate and appropriate patient data are transferred to the system. As mentioned before, the data of a total of 268 different breast cancer patients associated with *BRCA1* and *BRCA2* were transferred to the system. 61 of these patients were *BRCA1*, 128 of them were *BRCA2*, 11 of them were both *BRCA1* and *BRCA2*, and the remaining 68 patients were associated with other genes.

The membership functions in the input clusters provided different possibilities for different patients as different degrees. These degrees were between 0 and 1 based on the binary system. With the help of these membership functions, we identified which data about the patients should be used as a risk factor in the study for the training of the system.

Sixteen different risk factors were determined for each patient's data. Each risk factor was divided into sub-groups known as membership functions. Membership functions of each risk factor for this model are shown in detail in Table 6.

Table 6.

*Values of Membership Functions in Input Clusters*

| **Input Clusters (Risk Factors)** | **Membership Functions** | **Values [0,1]** |
|---|---|---|
| Age | <15 | 0 |
| | 15-19 | 0.25 |
| | 20-39 | 0.5 |
| | 40-59 | 0.75 |
| | >=60 | 1 |
| Sex | Male | 0 |
| | Female | 1 |
| Consanguinuty | No | 0 |
| | Yes | 1 |
| Family History | No | 0 |
| | Yes | 1 |
| Number of Family Member | 0 | 0 |
| | 1 & 2 | 0.5 |
| | >=3 | 1 |
| Tumor Size | 0-19cm | 0 |
| | 20-39cm | 0.5 |
| | >=40cm | 1 |
| Lymph Node | Negative | 0 |
| | Positive | 1 |
| Degree of Malignancy | Grade 1 | 0 |
| | Grade 2 | 0.5 |
| | Grade 3 | 1 |

Table 6 (Continued).

| | | |
|---|---|---|
| Position | Other | 0.25 |
| | Right Breast | 0.5 |
| | Left Breast | 0.75 |
| | Both Breast | 1 |
| Estrogen Receptor | Negative | 0 |
| | Positive | 1 |
| Progesterone | Negative | 0 |
| | Positive | 1 |
| *BRCA1* Genes | Negative | 0 |
| | Positive | 1 |
| *BRCA2* Genes | Negative | 0 |
| | Positive | 1 |
| Other Genes | Negative | 0 |
| | Positive | 1 |
| Diagnosis | No | 0 |
| | Yes | 1 |
| Classification | Benign | 0 |
| | Likely Benign | 0.25 |
| | VUS | 0.5 |
| | Likely Pathogenic | 0.75 |
| | Pathogenic | 1 |

In fuzzy logic system, 43 different membership functions were created in a total of 16 different input clusters. All these membership functions are shown in Figures 17-32 below.

Figure 17

*Membership Functions of the Age Cluster in the System*

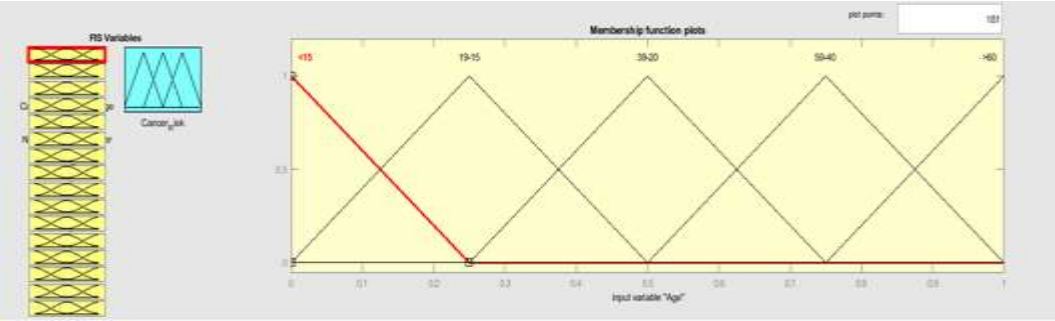


Figure 18

*Membership Functions of the Sex Cluster in the System*

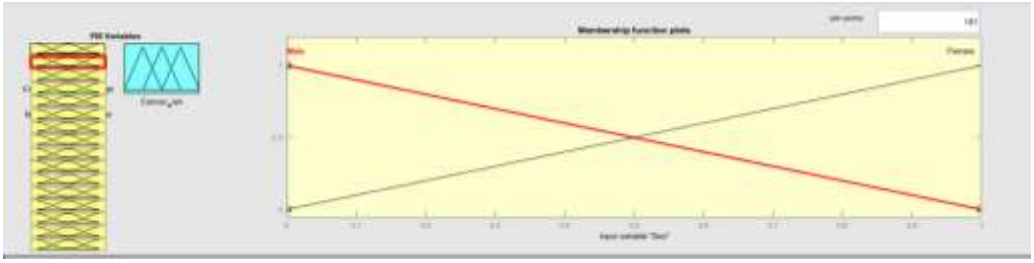


Figure 19

*Membership Functions of the Consanguinuty Cluster in the System*

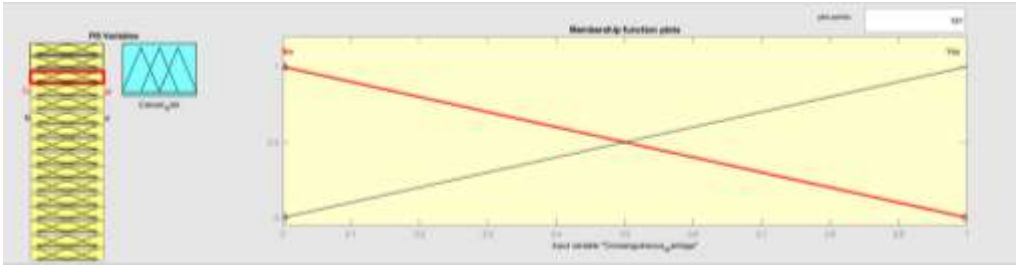


Figure 20

*Membership Functions of the Family History Cluster in the System*

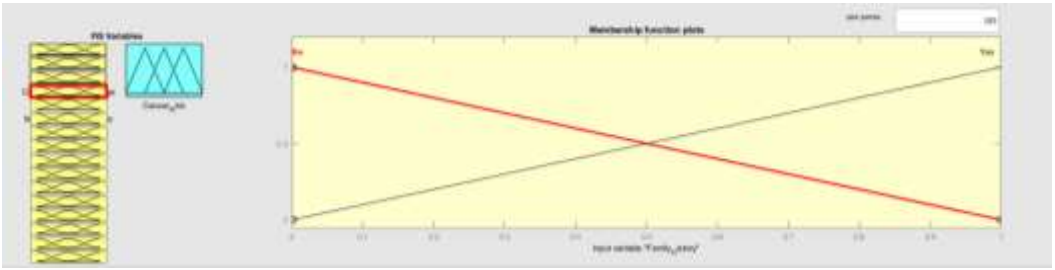


Figure 21

*Membership Functions of the Number of Family History Cluster in the System*

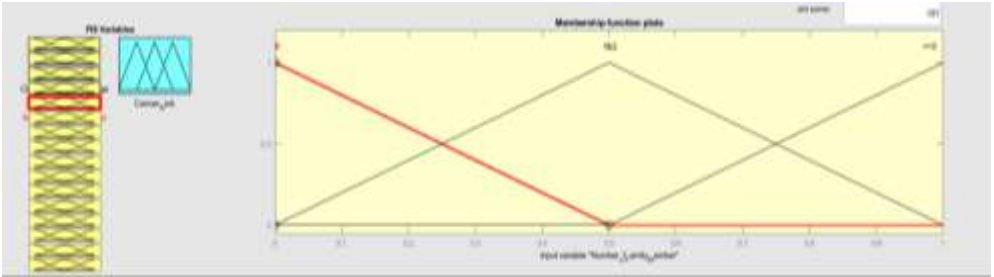


Figure 22

*Membership Functions of the Tumor Size Cluster in the System*

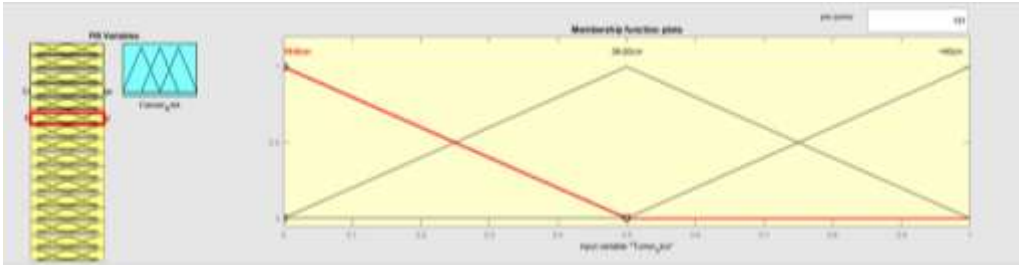


Figure 23

*Membership Functions of the Lymph Node Cluster in the System*

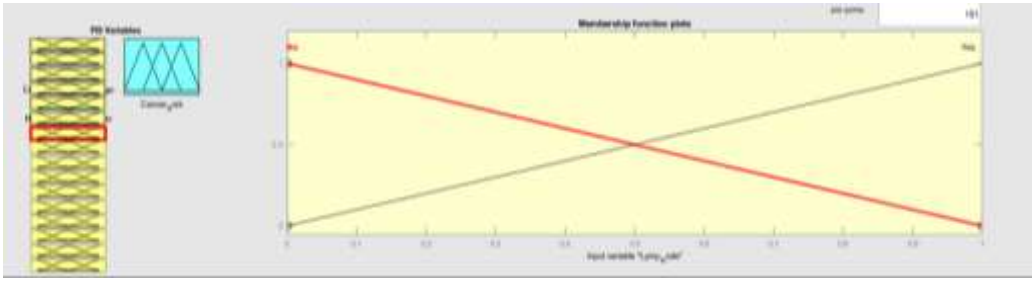


Figure 24

*Membership Functions of the Degree of Malignancy Cluster in the System*

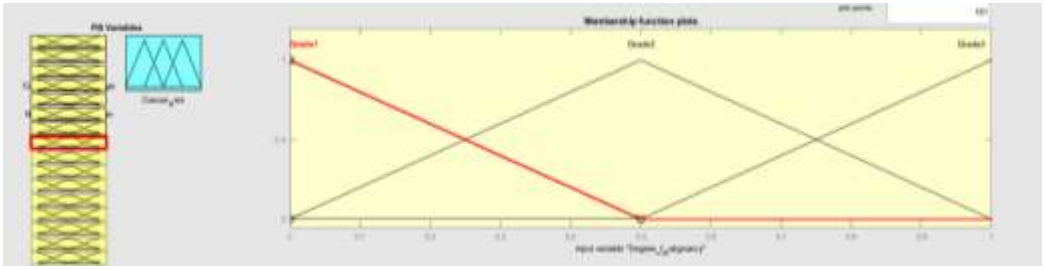


Figure 25

*Membership Functions of the Position Cluster in the System*

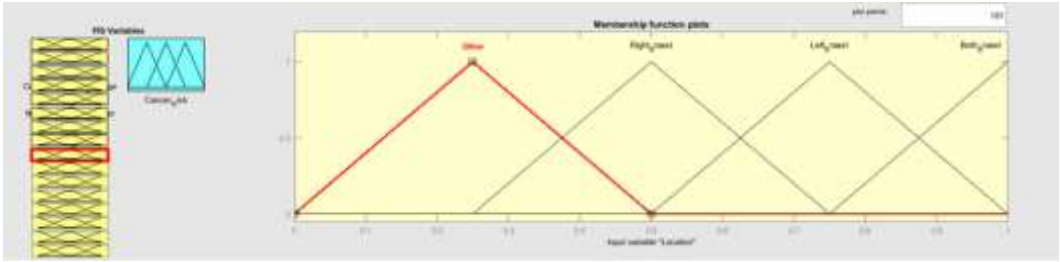


Figure 26

*Membership Functions of the Estrogen Receptor Cluster in the System*

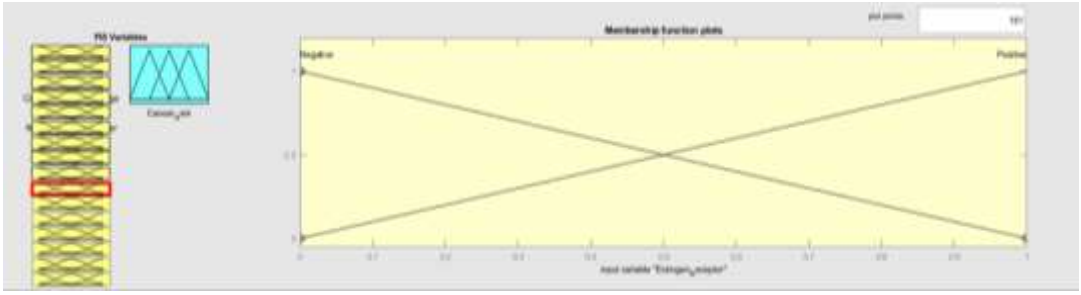


Figure 27

*Membership Functions of the Progesterone Cluster in the System*

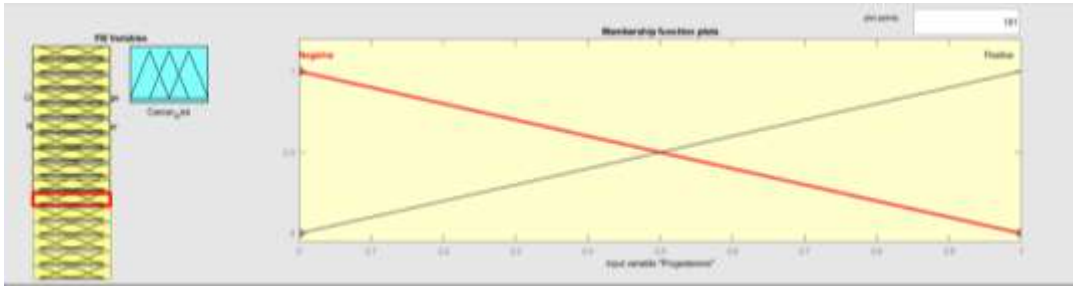


Figure 28

*Membership Functions of the BRCA1 Gene Cluster in the System*

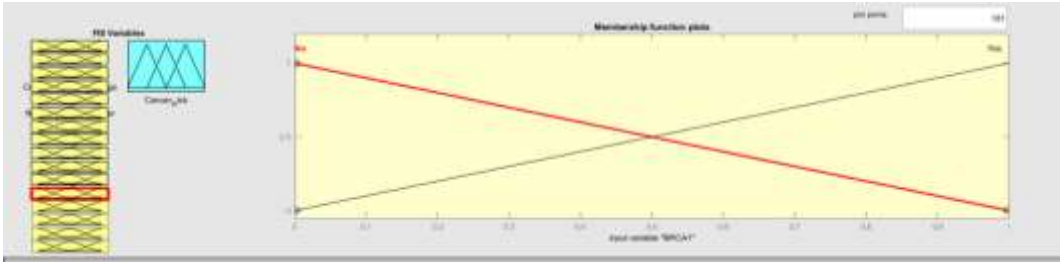


Figure 29

*Membership Functions of the BRCA2 Gene Cluster in the System*

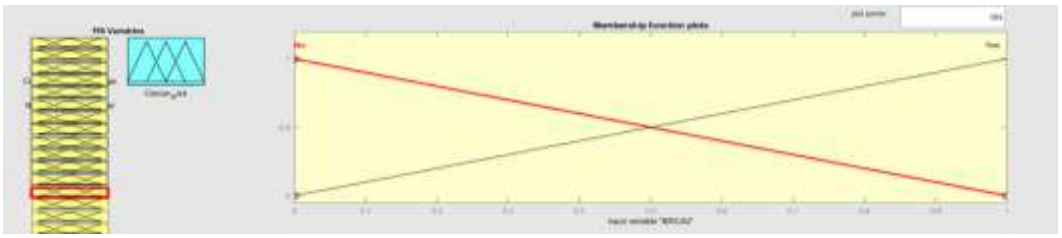


Figure 30

*Membership Functions of the Other Gene Cluster in the System*

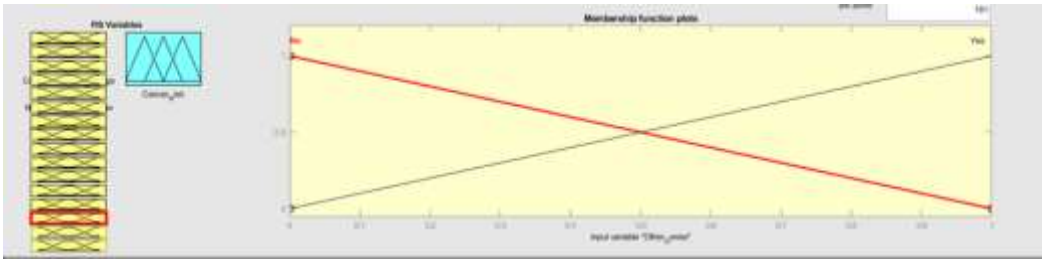


Figure 31

*Membership Functions of the Diagnosis Cluster in the System*

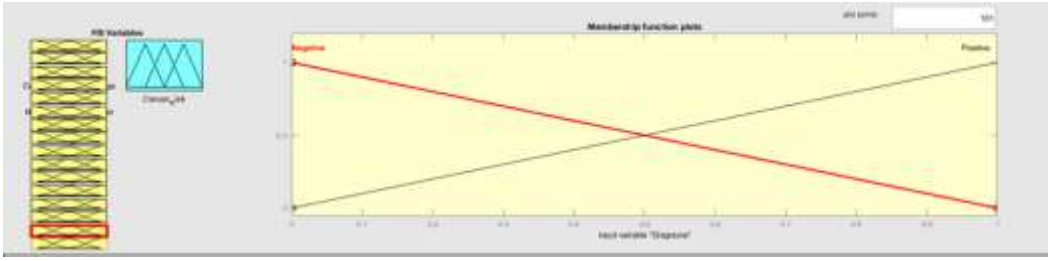
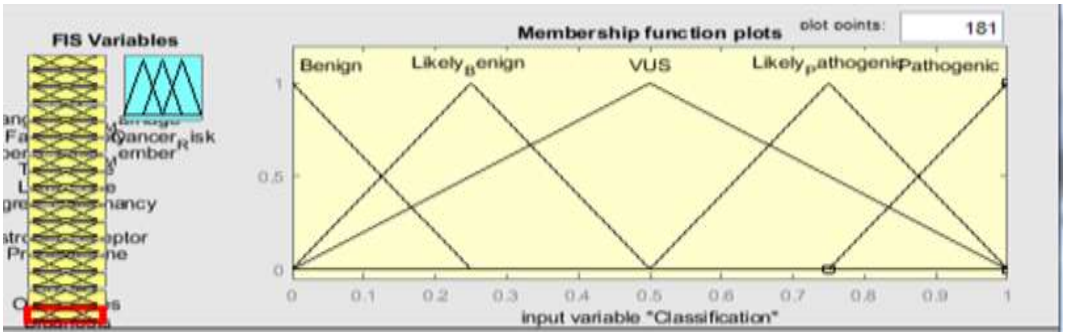


Figure 32

*Membership Functions of the Classification Cluster in the System*




In the rule section, the database that we have created from patients' data is taught to the system as a rule. Each patient was defined as a different rule, so we used a total of 268 different rules. Thus, these rules yield the system a different perspective and possibility. In other words, with these patients' data, the system aimed to evaluate with 268 different possibilities and perspectives to reach a more accurate and sensitive result. The more patient data generated in the system, they obtain greater the sensitivity and accuracy from the result.

Figure 33

*Rules Section in the System*

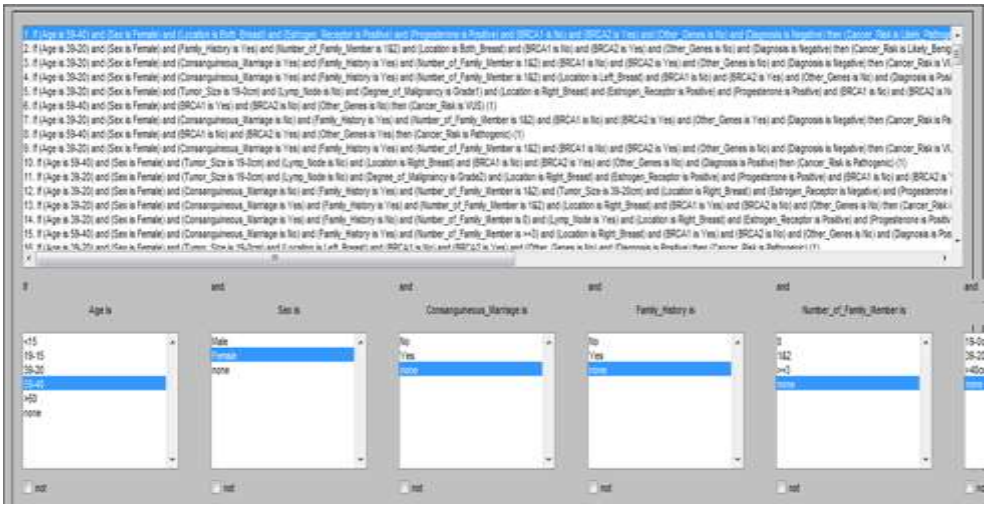



Figure 33 illustrates generated rules section within the Fuzzy Logic system. The upper rectangle box presents an example of the data from 268 patients used that used to train the system. Lower small square boxes show example the parameters (age, sex, consanguineous marriage, family history and number of family members) which were defined as input and membership functions within rule section.

Five membership functions, the values of which were given for each membership, were defined at the output section. A classification range was created to determine the variant pathogenicity which was predicted using *in silico* and variant analysis programs, previously (Table 7). The values in Table 7 were determined by their pathogenic classification according to ACMG (Richards, et al., 2015).

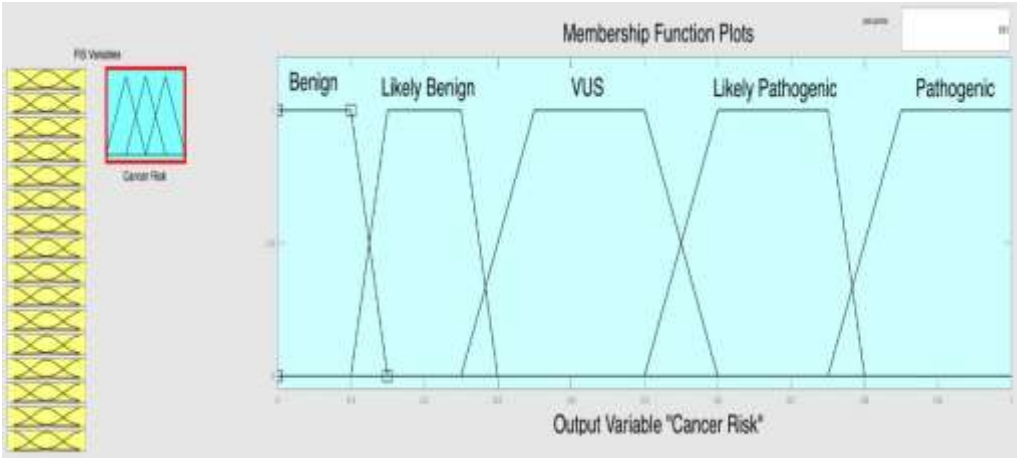
Table 7.

*Output Cluster Created in the System*

| **Membership Functions of Output Cluster** | **Values of Membership Functions** |
|---|---|
| Benign | 0 |
| Likely Benign | 0.25 |
| VUS | 0.5 |
| Likely Pathogenic | 0.75 |
| Pathogenic | 1 |

Figure 34

*The Appearance of the Output Cluster in the Fuzzy Logic Interface on MATLAB.*




The generated appearance of the output cluster using fuzzy logic interface on the MATLAB. Small-merged yellow boxes illustrate sixteen parameters that were introduced as inputs. The blue box shows the output part and determines five different variant classifications as membership functions. The Y-axis presents membership functions of output which can be determine according to the output score. The X-axis presents values of membership function between 0–1.

**Testing the Fuzzy System**

The designed software system was tested using an operation test. Six different tests were conducted for 12 different individuals in the test group to check the accuracy of the system. Best to notice that, these subjects were not introduced to the system before. Four patients had two different pathogenic variants within either *BRCA1* or *BRCA2*. Two subjects (subject 1 and 2) had pathogenic *BRCA2* c.7698deIC variant (Table 8) and the other two (subject 3 and 4) had *BRCA1* C.788dupG (p.Ser264*fs*1) pathogenic frameshift variant (Table 8). After entering the data for subject 1 and subject 2, the system calculated values as 0.9 and 0.89, respectively. According to the classification criteria and obtained values, the system was confirmed that both individuals were pathogenic. Subject 3 and subject 4 have also the same pathogenic variant which gave the same risk scores as 0.9. Subject 5 and 6 have the same variant as classified as likely pathogenic *BRCA1* c.4070_4071delAA (p.Glu135.7Glyfs*10). While the test was focused on two likely pathogenic variants, we got 0.661 obtained value for both patients.

Table 8.

*Obtained Results from Testing the System*

| Risk Factors | Test Subject 1 Variant: BRCA2 c.7698deIC Classification: Pathogenic | Test Subject 2 | Test Subject 3 Variant: *BRCA1* C.788dupG Classification: Pathogenic | Test Subject 4 | Test Subject 5 Variant: *BRCA1* c.4070_4071deIAA Classification: Likely Pathogenic | Test Subject 6 |
|---|---|---|---|---|---|---|
| **Age** | 43 | 36 | 44 | 42 | 34 | 33 |
| **Sex** | Female | Female | Female | Female | Female | Female |
| **Consanguineous Marriage** | Unknown | Unknown | Yes | Unknown | Yes | Unknown |
| **Family History** | Unknown | Unknown | Yes | Unknown | Yes | Yes |
| **Number of Affected Family Member** | Unknown | Unknown | 1 | Unknown | 1 | 3 |
| **Tumor Size** | 17.5 cm | 0-1 cm | Unknown | 6.6 cm | Unknown | Unknown |
| **Lymp Node** | No | No | Unknown | No | Unknown | Unknown |
| **Degree of Malignancy** | Unknown | Grade 2 | Unknown | Grade 3 | Unknown | Unknown |
| **Tumor Location** | Right Breast | Right Breast | Right Breast | Right Breast | Right Breast | Unknown |
| **Estrogen Receptor Hormone** | Unknown | Unknown | Unknown | Positive | Unknown | Unknown |
| **Progesterone Hormone** | Positive | Positive | Unknown | Negative | Unknown | Unknown |
| ***BRCA1*** | No | No | Yes | Yes | Yes | Yes |
| ***BRCA2*** | Yes | Yes | No | No | No | No |
| **Other Genes** | No | No | No | No | No | No |
| **Diagnosis** | Yes | Yes | Yes | Yes | Unknown | No |
| **Result** | 90% (0.900) | 89% (0.890) | 90% (0.900) | 90% (0.900) | 66.1% (0.661) | 66.1% (0.661) |

On the other hand, the system was tested for variants of unknown significance (VUS) such as *BRCA2* c.9924 A>G (p.Ile3312Val), *BRCA1* c.3368 A>G (p.Lys1290Glu) and *RAD50* c.379 G>A, respectively (Table 9). Therefore, we focused on making the correct estimation of individuals with VUS and the possible identification of VUS variants. In the fourth test, we checked *BRCA2* c.9924 A>G (p.Ile3312Val) VUS variant in both individuals (subject 7 and subject 8). The system predicted 0.425 and 0.489 cancer risk scores for both individuals, respectively. Accordingly, the system confirmed that two individuals had VUS as they have been obtained within the VUS threshold. Subject 9 and subject 10 carry both *BRCA1* c.3368 A>G (p.Lys1290Glu) VUS variants according to ACMG criteria. The

obtained value was 0.489 for subject 9. However, subject 10 gave 0.571 cancer risk score which was within the likely pathogenic threshold in our fuzzy logic system. Subject 11 and subject 12 carry same *RAD50* c.379 G>A VUS variant and system predicted 0.425 value for both individuals.

Table 9.

*Obtained VUS Results from Testing the System*

| **Risk Factors** | **Test Subject 7** | **Test Subject 8** | **Test Subject 9** | **Test Subject 10** | **Test Subject 11** | **Test Subject 12** |
|---|---|---|---|---|---|---|
| | ***BRCA2* c.9934 A>G** | | ***BRCA1* c.3368 A>G** | | **Variant: *RAD50* c.379 G>A** | |
| | **Classification: VUS** | | **Classification: VUS** | | **Classification: VUS** | |
| **Age** | 38 | 42 | 58 | 58 | 32 | 40 |
| **Sex** | Female | Female | Female | Female | Female | Female |
| **Consanguineous Marriage** | Unknown | No | Unknown | Unknown | Yes | No |
| **Family History** | No | No | Yes | No | No | No |
| **Number of Affected Family Member** | 0 | 0 | Unknown | 0 | 0 | 0 |
| **Tumor Size** | 3-4 cm | 0.5 cm | Unknown | Unknown | Unknown | 30 cm |
| **Lymp Node** | No | No | Unknown | Yes | Yes | Yes |
| **Degree of Malignancy** | Grade 3 | Grade 2 | Grade 2 | Grade 2 | Unknown | Grade 2 |
| **Tumor Location** | Right Breast | Right Breast | Right Breast | Right Breast | Right Breast | Both |
| **Estrogen Receptor Hormone** | Positive | Positive | Positive | Positive | Positive | Positive |
| **Progesterone Hormone** | Positive | Positive | Positive | Positive | Positive | Positive |
| ***BRCA1*** | Yes | No | Yes | Yes | No | Yes |
| ***BRCA2*** | Yes | Yes | No | No | No | No |
| **Other Genes** | No | No | No | No | Yes | Yes |
| **Diagnosis** | Yes | Yes | Yes | Yes | Yes | Yes |
| **Result** | 42.5% (0.425) | 48.9% (0.489) | 48.9% (0.489) | 57.1% (0.571) | 42.5% (0.425) | 42.5% (0.425) |

## Discussion

In similar fuzzy logic-based studies in the literature, five different risk factors were used in a study published in 2018 (Tawfeeq, 2018), and six different risk factors were used in another publication published in 2019 (Domingo, et al., 2019). However, we used 16 different risk factors in our study, as we mentioned before. These risk factors included age, gender, consanguinity, family history, number of

family members, tumor size, lymph node, grade of malignancy, position, estrogen receptor, progesterone, *BRCA1* gene, *BRCA2* gene, other genes, diagnosis, and classification.

Thus, we increased the reliability, accuracy, and consistency of our results, because we enabled our model to evaluate with a broad perspective. In addition, our study is unique in terms of the number of risk factors used and also the use of genetic risk factors.

Militello et al. (2022) investigated classical approaches in breast cancer using DCE-MRI images with an unsupervised method approach based fuzzy logic model. They achieved a sensitivity of 77.84% and a specificity of 87.10% with this model (Militello, et al., 2022). In another study published in 2019, they developed a fuzzy logic model using mammography images to detect breast cancer. They achieved accuracy, specificity, and sensitivity rates were 89.74, 88%, and 89, respectively (Padmavathy, et al., 2019). In our Mamdani's fuzzy logic model, the sensitivity rate was obtained as 98.69%. The specificity rate of the system was calculated as 95.71%. The accuracy rate was achieved as 95.5%. Thus, this accuracy result shows that the error rate of the system was quite low, as 0.045.

Finally, we confirmed our variant risk assessment results with the tests we conducted with the test group of 12 individuals, which had not been introduced to the model before. The results of these tests showed that the developed Mamdani Fuzzy logic model was successful in characterizing the pathogenicity of VUS variants.

# CHAPTER IV

## Results of Neural Network System

### Introduction

Artificial intelligence, the ability of a computer or a computer-controlled robot to perform various activities similar to intelligent creatures. Artificial intelligence is an artificial operating system that is specific to human intelligence, which is expected to display high cognitive functions or autonomous behaviors such as perception, learning, connecting plural concepts, thinking, conducting ideas, solving problems, communicating, making inferences and making decisions. Artificial Neural Networks is a subtitle formed under the concept of Artificial Intelligence (Poznyak, et al., 2018).

Artificial neural networks (ANN) is a computing technology inspired by the information processing technique of the human brain. ANN simulates the way the simple biological nervous system works. Imitated nerve cells contain neurons, and these neurons connect in various ways to form the network. These networks are capable of learning, storing and revealing the relationship between data (Poznyak, et al., 2018).

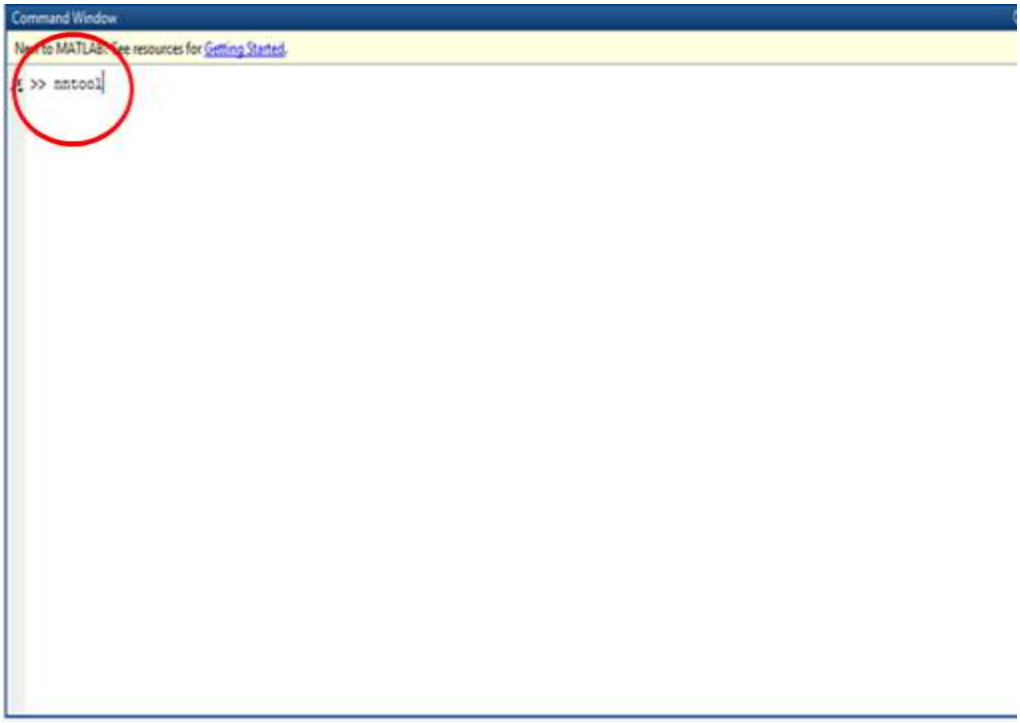
Learning in biological systems is by adjusting synaptic connections between neurons. In other words, humans go through a learning process by living from birth. During this process, the brain is constantly developing. As humans live and experience, synaptic connections are set and even new connections are created. In this way, learning takes place. This also applies to ANN. Learning takes place through the processing of input/output data using examples through training (Poznyak, et al., 2018).

### Neural Network System is Generated on MATLAB r2018a Edition

A neural network system was developed to variant risk assessment of breast cancer. MATLAB r2018a version was used in this study. Firstly, the neural network interface was accessed by the nntool command from the MATLAB command window (Figure 35).

Figure 35

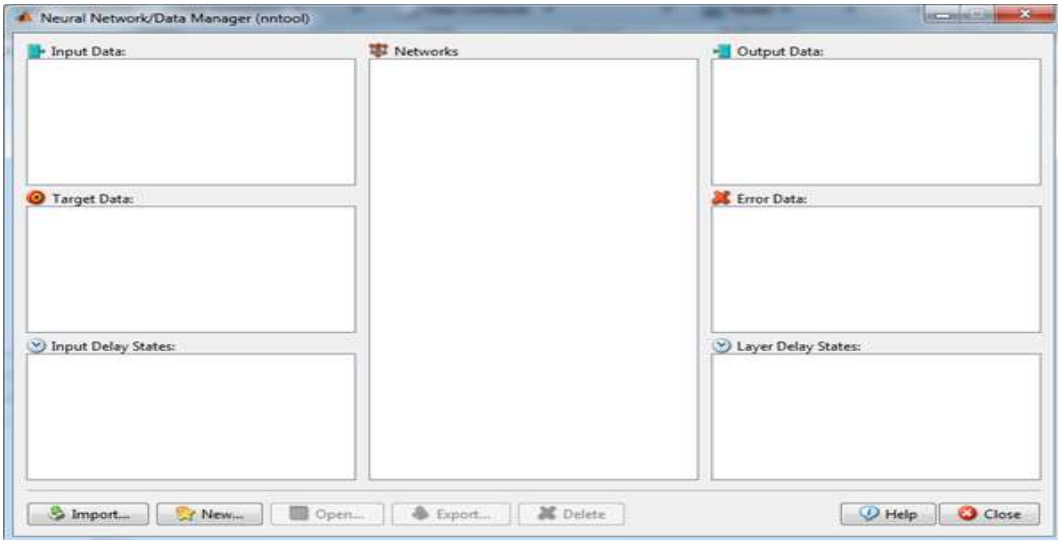
*Command Window of MATLAB.*




There are seven different parts to the standard neural network data manager. These are input data, target data, input delay states, networks, output data, error data and layer delay states (Figure 36). In the input data, patient data to be used to train the system was gathered to the system as input in this section. Then, training output data required for each patient defined as input in the system was transferred to the system in the target data section. Thus, after input and target data were transferred to the system, a new network was created to train the system and after that, this created network was accessed from the network section. Finally, after completing the system training, testing, and validation, the results were obtained from the output section.

Figure 36

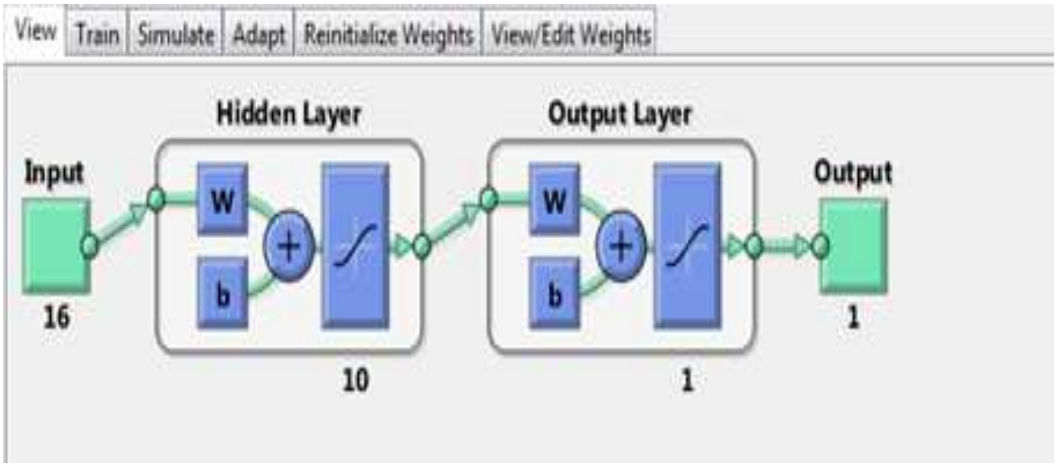
*Standard Neural Network Interface in MATLAB*




An unlimited number of input and target parameters can be added to the system. In addition, an unlimited number of data can be used for the training system. 16 different inputs and output clusters added to the model are shown in Figure 37.

Figure 37

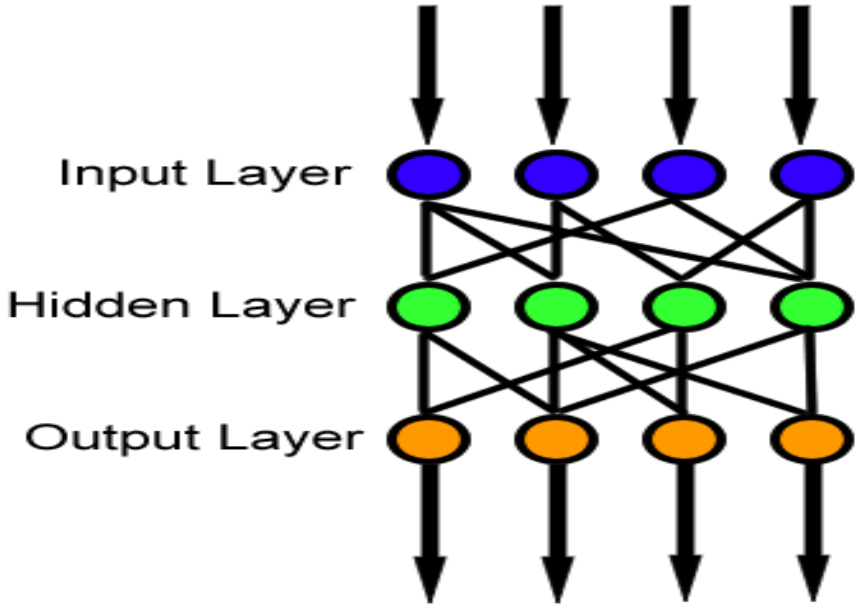
*Input and Output Parameters of the Developed Neural Network System*




In our model, the feedforward method was used to create artificial neural networks. In this method, 16 input layers were connected to the hidden layer and the hidden layer was connected to the output layer to provide a variant risk assessment of breast cancer (Figure 37).

Figure 38

*Simple Feed-Forward Method* (Poznyak *et al.*, 2018).




Feedforward neural networks were mainly used for supervised learning in states where the data to be learned is neither sequential nor time-dependent. In this method, the information moves in only one direction in the network. Data were forwarded from the input nodes to the output nodes through hidden nodes therefore, data were moved only in the forward direction towards the hidden and the output layer (Figure 38).

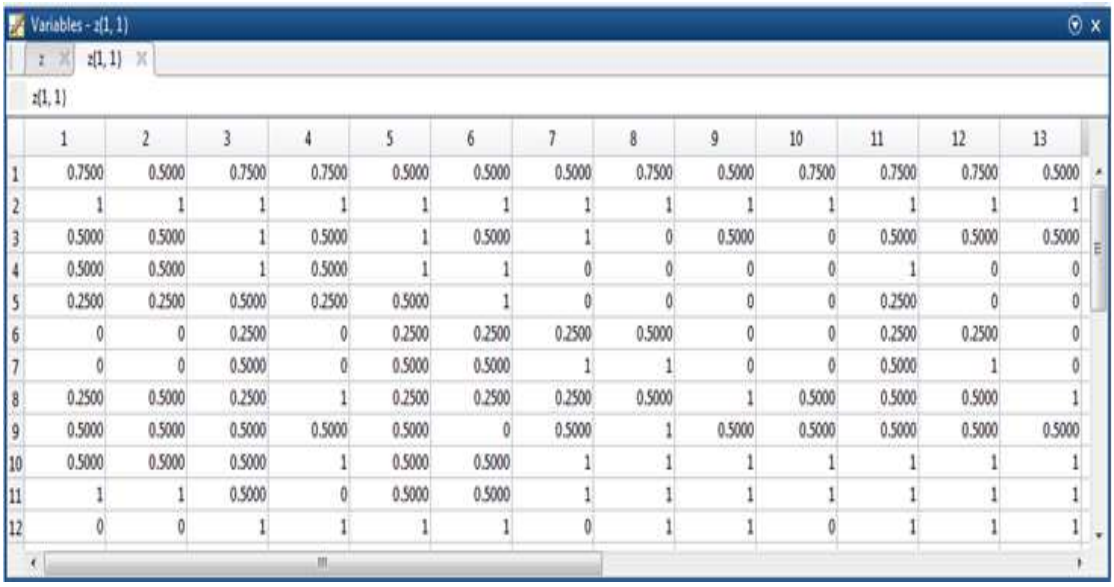
**Gathering Patient Data to the Generated Neural Network System**

Firstly, normalization was performed to enter the input and output data of the patients to be used to train the system. As in our fuzzy logic model, 16 input clusters and 43 membership functions introduced into these clusters were used. In addition, patient data used as a rule were gathered in excel format and merged with input clusters, input (Figure 39) and output (Figure 40) data were gathered into two different excel formats.

Randomly selected 160 patients were used to train the neural network system, 54 patients were tested for the system, and finally, the remaining 54 patients, therefore, were used to validate the created data.

Figure 39

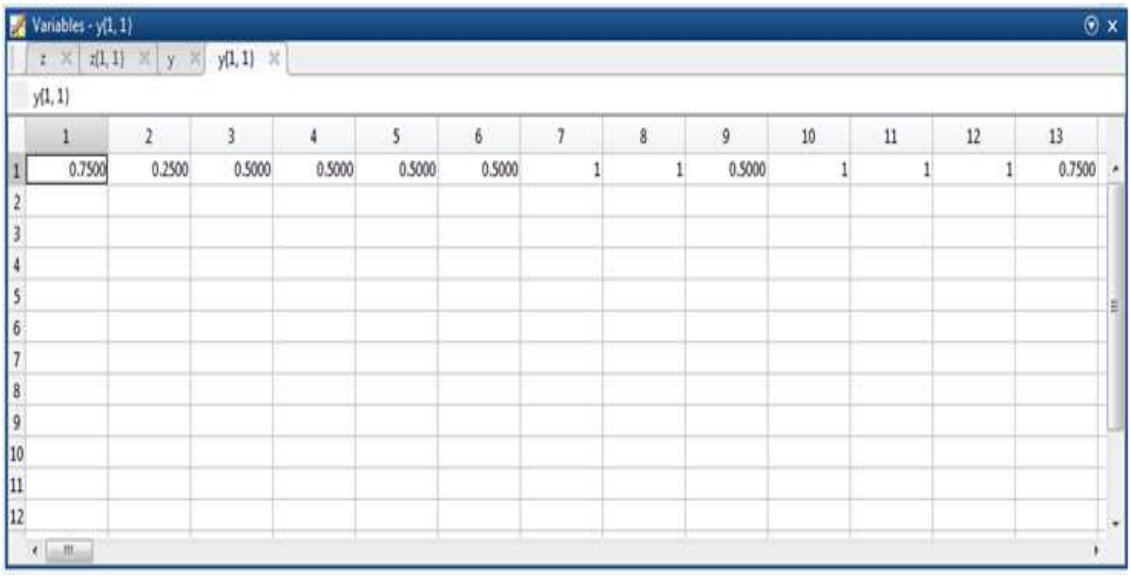
*Input Data for Training to Neural Network System*




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.7500 | 0.5000 | 0.7500 | 0.7500 | 0.5000 | 0.5000 | 0.5000 | 0.7500 | 0.5000 | 0.7500 | 0.7500 | 0.7500 | 0.5000 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3 | 0.5000 | 0.5000 | 1 | 0.5000 | 1 | 0.5000 | 1 | 0 | 0.5000 | 0 | 0.5000 | 0.5000 | 0.5000 |
| 4 | 0.5000 | 0.5000 | 1 | 0.5000 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 0.2500 | 0.2500 | 0.5000 | 0.2500 | 0.5000 | 1 | 0 | 0 | 0 | 0 | 0.2500 | 0 | 0 |
| 6 | 0 | 0 | 0.2500 | 0 | 0.2500 | 0.2500 | 0.2500 | 0.5000 | 0 | 0 | 0.2500 | 0.2500 | 0 |
| 7 | 0 | 0 | 0.5000 | 0 | 0.5000 | 0.5000 | 1 | 1 | 0 | 0 | 0.5000 | 1 | 0 |
| 8 | 0.2500 | 0.5000 | 0.2500 | 1 | 0.2500 | 0.2500 | 0.2500 | 0.5000 | 1 | 0.5000 | 0.5000 | 0.5000 | 1 |
| 9 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0 | 0.5000 | 1 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 |
| 10 | 0.5000 | 0.5000 | 0.5000 | 1 | 0.5000 | 0.5000 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 11 | 1 | 1 | 0.5000 | 0 | 0.5000 | 0.5000 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 12 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |

Figure 40

*Output Data for Training to Neural Network System*




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.7500 | 0.2500 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 1 | 1 | 0.5000 | 1 | 1 | 1 | 0.7500 |
| 2 | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | |
| 7 | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | |
| 11 | | | | | | | | | | | | | |
| 12 | | | | | | | | | | | | | |

Sixteen different risk factors were determined for each patient's data. Each risk factor was divided into sub-groups known as membership functions. Membership functions of each risk factor for the neural network model is shown in Table 10.

Table 10.

*Values of Input Layers in Neural Network System*

| Input Layers (Risk Factors) | Members of Input Layers | Values [0,1] |
|---|---|---|
| Age | <15 | 0 |
| | 15-19 | 0.25 |
| | 20-39 | 0.5 |
| | 40-59 | 0.75 |
| | >=60 | 1 |
| Sex | Male | 0 |
| | Female | 1 |
| Consanguinuty | No | 0 |
| | Yes | 1 |
| Family History | No | 0 |
| | Yes | 1 |
| Number of Family Member | 0 | 0 |
| | 1 & 2 | 0.5 |
| | >=3 | 1 |
| Tumor Size | 0-19cm | 0 |
| | 20-39cm | 0.5 |
| | >=40cm | 1 |
| Lymph Node | Negative | 0 |
| | Positive | 1 |
| Degree of Malignancy | Grade 1 | 0 |
| | Grade 2 | 0.5 |
| | Grade 3 | 1 |
| Position | Other | 0.25 |
| | Right Breast | 0.5 |
| | Left Breast | 0.75 |
| | Both Breast | 1 |
| Estrogen Receptor | Negative | 0 |
| | Positive | 1 |
| Progesterone | Negative | 0 |
| | Positive | 1 |
| *BRCA1* Genes | Negative | 0 |
| | Positive | 1 |
| *BRCA2* Genes | Negative | 0 |
| | Positive | 1 |
| Other Genes | Negative | 0 |
| | Positive | 1 |
| Diagnosis | No | 0 |
| | Yes | 1 |

Table 10 (Continued).

| | | |
|---|---|---|
| Classification | Benign | 0 |
| | Likely Benign | 0.25 |
| | VUS | 0.5 |
| | Likely Pathogenic | 0.75 |
| | Pathogenic | 1 |

After the input and output data were transferred to the system, MATLAB operations were started, first we used the "nntool" command in the command window to access the Neural Network toolbox via MATLAB (Figure 35).

In the first step, the new button was clicked to create a new network and the data of the patients were defined as input and output to the system. Then, the input data to the system was defined as the variable x and the target (output) data as the variable y. In the second step, the system was taught that the x variable is the input data (Figure 42) and the y variable is the target data (Figure 43). Thus, the data were classified as input and target within the system. (Figure 44).

Figure 41

*First Step for Create New Network*

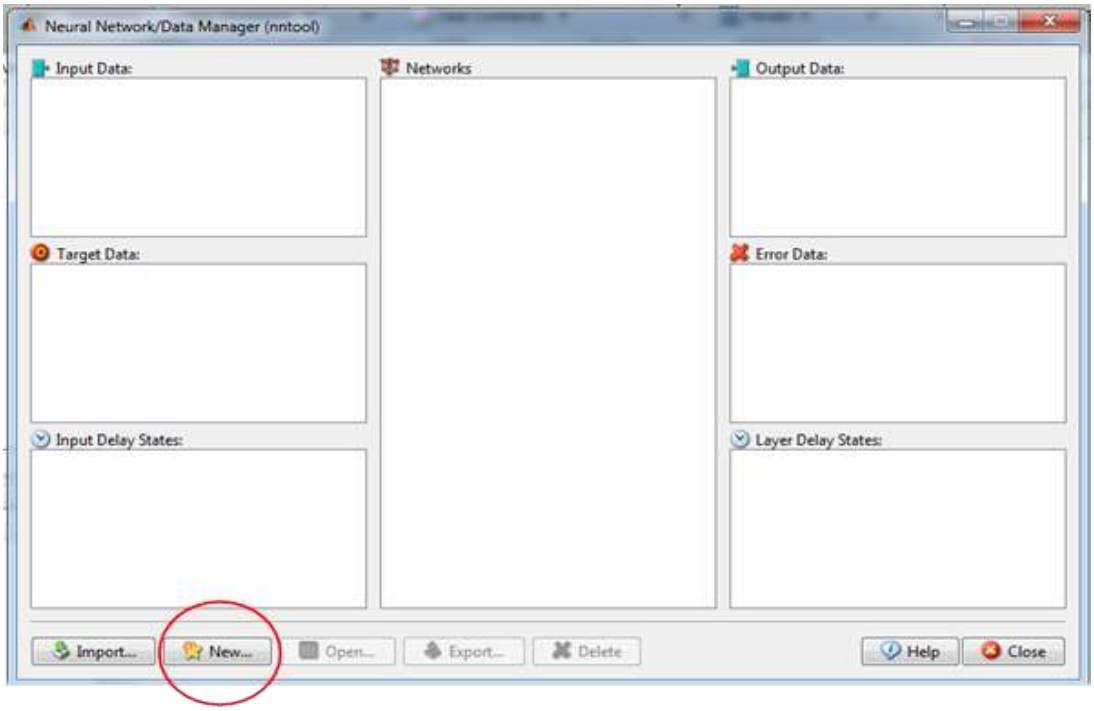


Figure 42

*Input Data on Network*

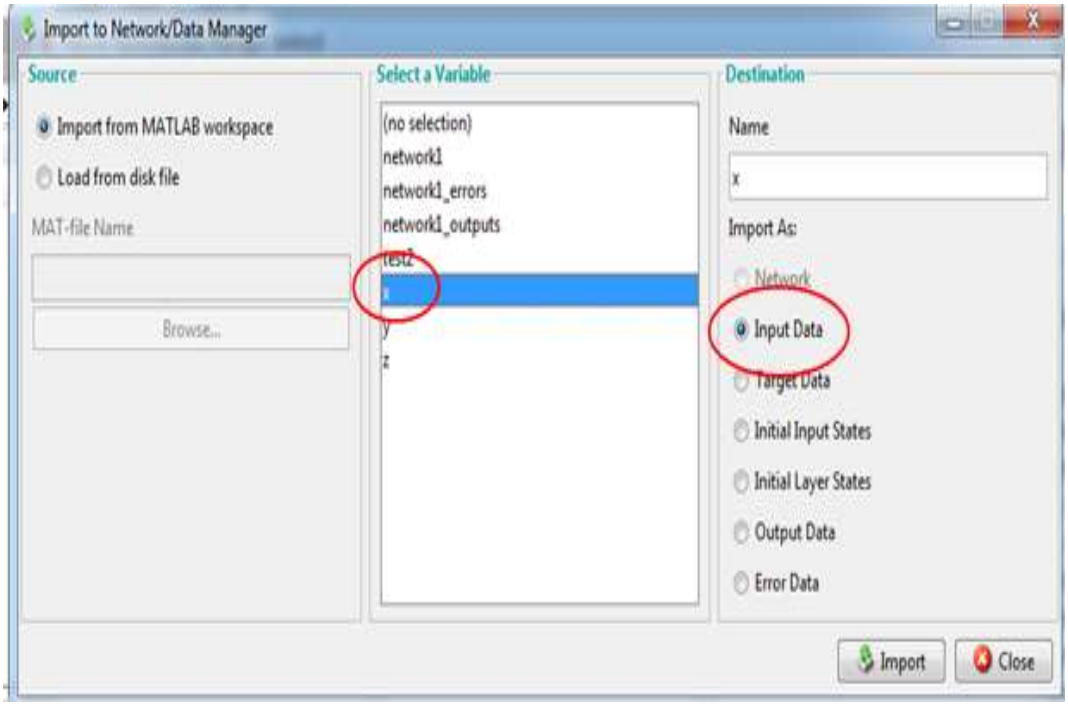



Figure 43

*Target Data on Network*

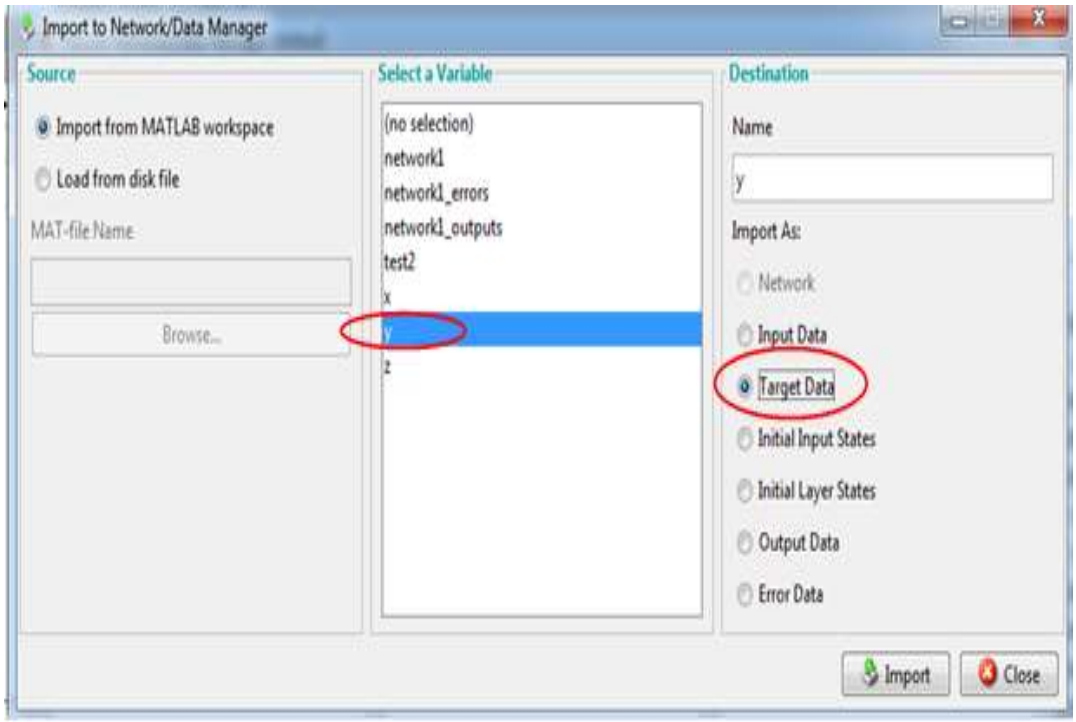
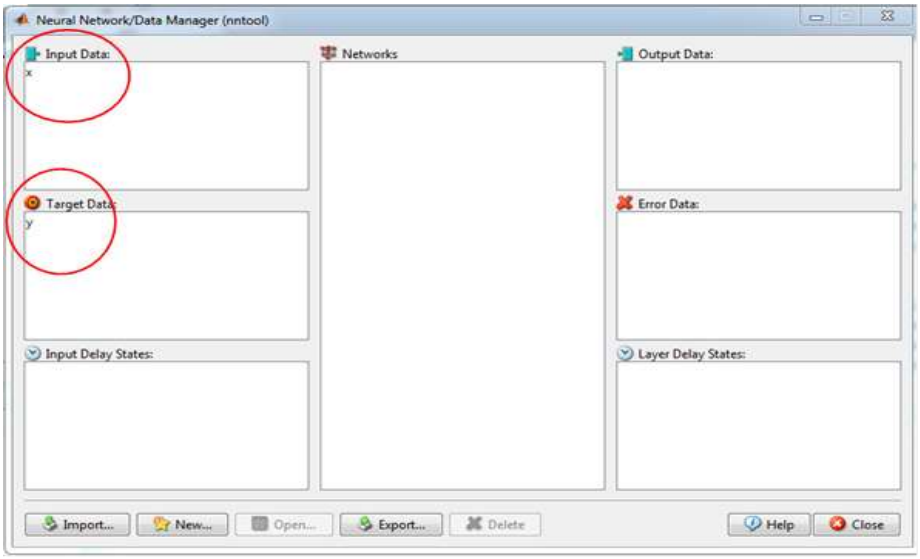


Figure 44

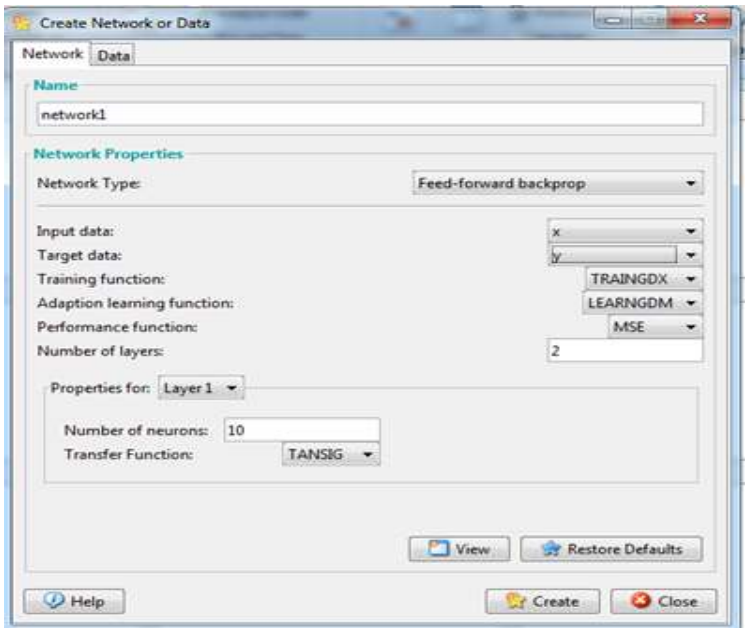
*Input and Output Data are Introduced to the System*




After the input and output data were introduced to the system, a new network was created for training. The transfer function was selected as LOGSIG because all data were positive (Figure 45). Finally, after all system settings were completed, the system was trained with a total of 268 different patient data.

Figure 45

*Method and System Settings of Network*

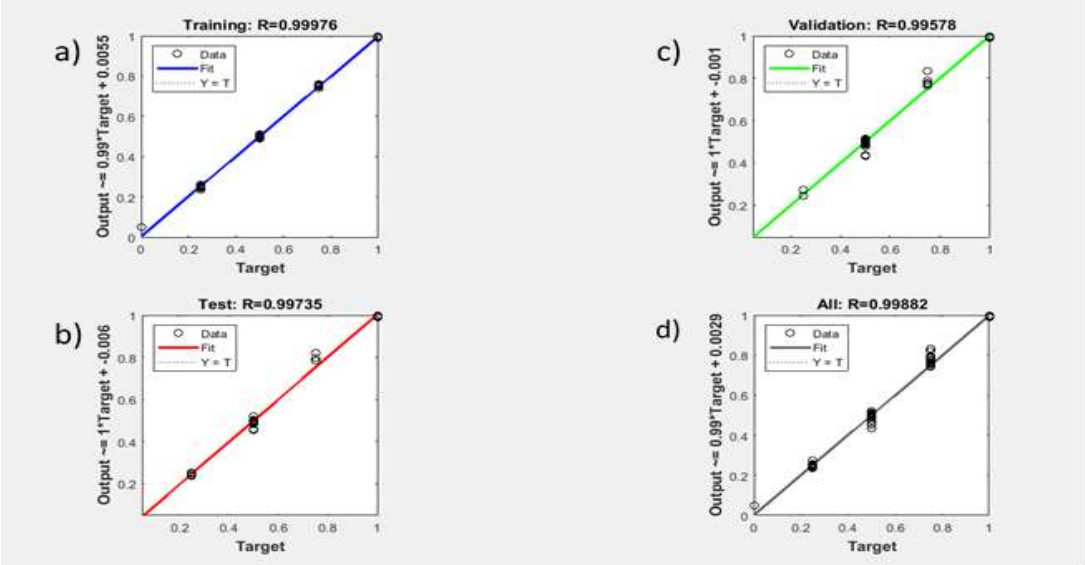


**Testing the Neural Network System**

Randomly selected 160 patients were used to train the neural network system, 54 patients were tested for the system and finally, remaining 54 patients therefore used to validate the created data. The training regression (success) was obtained as 99.9% (R = 0.99976). The test success of the system was calculated as 99.7% (R = 0.99735). The validation rate was achieved as ~99.6% (R = 0.99578). Thus, all regression were given as ~99.9% (R = 0.99882) (Figure 46). Thus, this overall result was compatible with the accurate result that obtained by fuzzy logic (95.5%).

Figure 46

*Neural Network Regression Results After Training*




In Figure 46 is shown Neural Network regression results of 268 patients. (a) The train success of the system using 160 patients (99.9%). (b) The test success of the system using 54 patient (99.7%). (c) The validation success of the system using remain 54 patients (~99.6%). (d) The overall success rate of the system (~99.9%). X-axis represented as output explain regressions data. Y-axis represented as target meaning success ratio between 0–1.

Six different tests were conducted for 12 different individuals in the test group to check the accuracy of the system. Best to notice that, these subjects were not introduced to the system before. Four patients had two different pathogenic variants within either *BRCA1* or *BRCA2*. Two subjects (subject 1 and 2) had

pathogenic *BRCA2* c.7698deIC variant (Table 11) and other two (subject 3 and 4) had *BRCA1* C.788dupG (p.Ser264*fs*1) pathogenic frameshift variant (Table 11). After entering the data for subject 1 and subject 2, the system calculated values as 0.99933 and 0.99938, respectively. According to the classification criteria and obtained values, the system was confirmed that both individuals were pathogenic. Subject 3 and subject 4, the system calculated values as 0.99933 and 0.99882, respectively. Subject 5 and 6 have the same variant as classified as likely pathogenic *BRCA1* c.4070_4071delAA (p.Glu135.7Glyfs*10). While the test was focused on two likely pathogenic variants, we got 0.77839 obtained value for subject 5 and 0.75161 for subject 6.

Table 11.

*Obtained Results from Testing the System*

| **Risk Factors** | **Test Subject 1 Variant: BRCA2 c.7698deIC Classification: Pathogenic** | **Test Subject 2** | **Test Subject 3 Variant: *BRCA1* C.788dupG Classification: Pathogenic** | **Test Subject 4** | **Test Subject 5 Variant: *BRCA1* c.4070_4071deIAA Classification: Likely Pathogenic** | **Test Subject 6** |
|---|---|---|---|---|---|---|
| **Age** | 43 | 36 | 44 | 42 | 34 | 33 |
| **Sex** | Female | Female | Female | Female | Female | Female |
| **Consanguineous Marriage** | Unknown | Unknown | Yes | Unknown | Yes | Unknown |
| **Family History** | Unknown | Unknown | Yes | Unknown | Yes | Yes |
| **Number of Affected Family Member** | Unknown | Unknown | 1 | Unknown | 1 | 3 |
| **Tumor Size** | 17.5 cm | 0-1 cm | Unknown | 6.6 cm | Unknown | Unknown |
| **Lymp Node** | No | No | Unknown | No | Unknown | Unknown |
| **Degree of Malignancy** | Unknown | Grade 2 | Unknown | Grade 3 | Unknown | Unknown |
| **Tumor Location** | Right Breast | Right Breast | Right Breast | Right Breast | Right Breast | Unknown |
| **Estrogen Receptor Hormone** | Unknown | Unknown | Unknown | Positive | Unknown | Unknown |
| **Progesterone Hormone** | Positive | Positive | Unknown | Negative | Unknown | Unknown |
| ***BRCA1*** | No | No | Yes | Yes | Yes | Yes |
| ***BRCA2*** | Yes | Yes | No | No | No | No |
| **Other Genes** | No | No | No | No | No | No |
| **Diagnosis** | Yes | Yes | Yes | Yes | Unknown | No |
| **Result** | 99,933% (0.99933) | 99,938% (0.99938) | 99,933% (0.99933) | 99,882% (0.99882) | 77,839% (0.77839) | 75,161% (0.75161) |

On the other hand, the system was tested for variants of unknown significance (VUS) such as *BRCA2* c.9924 A>G (p.Ile3312Val), *BRCA1* c.3368 A>G (p.Lys1290Glu) and *RAD50* c.379 G>A, respectively (Table 12). Therefore, we focused on making the correct estimation of individuals with VUS and possible identification of VUS variants. In the fourth test, we checked *BRCA2* c.9924 A>G (p.Ile3312Val) VUS variant in both individuals (subject 7 and subject 8). The system predicted 0.50281 and 0.49934 cancer risk scores for both individuals, respectively. Accordingly, the system confirmed that two individuals had VUS as they have been obtained within the VUS threshold. Subject 9 and subject 10 carry both *BRCA1* c.3368 A>G (p.Lys1290Glu) VUS variant according to ACMG criteria. The obtained value was 0.50542 for subject 9. However, subject 10 gave 0.5033 cancer risk score which was within the likely pathogenic threshold in our neural network system. Subject 11 and subject 12 carry same *RAD50* c.379 G>A VUS variant and system predicted 0.49998 value for subject 11 and 0.51068 for subject 12.

Table 12.

*Obtained VUS Results from Testing the System*

| **Risk Factors** | **Test Subject 7** | **Test Subject 8** | **Test Subject 9** | **Test Subject 10** | **Test Subject 11** | **Test Subject 12** |
|---|---|---|---|---|---|---|
| | ***BRCA2* c.9934 A>G** | | ***BRCA1* c.3368 A>G** | | **Variant: Variant: *RAD50* c.379 G>A** | |
| | **Classification: VUS** | | **Classification: VUS** | | **Classification: VUS** | |
| **Age** | 38 | 42 | 58 | 58 | 32 | 40 |
| **Sex** | Female | Female | Female | Female | Female | Female |
| **Consanguineous Marriage** | Unknown | No | Unknown | Unknown | Yes | No |
| **Family History** | No | No | Yes | No | No | No |
| **Number of Affected Family Member** | 0 | 0 | Unknown | 0 | 0 | 0 |
| **Tumor Size** | 3-4 cm | 0.5 cm | Unknown | Unknown | Unknown | 30 cm |
| **Lymp Node** | No | No | Unknown | Yes | Yes | Yes |
| **Degree of Malignancy** | Grade 3 | Grade 2 | Grade 2 | Grade 2 | Unknown | Grade 2 |
| **Tumor Location** | Right Breast | Right Breast | Right Breast | Right Breast | Right Breast | Both |
| **Estrogen Receptor Hormone** | Positive | Positive | Positive | Positive | Positive | Positive |
| **Progesterone Hormone** | Positive | Positive | Positive | Positive | Positive | Positive |

Table 12 (Continued).

| | | | | | | |
|---|---|---|---|---|---|---|
| ***BRCA1*** | Yes | No | Yes | Yes | No | Yes |
| ***BRCA2*** | Yes | Yes | No | No | No | No |
| **Other Genes** | No | No | No | No | Yes | Yes |
| **Diagnosis** | Yes | Yes | Yes | Yes | Yes | Yes |
| **Result** | 50,281% (0.50281) | 49,934% (0.49934) | 50,542% (0.50542) | 50,33% (0.5033) | 49,998% (0.49998) | 51,068% (0.51068) |

**Discussion**

In the study published in 2019, they developed a breast cancer diagnosis system using a hybrid model( a support vector machine and an artificial neural network). In this study, a total of 160 digital mammogram images of 80 normal breasts, 40 benign and 40 malignant cases were used to train the model. They achieved a 99.4% accuracy rate (Lim, et al., 2019). In another study published in 2019, Xie and Zhou developed a kernel support vector machine for cancer diagnosis. This model obtained 84% accuracy in diagnosis (Xie & Zhou, 2019).

On the other hand, a convolutional neural network (CNN) method was developed to boost the automatic identification of breast cancer by Alanazi and his colleagues (2021). The proposed system has been found successful by obtaining results with an accuracy of 87% (Alanazi, et al., 2021). Moreover, Zhang and his colleagues (2021) were used Convolutional Neural Network for breast cancer classification. They used 322 mammographic images to train the model, resulting in an accuracy of 96.10% (Zhang, et al., 2021).

In another study published recently, Kashyap (2022) used a deep learning-based classification model to classify histopathological images of breast cancer. He achieved a 95.65% accuracy rate (Kashyap, 2022). In our feed-forward neural network model, the success (accuracy) rate was achieved as ~99.9%.

Finally, we confirmed our variant risk assessment results, as in our fuzzy logic model, with the tests we performed with a test group of 12 individuals, which had not been introduced to the model before. In both models, variant risk assessment was performed with high success rates.

# CHAPTER V
# Discussions

## Introduction

Breast cancer is the most common cancer among women, affecting approximately 2.1 million women each year, causing the highest number of cancer-related deaths among women (Bray, et al., 2018). According to World Health Organization (WHO) records, it is estimated that more than 2.26 million new cases and 685,000 women died from breast cancer in 2020 (https://www.iarc.who.int/). Breast cancer is a disease caused by a tumor caused by changes in the cell groups that make up the breast tissue and their uncontrolled proliferation. The cancerous tissue first spreads to its immediate surroundings, then to the lymph nodes near the breast. In patients who are not diagnosed and treated on time, cancer spreads to other organs and passes to an impossible stage for treatment (Li, et al., 2020).

Artificial intelligence allows for more cost-effective, speedier, and practical medical diagnostics. Artificial intelligence will become more widely used as technology advances, particularly in medical diagnostics. In medicine, rapid diagnosis and treatment are critical for the prevention of many diseases, including cancer. Artificial intelligence applications have become increasingly important in this field in recent years. The adoption of high-throughput sequencing methods has amassed massive amounts of genetic variant data, as well as clinical and laboratory data from patients, in the previous decade. As a result, it is expected that the utilization of accumulated data in artificial intelligence applications would establish risk score evaluation for the most frequent cancer in women, breast cancer. Therefore, in this study, we aimed to evaluate the risk assessment for *BRCA1*- and *BRCA2*- associated breast cancer using Mamdani's fuzzy logic and feed-forward neural networks systems

## Association of *BRCA1* and *BRCA2* Genes with Breast Cancer

Cancers that occur in the genetic structure of the parents and are called mutations, are called hereditary cancers due to genetic differences that can pass to the next generations, their children. About 10% of all cancers are hereditary.

Mutations in the *BRCA1* and *BRCA2* genes are the major cause of hereditary breast and ovarian cancers. Cancer-related genes are divided into three classes. These are tumor suppressor genes, oncogenes and DNA repair genes (Scott, et al., 2019).

Tumor suppressor genes are protective genes. These genes normally limit the rate of cell division, repair damaged DNA, and control cell death. When these genes mutate, the cells continue to grow and eventually the tumor is formed. Approximately 30 tumor suppressor genes have been identified, including *BRCA1*, *BRCA2* (breast cancer), and *p53* genes (Hyder, et al., 2020).

Oncogenes are genes that transform healthy cells into cancerous cells. *HER2 / neu* (breast cancer) and *ras* are the two most common oncogenes (Vaklavas, et al., 2020). Finally, DNA repair genes repair errors that occur during DNA replication. Unrepaired errors lead to mutation and eventually cancer can develop (Özgöz, et al., 2019).

All humans have two copies of the *BRCA1* and *BRCA2* genes, one from the mother and the other from the father. The task of these two genes is to prevent the development of cancer in different parts of the body. Humans who have mutations in the *BRCA1* and *BRCA2* genes are at high risk of occurring breast cancer. Mutations in the *BRCA1* and *BRCA2* genes are the most important causes of hereditary breast cancers. Also, in individuals with the *BRCA1* and *BRCA2* mutation, cancer occurs earlier than other individuals, and there is a higher risk of developing secondary cancer in the same or a different organ (Hyder, et al., 2020).

However, we mainly focused on breast cancer associated with *BRCA1* and *BRCA2* genes. Fuzzy logic and neural network systems were trained by merging relevant risk factors and *BRCA* positive breast cancer patients' data. As a result, we achieved positive results for the use of physicians and clinicians for breast cancer variant risk assessment of breast cancer.

**Determining Risk Factors Caused Breast Cancer**

Risk factors are all factors that affect to occurring cancer. There are many different risk factors for breast cancer. Some factors, such as a person's age or race, cannot be changed, some others are linked to environmental factors that cause cancer. Additionally, some factors such as smoking, alcohol, and diet are related to

personal lifestyle, however, there are also genetic and hereditary factors that cause or increase the susceptibility to breast cancer.

In a study conducted, Ho and his colleagues (2020) worked on the risk factors of breast cancer. They examined the impact of risk factors, including body mass index (BMI), reproductive and hormonal risk factors, and breast density on the incidence of breast cancer. Also, they said reproductive, menstrual as menopausal status and age at menarche, hormone replacement therapy, and alcohol consumption are important risk factors for occurring breast cancer (Ho, et al., 2020).

In another study conducted in 2020, Hiller and his colleagues argued that there is a protective relationship between exposure to solar ultraviolet radiation (UVR) and the development of breast cancer. The main aim of the researchers in this study was to conduct a systematic literature review of the studies examining the relationship between sun UVR exposure and breast cancer risk. Accordingly, they conducted a meta-analysis with 13 different studies. Thus, they considered a breast cancer risk decreased for individuals who spend≥1h/d in the sun during summer months over a lifetime or adulthood period (Hiller, et al., 2020).

In another study conducted in 2017, Açıkgöz and Yıldız worked on the risk factors of breast cancer. They first mentioned gender and age risk. Compared to men, they argued that women had a lifetime risk of getting breast cancer nearly 100 times higher. Secondly, they focused on genetic predisposition and family history. They said that if the individuals have any cancer patients in their family, it increases cancer risk. In addition, they mentioned that many main and side factors such as menarche, menopause, obesity, radiation exposure, alcohol use, smoking were effective (Açıkgöz & Yıldız, 2017).

We analyzed data from 268 patients that we used to train our models, and 16 different risk factors were identified that suitable these data. These risk factors were age, sex, consanguinity, family history, number of family members, tumor size, lymph node, degree of malignancy, position, estrogen receptor, progesterone, *BRCA1* gene, *BRCA2* gene, other genes, diagnosis, and classification.

**Generating System Database and Creating Mamdani's Fuzzy Logic System on MATLAB**

Machine learning based on artificial intelligence was successfully used to classify cancer risk scores by Kaya and Turk (2020). They used a total of 140 data to test including 130 for test performance analysis and the remaining 10 for status determination (Kaya & Turk, 2020). In the current study, 268 different patients' data were trained in both fuzzy logic and neural network systems. Therefore, broader perspectives were used in both systems for the decision-making section whereas the risk of making errors were reduced.

A previous study was focused on cytological and histological image analysis in breast diseases for diagnostic outcomes using the fuzzy logic system on MATLAB (Berezsky, et al., 2019).

In 2018, A fuzzy logic system was used to predict breast cancer mortality with only five risk factors such as age, personal history, grade, malignant tumour classification (TNM) stage and multicentricity (Tawfeeq, 2018). Furthermore, Domingo et al. (2019) only used six risk factors on fuzzy logic for predicting the stages of breast cancer (Domingo, et al., 2019). As the variety of risk factors was quite low, they mainly focused on lymph nodes and tumours with a narrow perspective.

Mandang (2019) developed a program that could show the risk of breast cancer based on the fuzzy logic method using five histological risk factors for only young women (Mamdang, 2019). Controversially, the developed systems in this study can applied any age, gender.

Genetic variants are classified as pathogenic, likely pathogenic, VUS, likely benign, and benign according to ACMG (Richards, et al., 2015). However, problems arising from the evaluation and diagnosis of VUS variants have been a major challenge for physicians and geneticists today. More importantly, VUS variant carrier cancer patients cannot benefit from treatment processes. A study was designed to classify BRCA gene-related VUS variants in breast cancer using a statistical method, previously (Lindor, et al., 2011). In their study, the VUS variant was classified as either pathogenic or non-pathogenic.

VUS variants are crucial as the potential effect of the variant in the protein structure is either unknown or rare in the population or has not been registered before. VUS variant carrier cancer patients cannot benefit from treatment processes. Thus, the identification of VUS variants is important for precise treatment and targeted therapies. This developed Mamdani's Fuzzy logic model has been successful in characterization the pathogenicity of VUS variants.

Mamdani's fuzzy logic system in this study was designed and trained to give risk scores to VUS variants using other clinical outcomes of the patient. Moreover, systems were tested for 12 new individuals including two pathogenic, one likely pathogenic and three VUS. Tests scores obtained, respectively: *BRCA2* c.7698deIC (90%, 89%), *BRCA1* C.788dupG (90%, 90%), *BRCA1* c.4070_4071deIAA (66.1%, 66.1%), *BRCA2* c.9934 A>G (42.5%, 48.9%), *BRCA1* c.3368 A>G (48.9%, 57.1%), *RAD50* c.379 G>A (42.5%, 42.5%). The fuzzy logic system accuracy rate was achieved as 95.5%.

**Generating System Database and Creating Feed-Forward Neural Network System on MATLAB**

On the other hand, the neural network system was previously proposed to diagnose breast cancer patterns using histological and demographic characteristics, such as Toğaçar et al. (2020) investigated the diagnostic process based on histological image analysis in breast cancer using deep learning and a convolutional neural network giving a success rate of 98.80% (Toğaçar, et al., 2020). In another study, a hybrid deep neural network with artificial intelligence was successfully used to classify breast cancer risk scores by Yan et al. (2021) based on histological image classification and an average accuracy was 91.3% (Yan, et al., 2020). Zhang et al. (2020) investigated three breast cancer molecular subtypes based on DCE-MRI images using a convolutional neural network (Zhang, et al., 2020). Thus, in this study, 16 different risk factors were used with the aim of obtaining more accurate results affecting breast cancer with a broader perspective to give more significant value than similar studies in the literature.

In another study, the authors investigated the diagnosis based on automated breast ultrasound images in breast cancer using the convolutional neural network (CNN) technique, which is classified as benign or malignant in breast cancer lesions

(Wang, et al., 2020). The subjective decision of physicians was the major disadvantage of diagnosis. Hence, it is critical to developing a model to verify the preliminary diagnosis.

In 2019, Choi *et al.* were used a convolutional neural network system for predicting the pathological response to neoadjuvant chemotherapy in advanced breast cancer (Choi, et al., 2019), and genetic factors were not included. Furthermore, Zhou *et al.* were used a deep learning technique to predict the negative axillary lymph node metastasis in primary breast cancer patients in 2020 (Zhou, et al., 2020) and genetic factors were not included.

The most important key point of the study was the risk assessment of two designed different artificial intelligence methods were based on cancer-associated genes and gene variants. Two recent studies aimed to predict breast cancer using histopathology and radiology images for *BRCA*-mutation carriers using deep learning (Wang, et al., 2021) and machine learning (Gullo, et al., 2020), respectively. This current study mainly focused on gene variant-based risk assessment in cancer.

The feed-forward Neural network system in this study was designed and trained to give risk scores to VUS variants using other clinical outcomes of the patient. Therefore, physicians can evaluate VUS variant with a given risk score 87 patients with pathogenic, 23 with likely pathogenic, 128 VUS, 29 likely benign, and 1 benign BRCA1 and BRCA2 gene variants together with 14 other clinical breast cancer risk factors. Moreover, systems were tested for 12 new individuals including two pathogenic, one likely pathogenic, and three VUS. Tests scores obtained, respectively: *BRCA2* c.7698deIC (99,933%, 99,938%), *BRCA1* C.788dupG (99,933%, 99,882%), *BRCA1* c.4070_4071deIAA (77,839%, 75,161%), *BRCA2* c.9934 A>G (50,281%, 49,934%), *BRCA1* c.3368 A>G (50,542%, 50,33%), *RAD50* c.379 G>A (49,998%, 51,068%). The neural network system overall success rate was achieved as 99.9% whereas training success (99.9%), evaluating validation success (99.6%), test success (99.7%).

# CHAPTER VI
## Conclusion

Firstly, two large clinical data from two different universities were collected and analyzed. Then to determine risk factors that are associated with *BRCA1* and *BRCA2* breast cancers. Additionally, determined risk factors for each patient data have been merged using two different artificial intelligence models on MATLAB.

Secondly, models were trained using these clinical data and risk factors. After that, models took on a structure that thought and decided by themselves. At the last stage, the testing phase was started to check that these developed models work correctly. During the test phase, a test group consisting of 12 individuals who had not been introduced to the system before was used.

Finally, in this study, developed fuzzy logic and neural networks models were found to be successful in predicting correct risk scores for *BRCA1* and *BRCA2* associated breast cancers, especially classifying VUS variants. Thus, we believe that the generated fuzzy logic system will become a good source for the identification of VUS variants in breast cancer diagnosis. To conclude, the artificial intelligence model will provide significant advantages considering an early diagnosis and personalized therapy are vital in cancer.

To conclude, the developed artificial intelligence models will facilitate the work of physicians as improving preventive medicine approaches as well as and is a unique tool for today's personalized medicine software. Furthermore, our models provide quick results to variant risk assessment of breast cancer for clinicians and physicians.

## References

Açıkgöz, A., & Yıldız, E. A. (2017). Etiology and Risk Factors of Breast Cancer. *Ergoterapi ve Rehabilitasyon Dergisi, 5(1),* 45-56.

Adachi, M., Fujioka, T., Mori, M., Kubota, K., Kikuchi, Y., Xiaotong, W., ... & Tateishi, U. (2020). Detection and diagnosis of breast cancer using artificial intelligence based assessment of maximum intensity projection dynamic contrast-enhanced magnetic resonance images. *Diagnostics*, *10*(5), 330.

Adalı, T., & Şekeroğlu, B. (2012). Analysis of micrornas by neural network for early detection of cancer. *Procedia Technology*, *1*, 449-452.

Aggarwal, S., Sharma, R., & Aggarwal, A. (2018). Abstract P4-06-14: Voice of cancer patients: Association of family history with genetic mutations in breast and ovarian cancer. *Cancer Research, 78(4),* 4-14.

Alanazi, S. A., Kamruzzaman, M. M., Islam Sarker, M. N., Alruwaili, M., Alhwaiti, Y., Alshammari, N., & Siddiqi, M. H. (2021). Boosting breast cancer detection using convolutional neural network. *Journal of Healthcare Engineering*, *2021*.

An, K.Y., Morielli, A.R., Kang, D.W., Friedenreich, C.M., McKenzie, D.C., Gelmon, K., Mackey, J.R., Reid, R.D., & Courneya, K.S. (2020). Effects of exercise dose and type during breast cancer chemotherapy on longer-term patient-reported outcomes and health-related fitness: A randomized controlled trial. *International journal of cancer, 146(1),* 150-160.

Arthur, R.S., Wang, T., & Rohan, T. (2019). The interplay between lifestyle-related factors and genetics with risk of invasive breast cancer among postmenopausal women from the UK Biobank. *AACR Annual Meeting, 79(13),* 18-963.

Ayhan, A. (2013). Kanserin Moleküler Temeli ve Jinekolojik Kanserlerde Önemi. Ankara: Güneş Kitabevi.

Bakr, M. A. H. A., Al-Attar, H. M., Mahra, N. K., & Abu-Naser, S. S. (2020). Breast Cancer Prediction using JNN. *International Journal of Academic Information Systems Research (IJAISR), 4*(10).

Baratloo, A., Hosseini, M., Negida, A., & El Ashal, G. (2015). Part 1: simple definition and calculation of accuracy, sensitivity and specificity.

Belciug, S. (2020). Logistic regression paradigm for training a single-hidden layer feedforward neural network. Application to gene expression datasets for cancer research. *Journal of Biomedical Informatics, 102*, 103373.

Berezsky, O., Dubchak, L., Batryn, N., Berezska, K., Pitsun, O., Batko, Y., & Datsko, T. (2019). Fuzzy system for breast disease diagnosing based on image analysis. In *CEUR Workshop Proceedings* (pp. 69-83).

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians, 68*(6), 394-424.

Cheung, W., McLenachan, S., Babu, G., & Smith, M. (2015). Accuracy of axillary nodal ultrasound and ultrasound fine needle aspiration/core biopsy in the preoperative staging of patients with invasive breast cancer. *Breast Cancer Research, 17(1),* 1-13.

Choi, J.H., Kim, W., Lim, I, Woo, S.K., Lee, I, Byun, B.H., Noh, W.C., Ah, K.H., Seong, M.K., Kim, B.I, Choi, C.W., & Lim, S.M. (2019). Early Prediction of Pathologic Response to Neoadjuvant Chemotherapy in Advanced Breast Cancer using Convolutional Neural Network Algorithm. *The Journal of Nuclear Medicine, 60(1),* 599-599.

Copson, E.R., Maishman, T.C., Tapper, W.J., Cutresss, R.I., Heygate, S.G., Altman, D.G., Eccles, B., Gerty, S., Durcan, L.T., Jones, L., Evans, G., Thompson, A.M., Pharoah, P., Easton, D.F., Dunning, A.M., Hanby, A., Lakhani, S., Eeles, R., Gilbert, F.J., Hamed, H., Hodgson, S., Simmonds, P., Stanton, L., & Eccles, D.M. (2018). Germline *BRCA* mutation and outcome in young-

onset breast cancer (POSH): a prospective cohort study. *The Lancet Oncology, 19(2),* 169-180.

Dale, R. (2021). GPT-3: What's it good for?. *Natural Language Engineering*, *27*(1), 113-118.

De Silva, S., Tennekoon, K. H., & Karunanayake, E. H. (2019). Overview of the genetic basis toward early detection of breast cancer. *Breast Cancer: Targets and Therapy*, *11*, 71.

Domingo, M. J., Gerardo, B. D., & Medina, R. P. (2019, November). Fuzzy decision tree for breast cancer prediction. In *Proceedings of the International Conference on Advanced Information Science and System* (pp. 1-6).

E Ellsworth, R., J Decewicz, D., D Shriver, C., & L Ellsworth, D. (2010). Breast cancer in the personal genomics era. *Current genomics*, *11*(3), 146-161.

Feng, Y., Pan, L., Zhang, B., Huang, H., & Ma, H. (2019). BATF acts as an oncogene in non-small cell lung cancer. *Oncology Letters, 19(1),* 205-210.

GBD 2015 Risk Factors Collaborators (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet, 388 (10053),* 1659-1724.

Ghazal, T. M. (2021). Internet of Things with Artificial Intelligence for Health Care Security. *Arabian Journal for Science and Engineering*, 1-12.

Gresner, P., Jablonska, E., & Gromadzinska, J. (2020). Rad51 paralogs and the risk of unselected breast cancer: A case-control study. *PloS one*, *15(1),* 1-29.

Gullo, R. L., Daimiel, I., Saccarelli, C. R., Bitencourt, A., Gibbs, P., Fox, M. J., ... & Pinker, K. (2020). Improved characterization of sub-centimeter enhancing breast masses on MRI with radiomics and machine learning in BRCA mutation carriers. *European radiology*, *30*(12), 6721-6731.

Hiller, T.W., O'Sullivan, D.E., Brenner, D.R., Peters, C.E., & King, W.D. (2020). Solar Ultraviolet Radiation and Breast Cancer Risk: A Systematic Review and Meta-Analysis. *Environmental Health Perspectives, 128(1),* 016002.

Ho, P.J., Lau, H.S.H., Ho, W.K., Wong, F.Y., Yang, Q., Tan, K.W., Tan, M.H., Chay, W.Y., Chia, K.S., Hartman, M., & Li, J. (2020). Incidence of breast cancer attributable to breast density, modifiable and non-modifiable breast cancer risk factors in Singapore. *Scientific Reports, 10(1),* 1-11.

HomaieFasih, A., Atashi, A., & Ahsani Tehrani, H. (2020). Prediction of Breast Cancer Metastasis Using Fuzzy Models based on Data from Iranian Breast Cancer Patients. *Journal of Health and Biomedical Informatics*, *7*(2), 181-189.

Hu, W., Wu, X., Tang, J., Zhao, G., Xiao, N., Zhang, L., & Li, S. (2019). Anti-cancer targets of formononetin and molecular mechanisms in osteosarcoma: Findings of bioinformatic and experimental assays. *Journal of Cellular and Molecular Medicine*, *23(5),* 3505-3511.

Hyder, Z., Harkness, E.F., Woodward, E.R., Bowers, N.L., Pereira, M., Wallace, A.J., Howell, S.J., Howell, A., Lalloo, F., Newman, W.G., Smith, M.J., & Evans, D.G. (2020). Risk of Contralateral Breast Cancer in Women with and without Pathogenic Variants in BRCA1, BRCA2, and TP53 Genes in Women with Very Early-Onset (< 36 Years) Breast Cancer. *Cancers, 12(2),* 378.

Jijitha, S., & Amudha, T. (2021). Breast cancer prognosis using machine learning techniques and genetic algorithm: experiment on six different datasets. In *Evolutionary Computing and Mobile Sustainable Networks* (pp. 703-711). Springer, Singapore.

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research*.

Kang, T., Yau, C., Benz, S., Krings, G., Camarda, R., Henry, J.E., Powell, M., & Benz, C.C. (2019). Normal breast tissue at risk for cancer development: A

breast cancer initiating role for mammary adipocytes. *AACR Annual Meeting, 79(13),* 3314.

Kashyap, R. (2022). Breast Cancer Histopathological Image Classification Using Stochastic Dilated Residual Ghost Model. *International Journal of Information Retrieval Research (IJIRR), 12*(1), 1-24.

Kaya, D., & Turk, M. (2020). LabVIEW based robust cascade predictive model for evaluating cancer prognosis. *Physica A: Statistical Mechanics and its Applications, 549,* 123978.

Kotsopoulos, J., Lubinski, J., Lynch, H.T., Tung, N., Armel, S., Senter, L., Singer, C.F., Fruscio, R., Couch, F., Weitzel, J.N., Karlan, B., Foulkes, W.D., Moller, P., Eisen, A., Ainsworth, P., Neuhausen, S.L., Olopade, O., Sun, P., Gronwald, J., & Narod, S.A. (2019). Oophorectomy and risk of contralateral breast cancer among *BRCA1* and *BRCA2* mutation carriers. *Breast Cancer Research and Treatment, 175(2),* 443-449.

Kumar, R., Singh, A. K., & Mukaherjee, S. (2021). Review of an EMG-Controlled Prosthetic Arm. In *Computational Methodologies for Electrical and Electronics Engineers* (pp. 85-92). IGI Global.

Li, M., Warner-Smith, M., McGill, S., Roder, D., & Currow, D. (2020). History of screening by BreastScreen New South Wales of women with invasive breast cancer. *Cancer Epidemiology, 64,* 101659.

Lilyquist, J., Ruddy, K.J., Vachon, C.M., & Fergus J. Couch, F.J. (2018). Common Genetic Variation and Breast Cancer Risk Past, Present, and Future. *Cancer Epidemiology, Biomarkers & Prevention, 27(4),* 1117-1144.

Lim, T. S., Tay, K. G., Huong, A., & Lim, X. Y. (2021). Breast cancer diagnosis system using hybrid support vector machine-artificial neural network. *International Journal of Electrical & Computer Engineering (2088-8708), 11*(4).

Lindor, N. M., Guidugli, L., Wang, X., Vallée, M. P., Monteiro, A. N., Tavtigian, S., ... & Couch, F. J. (2012). A review of a multifactorial probability-based

model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Human mutation*, *33*(1), 8-21.

Mandang, I. (2019). Diagnosis study of carcinoma mammae (breast cancer) disease using fuzzy logic method. In *Journal of Physics: Conference Series* (Vol. 1277, No. 1, p. 012039). IOP Publishing.

Maxwell, K.N., Hart, S.N., Vijai, J., Schrader, K.A., Slavin, T.P., Thomas, T., Wubbenhorst, B., Ravichandran, V., Moore, R.M., Hu, C., Guidugli, L., Wenz, B., Domchek, S.M., Robson, M.E., Szabo, C., Neuhausen, S.L., Weitzel, J.N., Offit, K., Couch, F.J., & Nathanson, K.L. (2016). Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer Associated Genes in Families Affected by Breast Cancer. *The American Journal of Human Genetics, 98*, 801–817.

McGrath, J., Panzica, L., Ransom, R., Withers, H.G., & Gelman, I.H. (2019). Identification of Genes Regulating Breast Cancer Dormancy in 3D Bone Endosteal Niche Cultures. *Molecular Cancer Research,* doi: 10.1158/1541-7786.MCR-18-0956.

Militello, C., Ranieri, A., Rundo, L., D'Angelo, I., Marinozzi, F., Bartolotta, T. V., ... & Russo, G. (2022). On Unsupervised Methods for Medical Image Segmentation: Investigating Classic Approaches in Breast Cancer DCE-MRI. *Applied Sciences*, *12*(1), 162.

Nykamp, K., Anderson, M., Powers, M., Garcia, J., Herrera, B., Ho, Y. Y., ... & Topper, S. (2017). Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria. Genetics in Medicine, 19(10), 1105-1117.

Özgöz, A., Öztürk, K.H., Yükseltürk, A., Şamlı, H., Başkan, Z., İçduygu, F.M., & Bacaksız, M. (2019). Genetic variations of DNA repair genes in breast cancer. *Pathology & Oncology Research, 25(1),* 107-114.

Öztemur, Y., Aydos, A., & Gürdedeoğlu, B. (2015). Combination of breast cancer microarray data by using bioinformatic methods - Meta-analysis approaches. *Türk Hijyen ve Deneysel Biyoloji Dergisi, 71(2),* 155-162.

Padmavathy, T. V., Vimalkumar, M. N., & Bhargava, D. S. (2019). Adaptive clustering based breast cancer detection with ANFIS classifier using mammographic images. *Cluster Computing*, *22*(6), 13975-13984.

Phillips, K.A., Milne, R.L., Bassett, J.K., Hopper, J.L., Buys, S.S., Daly, M.B., Hooning, M.J., Mooji, T.M., Andrieu, N., Antoniou, A.C., Rookus, M.A., Easton, D.F., & Beth, T.M. (2018). Tamoxifen and contralateral breast cancer (CBC) risk for *BRCA1* and *BRCA2* mutation carriers: An updated analysis of data from the Kathleen Cuningham Foundation consortium for research into familial breast cancer, the International *BRCA1* and *BRCA2* Carrier cohort study and the breast cancer family registry. *Cancer Research, 78(4),* 10.

Piccialli, F., Giampaolo, F., Prezioso, E., Camacho, D., & Acampora, G. (2021). Artificial intelligence and healthcare: Forecasting of medical bookings through multi-source time-series fusion. *Information Fusion*, *74*, 1-16.

Potter M.C., Lessing J.L., & Aboufadel E.F. (2019). Matrix Applications. In: Advanced Engineering Mathematics, Switzerland: Springer.

Pourjavad, E., & Mayorga, R.V. (2019). A comparative study and measuring performance of manufacturing systems with Mamdani fuzzy inference system. *Journal of Intelligent Manufacturing, 30(3),* 1085-1097.

Poznyak, T., Oria, J.I.C., & Poznyak, A. (2018). Ozonation and Biodegradation in Environmental Engineering: Dynamic Neural Network Approach. Amsterdam: Elsevier Science.

Quy, H. D., Van Kien, C., Anh, H. P. H., & Son, N. N. (2020, November). Machine learning-based evolutionary neural network approach applied in breast cancer tumor classification. In *International conference on green technology and sustainable development* (pp. 72-83). Springer, Cham.

Rahman, M. A., chandren Muniyandi, R., Albashish, D., Rahman, M. M., & Usman, O. L. (2021). Artificial neural network with Taguchi method for robust classification model to improve classification accuracy of breast cancer. *PeerJ Computer Science*, *7*, e344.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, B., Foster, J. G., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine, 17,* 405-423.

Rudolph, A., Song, M., Brook, M.N., Milne, R.L., Mavaddat, N., Michailidou, K., Bolla, M.K., Wang, Q., Dennis, J., Hopper, J.L., Southey, M.C., Keeman, R., Fasching, P.A., Beckmann, M.W., Dominguez, M.G., Castelao, J.E., Guénel, P., Truong, T., Bojesen, S.E., Flyger, H., Brenner, H., Arndt, V., Brauch, H., Brüning, T., Mannermaa, A., Kosma, V., Lambrechts, D., Keupers, M., Couch, F.J., Vachon, C., Giles, G.G., MacInnis, R.J., Figueroa, J., Brinton, L., Czene, K., Brand, J.S., Gabrielson, M., Humphreys, K., Cox, A., Cross, S.S., Dunning, A.M., Orr, N., Swerdlow, A., Hall, P., Pharoah, P.D.P., Schmidt, M.K., Easton, D.F., Chatterjee, N., Claude, J.C., & Closas, M.G. (2018). Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. *International Journal of Epidemiology, 47(2),* 526-536.

Rué i Monné, M., Vilaprinyo Terré, E., Lee, S., Martínez Alonso, M., Carles Lavila, M., Marcos-Gragera, R., ... & Espinàs, J. A. (2009). Effectiveness of early detection on breast cancer mortality reduction in Catalonia (Spain). *BMC Cancer, 2009, vol. 9, núm. 326, p. 1-10*.

Sandeep, D., & Bethel, G. B. (2021, November). Accurate Breast Cancer Detection and Classification by Machine Learning Approach. In *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 366-371). IEEE.

Scott, C.J., Schoeman, M., & Urban, M.F. (2019). Cancer genetics: An approach to suspected hereditary breast or colorectal cancer. *SAMJ: South African Medical Journal, 109(4),* 214-218.

Shaheen, M. Y. (2021). Adoption of machine learning for medical diagnosis. *ScienceOpen Preprints*.

Shoumi, M. N. M., & Syulistyo, A. R. (2021). Analisis Resiko Kanker Payudara (Breast Cancer) Menggunakan Fuzzy Inference System (FIS) Model Mamdani. *Informasi Interaktif*, *6*(1), 25-30.

Smith, S.G., Sestak, I., Forster, A., Partridge, A., Side, L., Wolf, M.S., Horne, R., Wardle, J., & Cuzick, J. (2015). Factors affecting uptake and adherence to breast cancer chemoprevention: a systematic review and meta-analysis. *Annals of Oncology, 27(4),* 575-590.

Stewart, B.W., & Wild C.P. (2014). World cancer report 2014 Lyon. *International Agency for Research on Cancer*.

Sultana, Z., Khan, M. R., & Jahan, N. (2021). Early breast cancer detection utilizing artificial neural network. *WSEAS Transactions on Biology and Biomedicine, 18,* 32-42.

Tavtigian, S. V., Greenblatt, M. S., Harrison, S. M., Nussbaum, R. L., Prabhu, S. A., Boucher, K. M., & Biesecker, L. G. (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genetics in Medicine, 20(9), 1054-1060.

Tawfeeq, F.N. (2018). Development of Prognosis Factors in a Scoring System for Predicting of Breast Cancer Mortality. *Journal of Information Engineering and Applications, 8(4),* 43-50.

Toğaçar, M., Özkurt, K. B., Ergen, B., & Cömert, Z. (2020). BreastNet: a novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. *Physica A: Statistical Mechanics and its Applications*, *545*, 123592.

Tran, W. T., Sadeghi-Naini, A., Lu, F. I., Gandhi, S., Meti, N., Brackstone, M., ... & Curpen, B. (2021). Computational radiology in breast cancer screening and diagnosis using artificial intelligence. *Canadian Association of Radiologists Journal*, *72*(1), 98-108.

Tuan, T. M., Lan, L. T. H., Chou, S. Y., Ngan, T. T., Son, L. H., Giang, N. L., & Ali, M. (2020). M-CFIS-R: Mamdani complex fuzzy inference system with rule reduction using complex fuzzy measures in granular computing. *Mathematics*, *8*(5), 707.

Vaklavas, C., Blume, S.W., & Grizzle, W.E. (2020). Hallmarks and Determinants of Oncogenic Translation Revealed by Ribosome Profiling in Models of Breast Cancer. *Translational Oncology, 13(2),* 452-470.

Vijayakumar, K., Kadam, V. J., & Sharma, S. K. (2021). Breast cancer diagnosis using multiple activation deep neural network. *Concurrent Engineering*, *29*(3), 275-284.

Wang, X., Zou, C., Zhang, Y., Li, X., Wang, C., Ke, F., ... & Zhang, Y. (2021). Prediction of BRCA Gene Mutation in Breast Cancer Based on Deep Learning and Histopathology Images. *Frontiers in Genetics*, 1147.

Wang, Y., Choi, E. J., Choi, Y., Zhang, H., Jin, G. Y., & Ko, S. B. (2020). Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. *Ultrasound in medicine & biology*, *46*(5), 1119-1132.

Wasif, N., Neville, M., Gray, R., Cronin, P., & Pockaj, B.A., (2019). Competing Risk of Death in Elderly Patients with Newly Diagnosed Stage I Breast Cancer. *Journal of the American College of Surgeons, 229(1),* 30-36.

Wolfe, C. R., Reyna, V. F., Widmer, C. L., Cedillos, E. M., Fisher, C. R., Brust-Renck, P. G., & Weil, A. M. (2015). Efficacy of a web-based intelligent tutoring system for communicating genetic risk of breast cancer: A fuzzy-trace theory approach. *Medical Decision Making*, *35*(1), 46-59.

Wu, S., Powers, S., Zhu, W., & Hannun, Y. A. (2016). Substantial contribution of extrinsic risk factors to cancer development. *Nature, 529,* 43-47.

Xie, T., Yao, J., & Zhou, Z. (2019). DA-based parameter optimization of combined kernel support vector machine for cancer diagnosis. *Processes*, *7*(5), 263.

Yan, R., Ren, F., Wang, Z., Wang, L., Zhang, T., Liu, Y., ... & Zhang, F. (2020). Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*, *173*, 52-60.

Yang, Z., Gan, Z., Wang, J., Hu, X., Lu, Y., Liu, Z., & Wang, L. (2021). An empirical study of gpt-3 for few-shot knowledge-based vqa. *arXiv preprint arXiv:2109.05014*.

Zhang, Y., Chen, J. H., Lin, Y., Chan, S., Zhou, J., Chow, D., ... & Su, M. Y. (2021). Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers. *European Radiology*, *31*(4), 2559-2567.

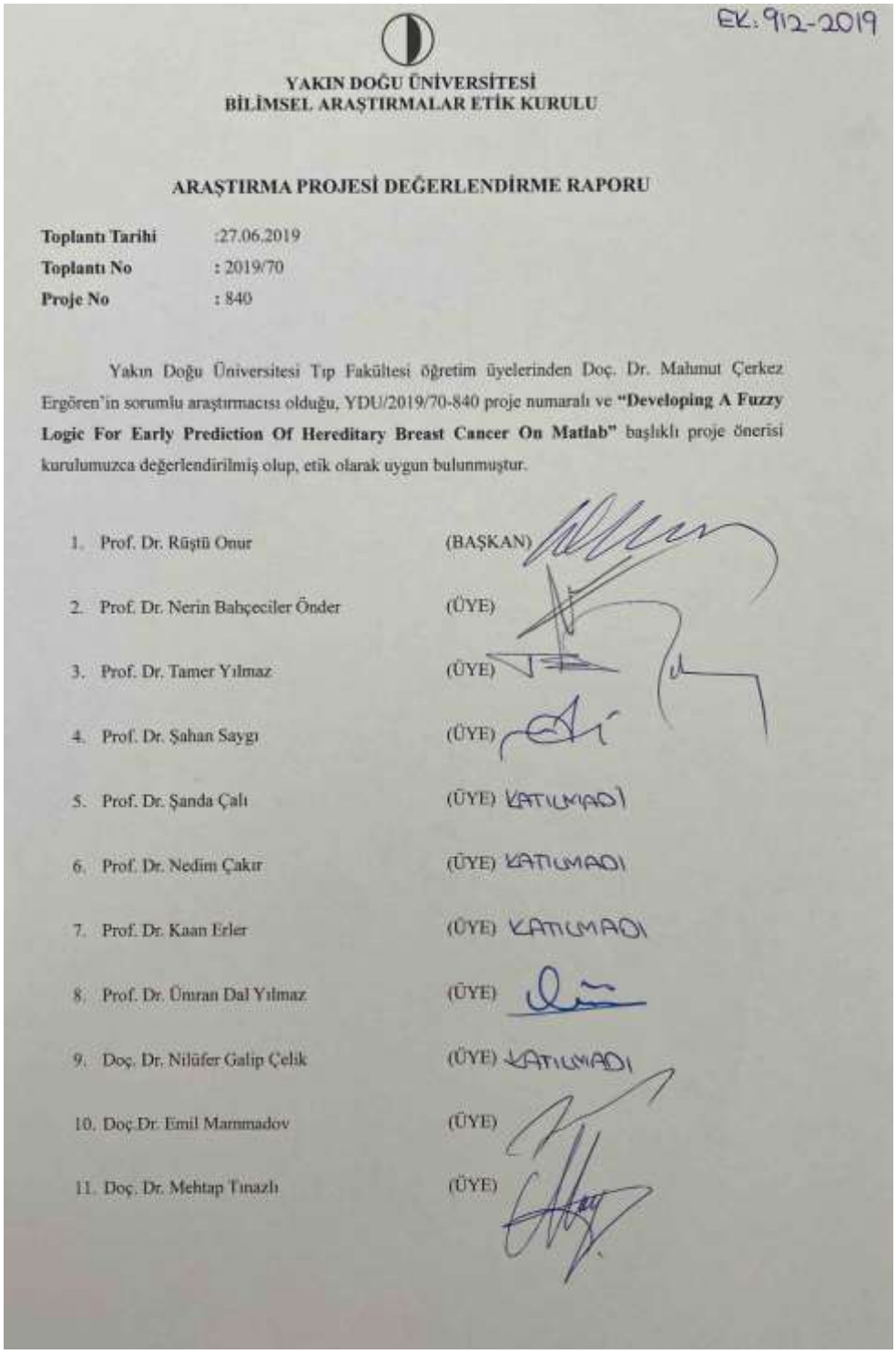
Zhang, Y. D., Satapathy, S. C., Guttery, D. S., Górriz, J. M., & Wang, S. H. (2021). Improved breast cancer classification through combining graph convolutional network and convolutional neural network. *Information Processing & Management*, *58*(2), 102439.

Zhou, L.Q., Wu, X.L., Huang, S.Y., Wu, G.G., Ye, H.R., Wei, Q., Bao, L.Y., Deng, Y.B., Li, X.R., Cui, X.W., & Dietrich, C.F. (2020). Lymph node metastasis prediction from primary breast cancer US images using deep learning. *Radiology*, *294(1),* 19-28.

# Appendices

## Appendix A

## Ethical Approval Document



EK: 912-2019

YAKIN DOĞU ÜNİVERSİTESİ
BİLİMSEL ARAŞTIRMALAR ETİK KURULU

ARAŞTIRMA PROJESİ DEĞERLENDİRME RAPORU

**Toplantı Tarihi** :27.06.2019
**Toplantı No** : 2019/70
**Proje No** : 840

Yakın Doğu Üniversitesi Tıp Fakültesi öğretim üyelerinden Doç. Dr. Mahmut Çerkez Ergören'in sorumlu araştırmacısı olduğu, YDU/2019/70-840 proje numaralı ve **"Developing A Fuzzy Logic For Early Prediction Of Hereditary Breast Cancer On Matlab"** başlıklı proje önerisi kurulumuzca değerlendirilmiş olup, etik olarak uygun bulunmuştur.

1. Prof. Dr. Rüştü Onur (BAŞKAN)
2. Prof. Dr. Nerin Bahçeciler Önder (ÜYE)
3. Prof. Dr. Tamer Yılmaz (ÜYE)
4. Prof. Dr. Şahan Saygı (ÜYE)
5. Prof. Dr. Şanda Çalı (ÜYE) KATILMADI
6. Prof. Dr. Nedim Çakır (ÜYE) KATILMADI
7. Prof. Dr. Kaan Erler (ÜYE) KATILMADI
8. Prof. Dr. Ümran Dal Yılmaz (ÜYE)
9. Doç. Dr. Nilüfer Galip Çelik (ÜYE) KATILMADI
10. Doç.Dr. Emil Mammadov (ÜYE)
11. Doç. Dr. Mehtap Tınazlı (ÜYE)

# Appendix B

## Signed Similarity Report

| AUTHOR | TITLE | SIMILARITY |
|---|---|---|
| Niyazi Şentürk | Abstract | 0% |
| Niyazi Şentürk | Chapter 3 - Results of Fuzzy Logic Syste... | 0% |
| Niyazi Şentürk | Chapter 4 - Results of Neural Network Sy... | 0% |
| Niyazi Şentürk | Chapter 6 - Conclusion | 0% |
| Niyazi Şentürk | Chapter 5 - Discussions | 10% |
| Niyazi Şentürk | Chapter 2 - Materials and Methods | 12% |
| Niyazi Şentürk | Entire Thesis | 13% |
| Niyazi Şentürk | Chapter 1 - Introduction | 14% |

Assoc. Prof. Mahmut Çerkez Ergören

# Appendix C

## Curriculum Vitae

### PERSONAL INFORMATIONS

| | |
|---|---|
| Surname, Name: | Şentürk, Niyazi |
| Date of Birth: | 26 April 1993 |
| Place of Birth: | Nicosia, Cyprus |

### EDUCATION

| Degree | Department/Program | University | Year of Graduation |
|---|---|---|---|
| M.Sc. | Biomedical Engineering | Near East University | 2018 |
| B.Sc. | Information Systems Engineering | Eastern Mediterranean University | 2016 |

**Master Thesis Title:** Developing an Online Portal for Unraveling Genomic Signature of Archaic DNA that are Related to Modern Human Genetic Disease.

### WORK EXPERIENCE

| Title | Place | Year |
|---|---|---|
| Lecturer | NEU, Faculty of Engineering, Department of Biomedical Engineering | 2018-present |
| Research Assistant | NEU, Faculty of Engineering, Department of Biomedical Engineering | 2016-2017 |

### FOREIGN LANGUAGES

Fluent spoken and written English

### PUBLICATIONS IN INTERNATIONAL REFERED JOURNALS

- "*BRCA* Variations Risk Assessment in Breast Cancers Using Different Artificial Intelligence Models", Genes, Vol. 12(11) 1774. (2021) (**Niyazi Şentürk**, Gulten Tuncel, Berkcan Dogan, Lamiya Aliyeva, Mehmet S. Dündar, Sebnem Sag, Gamze Mocan, Sehime G. Temel, Munis Dundar, Mahmut Çerkez Ergören)
- "Online medicine tracking system of pharmacy's in North Cyprus", Cyprus journal of Medical Sciences, Vol. 6(3), pp. 244-248. (2021) (**Niyazi Şentürk**)
- "Developing an online portal for determining the genomic signature of Archaic DNA that are associated to modern human genetic diseases: A meta-

analysis study", The Eurasian Journal of Medicine, Vol. 52(2), pp. 153-160. (2020) ( **Niyazi Şentürk**, Mahmut Çerkez Ergören)

- "Developing an in silico analysis portal for unrevealing genomic signature of Archaic DNA that are related to modern human genetic diseases", Journal of Biotechnology, Vol. 256, pp.53-54. (2017) (**Niyazi Senturk**, Mahmut Cerkez Ergoren)
- "Evaluating Cancer Treatment Alternatives using Fuzzy PROMETHEE Method", International Journal of Advanced Computer Science and Applications, Vol. 8, pp. 177-182 (2017) (Dilber Uzun Ozsahin, Berna Uzun, Musa Sani Musa, Abdulkader Helwan, Chidi Nwekwo Wilson, Fatih Veysel Nurçin, **Niyazi Şentürk**, Ilker Ozsahin)
- "Evaluating nuclear medicine imaging devices using fuzzy PROMETHEE method", Procedia Computer Science, Vol. 120, pp. 699-705 (2017) (Dilber Uzun Ozsahin, Berna Uzun, Musa Sani Musa, **Niyazi Şentürk**, Fatih Veysel Nurçin, Ilker Ozsahin)

**BULLENTING PRESENTED IN INTERNATIONAL ACADEMIC MEETINGS AND PUBLISHED IN PROCEEDINGS BOOK**

- **Niyazi Senturk** and Mahmut Cerkez Ergoren. How Neanderthal are you? In silico model-based analysis of Introgressed Neanderthal Ancestry in modern humans. Second International Biomedical Engineering Congress (IBMEC 2018) 24th-27th May 2018 Nicosia, Cyprus
- **Senturk N** and Ergoren MC. Developing an in silico analysis portal for unrevealing genomic signature of Archaic DNA that are related to modern human genetic diseases Journal of Biotechnology Volume 256, Supplement, Pages S1-S116 (30 August 2017)