



**SAĞKALIM ANALİZİNDE COX REGRESYON VE YAPAY SİNİR AĞLARI  
KULLANIMI**

**ELIAS ABDULLAH AL-SAMAI**

**Doktora Tezi**

**İstatistik Anabilim Dalı**

**Uygulamalı İstatistik Bilim Dalı**

**Danışman: Prof. Dr. Sevil ŞENTÜRK**

**Eskişehir**

**Anadolu Üniversitesi**

**Fen Bilimleri Enstitüsü**

**Ocak 2024**

*Bu tez çalışması BAP Komisyonu tarafından kabul edilen ..... no.lu  
proje kapsamında desteklenmiştir.*

## JÜRİ VE ENSTİTÜ ONAYI

ELIAS ABDULLAH AL-SAMAI'nin SAĞKALIM ANALİZİNDE COX REGRESYON VE YAPAY SİNİR AĞLARI KULLANIMI başlıklı çalışması 08/01/2024 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Eskişehir Teknik Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği"nin ilgili maddeleri uyarınca, İstatistik Anabilim dalında Doktora Tezi olarak kabul edilmiştir.

### Jüri Üyeleri

### Unvan Adı Soyadı

### İmza

|     |                              |  |
|-----|------------------------------|--|
| Üye | : Prof.Dr. Sevil ŞENTÜRK     |  |
| Üye | : Prof.Dr. Zeki YILDIZ       |  |
| Üye | : Prof. Dr. Nihal EGE        |  |
| Üye | : Doç.Dr. Şükrü ACITAŞ       |  |
| Üye | : Dr.Öğr. Üyesi Özer ÖZAYDIN |  |

Prof. Dr. Semra KURAMA

Lisansüstü Eğitim Enstitüsü Müdürü

## DANIřMAN ONAYI

Daniřmanlıęını yrttęm Doktora ęrencisi ELIAS ABDULLAH AL-SAMAI, SAęKALIM ANALİZİNDE COX REGRESYON VE YAPAY SİNİR AęLARI KULLANIMI bařlıklı tez alıřmasını tamamlamıřtır. Hazırlamıř olduęu tez tarafımda incelenmiř ve ęrencinin tez savunma sınavına alınması bilimsel ve etik aıdan uygun grlmřtr.

Tez Daniřmanı

Prof. Dr. Sevil řENTRK

**ÖZET**  
**SAĞKALIM ANALİZİNDE COX REGRESYON VE YAPAY SİNİR AĞLARI**  
**KULLANIMI**

ELIAS ABDULLAH AL-SAMAI

İstatistik Anabilim Dalı

Uygulamalı İstatistik Bilim Dalı

Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, Ocak 2024

Danışman: Prof. Dr. Sevil ŞENTÜRK

Sağkalım analizinde en yaygın kullanılan regresyon modeli Cox Regresyon Modelidir. Bu çalışmada, sağkalım analizinde cox regresyon modeli ve yapay sinir ağları modelleri kullanılarak Lösemi hastalarına ait veriler analiz edilmiştir. Çalışmada Yemen Cumhuriyeti'ndeki 2017 yılı Ocak ayı ve 2022 yılı Şubat ayı arasında 1168 Lösemi hastasına ait bilgiler kullanılmıştır. Bu tezin amacı, en yaygın hastalıklardan biri olan kanser hastalığı ile başa çıkmada Yemen Cumhuriyeti'ndeki lösemi hastalarının sağkalım sürelerini etkileyen en önemli faktörleri (risk faktörleri) belirlemektir. Ayrıca, sağkalım analizinde Yapay Sinir Ağlarını kullanmak ve Yapay Sinir Ağları modeli ile Cox Regresyon Modeli arasındaki performans göstergelerini karşılaştırmaktır. Söz konusu amaçla Cox Regresyon Modeli ve Yapay Sinir Ağları modelleri sağkalım verilerinin analizinde kullanılmıştır. Önerilen Cox modellerinden en iyi model seçilmiş ve ardından aynı veriler yapay sinir ağları kullanılarak analiz edilip modellerden en iyi olanı da seçilmiştir. Cox regresyon modeli sonuçları ile yapay sinir ağlarının kullanımından elde edilen sonuçlar arasında karşılaştırma yapılmıştır. İki yöntem arasındaki karşılaştırma, Hata Kareler Ortalaması (MSE) ve Ortalama mutlak hata (MAE) kriterlerine göre gerçekleştirilmiştir. Cox regresyon modellerinden en iyi model ve yapay sinir ağı modellerinden en iyi model seçilip karşılaştırdıktan sonra yapay sinir ağları modelinin, daha önce belirtilen kriterlere göre sağkalım verilerini analiz etmekte Cox regresyon modelinden daha iyi performans gösterdiği sonucuna varılmıştır.

**Anahtar Sözcükler:** Sağkalım analizi, Cox regresyon, Yapay sinir ağları.

## ABSTRACT

### THE USE OF COX REGRESSION AND ARTIFICIAL NEURAL NETWORKS IN THE SURVIVAL ANALYSIS.

ELIAS ABDULLAH AL-SAMAI

Department of Statistics

Programme in Applied Statistics

Anadolu University, Institute of Sciences, January 2024

Supervisor: Prof. Dr. Sevil ŞENTÜRK

The Cox Regression Model is the most commonly used regression model in survival analysis. In this study, Cox regression and artificial neural networks were introduced for survival analysis, and these models were exemplified using data from 1168 leukemia patients in the Republic of Yemen between January 2017 and February 2022. The aim of this study is to identify the most significant factors (risk factors) influencing the survival times of leukemia patients in the Republic of Yemen, dealing with one of the most common diseases (cancer). Additionally, the study compares the performance indicators between the Artificial Neural Networks model and the Cox Regression model in survival analysis. To achieve this analysis, the Cox regression model was used as a classic method in analyzing survival data in addition to artificial neural network as a modern method. The best model was chosen from the proposed Cox models and then the same data was analyzed using artificial neural networks. A comparison was made between the results of the Cox regression model and the results obtained from the use of artificial neural networks. The comparison between the two methods was carried out according to the criteria of the least mean squared error (MSE) and the mean absolute error (MAE). After selecting the best model from Cox regression models and the best model from the artificial neural network models and comparing them, the artificial neural networks outperformed the Cox regression model in analyzing survival data according to the previously mentioned criteria.

**Keywords:** Survival analysis, Cox regression, Artificial neural networks.

## TEŞEKKÜR

Doktora sürecinin başından itibaren her türlü desteği ve imkanı sağlayan, manevi desteğini her zaman en içten hissettiğim kıymetli danışman hocam Sayın Prof.Dr.Sevil Şentürk'e; tezimin şekillenmesinde ve sonuçlanmasında katkı ve önerileri ile yoluma ışık tutan Tez İzleme Komitesi üyeleri kıymetli hocalarım Sayın Doç.Dr. Şükrü ACITAŞ ve Sayın Dr.Öğr. Üyesi Özer Özaydın'a emekleri için sonsuz teşekkür ve saygılarımı sunarım.

Doktora savunma jürisinde yer alan hocalarım Prof.Dr. Zeki YILDIZ ve Prof. Dr. Nihal EGE'ye çalışmaya yapmış oldukları değerli katkıları ve yapıcı eleştirilerinden dolayı teşekkürlerimi sunuyorum.

İçinde bulunduğum süreçte bana rahat bir çalışma ortamı sağlayan, benden desteğini esirgemeyen ve pes etmeme hiç izin vermeyen sevgili eşim Basma AL-HAKIMI'ye ve varlıklarıyla bana güç veren oğullarım Ayham Elias AL-SAMAI ve Luai Elias AL-SAMAI'ye en içten teşekkürlerimi sunarım.

Beni sabırla bekleyen, bu günlere gelmemin sebebi beni yetiştiren ailem babam, annem ve kardeşlerim... Kelimelerle anlatılamaz, emekleri asla ödenmez... Size sonsuz teşekkür ederim.

ELIAS ABDULLAH AL-SAMAI

## **ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ**

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Eskişehir Teknik Üniversitesi tarafından kullanılan “bilimsel intihal tespit programı”yla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

ELIAS ABDULLAH AL-SAMAI



## İÇİNDEKİLER

### Sayfa

|                                                             |      |
|-------------------------------------------------------------|------|
| BAŞLIK SAYFASI .....                                        | I    |
| JÜRİ VE ENSTİTÜ ONAYI.....                                  | II   |
| DANIŞMAN ONAYI .....                                        | III  |
| ÖZET .....                                                  | IV   |
| ABSTRACT.....                                               | V    |
| TEŞEKKÜR .....                                              | VI   |
| ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....            | VII  |
| İÇİNDEKİLER .....                                           | VIII |
| TABLolar DİZİNİ .....                                       | XI   |
| ŞEKİLLER DİZİNİ .....                                       | XII  |
| SİMGELER VE KISALTMALAR DİZİNİ .....                        | XIII |
| 1. GİRİŞ .....                                              | 1    |
| 1.1. Literatür Özeti.....                                   | 4    |
| 2. SAĞKALIM ANALİZİ .....                                   | 7    |
| 2.1. Sağkalım Analizi Modelleri.....                        | 7    |
| 2.2. Sağkalım Analizinde Kullanılan Temel Fonksiyonlar..... | 8    |
| 2.2.1. Sağkalım fonksiyonu .....                            | 8    |
| 2.2.2. Tehlike fonksiyonu .....                             | 9    |
| 2.2.3. Başarısızlığın olasılık yoğunluk fonksiyonu .....    | 11   |
| 2.3. Yaşam Süresi Verileri Ve Sansürleme. ....              | 12   |
| 2.3.1. Birinci tür sansürleme.....                          | 12   |
| 2.3.2. İkinci tür sansürleme .....                          | 13   |
| 3. COX REGRESYON MODELİ.....                                | 14   |
| 3.1. Cox Modelinin Varsayımları.....                        | 15   |
| 3.2. Cox Regresyon Dönüşümü .....                           | 15   |
| 3.3. Model Parametrelerinin Tahmini.....                    | 16   |

|                                                                        |    |
|------------------------------------------------------------------------|----|
| 3.4. Partial Likelihood Yöntemi.....                                   | 16 |
| 3.5. Cox Regresyon Modelinin Uyumunun Değerlendirmesi.....             | 17 |
| 3.5.1. Olabilirlik Oran Testi (Likelihood Ratio Test).....             | 17 |
| 3.5.2. Wald Testi (Wald Statistics).....                               | 18 |
| 3.5.3. Akaike ölçütü (Akaike information criterion-AIC) .....          | 18 |
| 4. YAPAY SİNİR AĞLARI .....                                            | 19 |
| 4.1. Yapay Sinir Ağının Bileşenleri.....                               | 22 |
| 4.1.1. Giriş katmanı (Input Layer).....                                | 22 |
| 4.1.2. Çıktı katmanı (Output Layer).....                               | 22 |
| 4.1.3. Gizli katman (Hidden Layer) .....                               | 23 |
| 4.1.4. Bağlantılar veya ağırlıklar (Connections).....                  | 23 |
| 4.2. Yapay Sinir Ağlarında İşlem Birimleri (Processing Elements) ..... | 23 |
| 4.2.1. Ağırlık katsayıları (Weight coefficients) .....                 | 24 |
| 4.2.2. Toplama fonksiyonu (summation function).....                    | 24 |
| 4.2.3. Transfer fonksiyonu (Transfer function) .....                   | 24 |
| 4.2.3.1. Sigmoid aktivasyon fonksiyonu (Sigmoid function).....         | 25 |
| 4.2.3.2. Adım fonksiyonu (Step function) .....                         | 26 |
| 4.2.3.3. Doğrusal fonksiyon (linear function) .....                    | 26 |
| 4.2.3.4. Sinyal fonksiyonu (Sign function) .....                       | 27 |
| 4.2.4. Çıkış fonksiyonu .....                                          | 27 |
| 4.3. Yapay Sinir Ağlarının Sınıflandırması .....                       | 28 |
| 4.3.1. Katman sayısına göre sınıflandırma .....                        | 28 |
| 4.3.1.1. Tek katmanlı Ağlar .....                                      | 28 |
| 4.3.1.2. Çok katmanlı ağlar .....                                      | 29 |
| 4.3.2. Öğrenme ve eğitim yöntemleri göre sınıflandırma .....           | 30 |
| 4.3.2.1. Gözetimli öğrenme (Supervised learning ) .....                | 30 |
| 4.3.2.2. Gözetimsiz öğrenme (Unsupervised learning ) .....             | 30 |
| 4.3.2.3. Hibrit öğrenme (Hybrid learning ).....                        | 30 |
| 4.3.3. Veri akışının doğasına göre sınıflandırma .....                 | 31 |
| 4.3.3.1. İleri beslemeli ( Feed forward neural networks) .....         | 31 |
| 4.3.3.2. Geri beslemeli (Feedback neural networks) .....               | 31 |
| 4.4. Sinir Ağlarında Öğrenme Algoritmaları .....                       | 31 |

|                                                                     |           |
|---------------------------------------------------------------------|-----------|
| 4.4.1. Geri yayılım algoritması (Back propagation algorithm) .....  | 31        |
| <b>5. GERÇEK VERİ UYGULAMASI VE BULGULAR .....</b>                  | <b>36</b> |
| 5.1. Çalışma Verileri .....                                         | 36        |
| 5.1.1. Bağımlı Değişkenler .....                                    | 36        |
| 5.1.2. Bağımsız Değişkenler (Risk Faktörleri) .....                 | 36        |
| 5.2. Tanımlayıcı İstatistikler.....                                 | 37        |
| 5.3. Göreceli Risk Varsayım Testi .....                             | 44        |
| 5.3.1. Grafik Yöntemi.....                                          | 45        |
| 5.3.2. Schoenfeld Artıklar (residuals) Testi.....                   | 45        |
| 5.4. Cox Modelinin Temel Fonksiyonlarının Tahmin Edilmesi.....      | 46        |
| 5.5. Cox Regresyon Modelinin Parametrelerinin Tahmin Edilmesi ..... | 49        |
| 5.6. Modele Dahil Edilen Değişkenlerin Anlamlılığını Testi .....    | 50        |
| 5.7. En İyi Modelin Belirlenmesi.....                               | 53        |
| 5.8. Sağkalıma Süreleri Analizi.....                                | 54        |
| 5.9. Yapay Sinir Ağlarını Kullanarak Veri Analizi .....             | 57        |
| 5.9.1. Eğitim (Training).....                                       | 65        |
| 5.9.2. Modeller arasında karşılaştırma.....                         | 65        |
| 5.9.3. Önerilen yapay sinir ağı modelini Oluşturmak .....           | 66        |
| 5.10. Cox Modeli İle YSA'ın Sonuçlarının Karşılaştırılması. ....    | 67        |
| <b>6. SONUÇ VE TARTIŞMA .....</b>                                   | <b>68</b> |
| <b>KAYNAKÇA.....</b>                                                | <b>71</b> |
| <b>ÖZGEÇMİŞ</b>                                                     |           |

## TABLÖLAR DİZİNİ

### Sayfa

|                    |                                                                        |    |
|--------------------|------------------------------------------------------------------------|----|
| <b>Tablo 5.1.</b>  | Cinsiyet ve dönem sonundaki durum çapraz tablosu.....                  | 37 |
| <b>Tablo 5.2.</b>  | Yaş ve dönem sonundaki durum çapraz tablosu.....                       | 38 |
| <b>Tablo 5.3.</b>  | Kan grubu ve dönem sonundaki durum çapraz tablosu.....                 | 39 |
| <b>Tablo 5.4.</b>  | Medeni durumu ve dönem sonundaki durum çapraz tablosu .....            | 40 |
| <b>Tablo 5.5.</b>  | İkamet ettiği şehir ve dönem sonundaki durum çapraz tablosu.....       | 41 |
| <b>Tablo 5.6.</b>  | Hastalığın türü ve dönem sonundaki durum çapraz tablosu .....          | 42 |
| <b>Tablo 5.7.</b>  | Kullanılan tedavi metodu ve dönem sonundaki durum çapraz tablosu ..... | 43 |
| <b>Tablo 5.8.</b>  | Tedavi düzenlilik ve dönem sonundaki durum çapraz tablosu.....         | 44 |
| <b>Tablo 5.9.</b>  | Schoenfeld artıkları göreceli risk testinin sonuçları .....            | 46 |
| <b>Tablo 5.10.</b> | Cox regresyon modelinin parametre tahminleri .....                     | 49 |
| <b>Tablo 5.11.</b> | Denklemdaki değişkenler .....                                          | 50 |
| <b>Tablo 5.12.</b> | Referans hücre kullanılarak hastalık türü için hazard oranı testi..... | 52 |
| <b>Tablo 5.13.</b> | En iyi modeli seçmek için olabilirlik oran testi sonuçları.....        | 53 |
| <b>Tablo 5.14.</b> | Aic ve olabilirlik oran testi .....                                    | 54 |
| <b>Tablo 5.15.</b> | Lösemili hastalar için sağkalım verilerinin yaşam tablosu .....        | 54 |
| <b>Tablo 5.16.</b> | Eğitim ve test verilerinin bölünmüş oranları .....                     | 58 |
| <b>Tablo 5.17.</b> | Modeller arasında karşılaştırma sonuçları .....                        | 59 |
| <b>Tablo 5.18.</b> | Ağ bilgileri.....                                                      | 60 |
| <b>Tablo 5.19.</b> | Model özeti .....                                                      | 62 |
| <b>Tablo 5.20.</b> | Sınıflandırma tablosu.....                                             | 62 |
| <b>Tablo 5.21.</b> | Eğri altında kalan alanı .....                                         | 63 |
| <b>Tablo 5.22.</b> | Bağımsız değişkenin önemi .....                                        | 64 |
| <b>Tablo 5.23.</b> | Modeller arasında karşılaştırma sonuçları .....                        | 65 |
| <b>Tablo 5.24.</b> | Cox modeli ile ysa'ın sonuçlarının karşılaştırılması.....              | 67 |

## ŞEKİLLER DİZİNİ

### Sayfa

|                                                                               |    |
|-------------------------------------------------------------------------------|----|
| Şekil 2.1. Tehlike fonksiyonu.....                                            | 10 |
| Şekil 2.2. Farklı tehlike fonksiyonun biçimleri .....                         | 10 |
| Şekil 4.1. Biyolojik nöronun genel yapısı .....                               | 20 |
| Şekil 4.2. Yapay nöronun genel yapısı .....                                   | 23 |
| Şekil 4.3. İkili sigmoid aktivasyon fonksiyonu .....                          | 25 |
| Şekil 4.4. Kutupsal ikili sigmoid aktivasyon fonksiyonu .....                 | 26 |
| Şekil 4.5. Adım fonksiyonu.....                                               | 26 |
| Şekil 4.6. Doğrusal fonksiyonu .....                                          | 27 |
| Şekil 4.7. Sinyal fonksiyonu.....                                             | 27 |
| Şekil 4.8. Tek katmanlı ağ.....                                               | 29 |
| Şekil 4.9. Çok katmanlı ağ.....                                               | 29 |
| Şekil 4.10. İleri yayılım ağı (forward propagation).....                      | 33 |
| Şekil 4.11. Geri yayılım ağı (back propagation) .....                         | 35 |
| Şekil 5.1. Cinsiyet ve dönem sonundaki durum çapraz tablosu .....             | 37 |
| Şekil 5.2. Yaş ve dönem sonundaki durum çapraz tablosu .....                  | 38 |
| Şekil 5.3. Kan grubu ve dönem sonundaki durum çapraz tablosu .....            | 39 |
| Şekil 5.4. Medeni durumu ve dönem sonundaki durum çapraz tablosu.....         | 40 |
| Şekil 5.5. İkamet ettiği şehir ve dönem sonundaki durum çapraz tablosu .....  | 41 |
| Şekil 5.6. Hastalığın türü ve dönem sonundaki durum çapraz tablosu.....       | 42 |
| Şekil 5.7. Kullanılan tedavi metodu ve dönem sonundaki durum .....            | 43 |
| Şekil 5.8. Tedavi düzenliliği.....                                            | 44 |
| Şekil 5.9. Grafik kullanarak göreceli risk varsayımı testi .....              | 45 |
| Şekil 5.10. Lösemili hastalara temel hazard fonksiyonu .....                  | 47 |
| Şekil 5.11. Cum hazard fonksiyonu.....                                        | 48 |
| Şekil 5.12. Sağkalım fonksiyonu (survival)eğrisi .....                        | 48 |
| Şekil 5.13. Referans hücre kullanılarak hastalık türü için hazard oranı ..... | 53 |
| Şekil 5.14. Yaşam tabloları yöntemine göre sağkalım fonksiyonu eğrisi .....   | 56 |
| Şekil 5.15. Yaşam tabloları yöntemine göre hazard fonksiyonu eğrisi .....     | 56 |
| Şekil 5.16. Birinci aşamadaki seçilen sinir ağın mimarisi.....                | 61 |
| Şekil 5.17. Roc eğrisi.....                                                   | 63 |
| Şekil 5.18. Bağımsız değişkenin önemi.....                                    | 64 |

## SİMGELER VE KISALTMALAR DİZİNİ

|          |                                                           |
|----------|-----------------------------------------------------------|
| $\alpha$ | : Alpha                                                   |
| $\beta$  | : Beta                                                    |
| MAE      | : Ortalama Mutlak Hata.                                   |
| MSE      | : Hata Kareler Ortalaması                                 |
| YSA      | : Yapay Sinir Ağları                                      |
| ANN      | : Artificial Neural Network                               |
| $S(t)$   | : Sağkalım Fonksiyonunu                                   |
| $h(t)$   | : Tehlike Fonksiyonu (Hazard Function)                    |
| Adam     | : Adaptive Moment Estimation) Algoritması                 |
| SGD      | : Stochastic Gradient Descent Algoritması                 |
| ALL      | : Akut Lenfositik Lösemi (Acute Lymphoblastic Leukemia)   |
| AM       | : Akut Miyeloid Lösemi (Acute Myeloid Leukemia)           |
| CML      | : Kronik Miyeloid Lösemi (Chronic Myeloid Leukemia)       |
| CLL      | : Kronik Lenfositik Lösemi (Chronic Lymphocytic Leukemia) |

## 1. GİRİŞ

Sağkalım analizi, ilgilenilen değişkenin olayın meydana gelmesine kadar sağkalım süresi olan verileri analiz etmek için bir takım istatistiksel teknik içeren, bir istatistik dalı olarak tanımlanır [1][2]. Başka bir deyişle, sağkalım analizi, orijinal zamandan belirli bir olayın veya son noktaya kadar geçen zaman biçimini alan verilerin analizini tanımlamak için kullanılan bir terimdir.

Yani, sağkalım analizi, olaya kadar geçen sürenin verileri analiz etmek için bir takım istatistiksel prosedürdür ve sürenin başlangıcından bu olayın meydana gelmesine veya çalışmasından çekilmesine kadar geçen süredir. Bu süre, günler, haftalar, aylar hatta yıllar olabilir. Olay ise, ölüm, bir hastalığın oluşumu, evlilik, boşanma, tedaviye yanıt vb olabilmektedir [3][4].

Sağkalım analizinin farklı terimleri bulunmaktadır. Tıbbi alanlarda ve sağlık bilimlerinde kritik olay ölüm olduğu için sağkalım fonksiyonlarının analizi olarak adlandırılır. Mühendislik çalışmalarında güvenilirlik terimi, sosyal bilimde ise olayın tarihi ile ilgili çalışmaları ifade etmektedir.

Daha açık bir şekilde, sağkalım analizi, zamanın modellenmesini, hastalığın teşhis edildiği andan olayın meydana geldiği ana kadar hastanın durumunu incelemeyi içerir. Olay, tıbbi araştırmalarda sağkalım analizi literatüründe ölümü temsil eder, sansür ise, iyileşme, sağlık durumunu bilmeden hastaneden çekilme, veya çalışmanın nedeni dışında ölümü içerir.

Geçen yüzyılın yetmişli yıllarında, bağımlı değişkenin sansüre tabi olduğu regresyon problemiyle başa çıkma yöntemleri çeşitlilik göstermektedir. Yöntemlerden bazıları, sağkalım sürelerinin dağılımına ilişkin varsayımlara dayanır. Diğerleri sağkalım sürelerinin dağılımı hakkındaki varsayımlara dayanmaz, sağkalım sürelerinin dağılımı ile ilgili varsayımların olmadığı bu yöntemlerden biri, göreceli risk (Proportional Hazards) modeline dayanan Cox Regresyon (Cox Regression) modelidir. 1972 yılında İngiliz bilim adamı (David Cox), açıklayıcı değişkenleri ile onlar için sağkalım süreleri arasındaki ilişkiyi belirleyen göreceli risk modelini veya sözde Cox Regresyon modelini tahmin etmiştir. Cox modeli, kritik olayın ortaya çıkmasına kadar geçen süre boyunca hazard fonksiyonuna önemli katkı sağlayan risk faktörlerini bilmeyi amaçladığından, sağkalım verilerinin analizinde, özellikle sansür verileri (Censored Survival Time) söz konusu

olduğunda en çok kullanılan modeldir. David Cox 1975 yılında, 1972'de önerdiği modelin parametrelerini tahmin etmek için bir yöntem önermiştir ve buna kısmi olasılık fonksiyonu adını vermiştir [5][6].

Yapay Sinir Ağları (YSA) ise, biyolojik sinir ağlarını insan düşünme biçimine ve beynin bilgiyi nasıl işlediğine dayalı matematiksel modeller olarak simüle eden bir bilgi işleme sistemi olarak sayılan elektronik bilgi işleme modelleridir. YSA belirli problemleri çözmek için uyum içinde çalışan, ayrıca bilgiyi işlemek için biyolojik sinir sisteminden ilham alan bir tarzda insanlar gibi yapay sinir ağlarını öğrenen birçok işlem elemanından oluşan modellerdir.

YSA, insan beyninden esinlenilerek geliştirilmiş, ağırlıklı bağıntılar aracılığıyla birbirine bağlanan ve her biri kendi bağına sahip işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır. Bir başka deyişle, biyolojik sinir ağlarını taklit eden bilgisayar programlarıdır. Demek oluyor ki yapay sinir ağları zaman zaman bağlantıcılık, paralel dağıtılmış işlem, sinirsel-işlem, doğal zekâ sistemleri ve makine öğrenme algoritmaları gibi isimlerle de anılmaktadır. YSA bir programcının geleneksel yeteneklerini gerektirmeyen, kendi kendine öğrenme düzenekleridir. Bu ağlar öğrenmenin yanı sıra, ezberleme ve bilgiler arasında ilişkiler oluşturma yeteneğine de sahiptir[7].

Yapay sinir ağları günümüzde birçok probleme çözüm üretebilecek yeteneğe sahiptirler. YSA tanımlarının birkaç ortak noktası bulunmaktadır. Bunların en başında yapay sinir ağlarının birbirine hiyerarşik olarak bağlı ve paralel olarak çalışabilen yapay hücrelerden oluşmaları gelmektedir. Proses elemanları da denilen bu hücrelerin birbirine bağlandıkları ve her bağlantının bir değerinin olduğu kabul edilmektedir.[8]

YSA'lar, birbirine bağlı bir yapay nöron grubu olarak birkaç katmandan oluşur. Ayrıca her katman, nöronların birbirleri üzerindeki etki miktarını gösteren bir ağırlığa sahiptir. Genellikle, bir YSA modelinde girdi, gizli ve çıktı katmanları adı verilen üç katman bulunur.

Bu çalışmanın önemi, sağkalım analizi çalışmalarında Cox regresyon modelini kullanmak ve bunları modern yöntemlerle, özellikle de bu çalışmayı diğer çalışmalar için bir başlangıç noktası haline getiren yapay sinir ağı modelleriyle karşılaştırmaktır. Ayrıca, uygulama kısmında ise en yaygın hastalıklardan biri olan kanserle başa çıkmada Yemen



Cumhuriyeti'ndeki lösemi hastalarının sağkalım sürelerini etkileyen en önemli faktörler (risk faktörleri) belirlenmeye çalışılmıştır.

Bu tez çalışmasının amacı ise, Yapay Sinir Ağları modelini sağkalım analizinde kullanmak ve Yapay Sinir Ağları modeli ile Cox regresyon modeli arasındaki performans göstergelerini karşılaştırmaktır. Bu amaçla, hem sağkalım analiz, cox regresyon ve hem de yapay sinir ağları ile ilgili teorik yapı tezde verilmiştir.

Tez çalışması beş bölümden oluşmaktadır ve birinci bölüm olan giriş bölümünde literatür taramasına yer verilmiştir. Çalışmanın ikinci bölümünde sağkalım analizi ve üçüncü bölümde ise Cox regresyon ile ilgili yapıya yer verilmiştir. Dördüncü bölümde ise yapay sinir ağlarının tanımını, tarihçesi, türleri, bileşenleri, ek olarak yapay sinir ağındaki işlem öğeleri, sinir ağlarının teorik yönü ele alınmıştır. YSA, veri akışı yöntemlerine, içindeki katman sayılarına ve öğrenme yöntemlerine göre sınıflandırılmıştır. Son olarak, sinir ağlarının performansını değerlendirmede kullanılan yöntemlere yer verilmiştir.

Çalışmanın ana kısmını oluşturan beşinci bölüm olan uygulama bölümünde, çalışma verilerine tanımlayıcı analiz uygulanmıştır. Ardından, hastaların sağkalım süresini etkileyen en önemli faktörleri belirlenmiş ve bu değişkenlerin anlamlılığını ölçmek için Cox regresyon modeli kullanılmıştır. Lösemi hastaları için sağkalım süresini anlamlı ölçüde etkileyen değişkenlerin belirlenmesi, sağkalım verileri için yaşam tablolarının hazırlanması ve çalışma verileri için yapay sinir ağları kullanılarak da analizler gerçekleştirilmiştir.

Sağkalım verilerinin analizinde YSA kullanmak için, veri işlemeden başlayarak ve en iyi yapay sinir ağı modelini elde etmek için verilerin analiz edildiği son ağ modelinin formülasyonu ile biten birkaç sıralı adım izlemiştir. Python ve SPSS paket programlarıyla veriler yapay sinir ağlarını kullanarak iki aşamada analiz edilmiştir. Birinci aşamada, veriler Zaman ve Hasta durumu (Olay, Sansür) olmak üzere iki bağımlı değişkene dayalı olarak analiz edilirken, her iki aşamada da sekiz bağımsız değişkenle çalışmıştır.

İkinci aşamada, verilere yapay sinir ağları uygulanmış, ve bu durumda bağımlı değişken risk fonksiyonu olurken, bağımsız değişkenler ise ilk aşamadaki ile aynı sekiz değişken (cinsiyet, yaş, kan grubu, medeni durum, il, hastalığın türü, kullanılan tedavi metodu ve tedavi düzenlilik) olarak ele alınmıştır. Ağın girdileri birinci adımda olduğu gibi bu bölümde ele alınan Cox Regresyon modelinde olduğu gibi Hazard fonksiyonunu

etkileyen sekiz bağımsız değişken olacaktır. Gizli katmanlar eğitim sürecinden çıkan en iyi mimariye göre belirlenmiştir. Ağ çıktıları ise tahmin edilecek olan Hazard fonksiyonu değişkenidir. Bu bölümde son olarak Cox regresyon modeli ile YSA'nın sonuçları karşılaştırılmıştır. Veriler her iki yöntem kullanılarak analiz edildikten sonra, her iki yöntemden en iyi model seçilmiştir. Sonrasında ise en önemli adım sayılan MSE ve MAE karşılaştırma kriterlerine göre, en iyi yöntem seçilmiştir. Çalışmanın son kısmını oluşturan beşinci bölümde ise sonuç ve tartışmaya yer verilerek tez çalışması tamamlanmıştır.

### 1.1. Literatür Özeti

Literatürde yer alan Cox regresyon modeli ve YSA ile yapılan çalışmalara yer verilecektir.

Sabouri vd. (2020) çalışmalarında, kolon kanserli hastalarda sağkalım ile ilgili risk faktörlerini belirlemek için Cox Regresyon modeli ve yapay sinir ağları uygulamıştır. 2006-2011 yılları arasında toplam 157 kolon kanserli hasta incelenmiş ve 2016 yılına kadar takip edilmiştir. Hastaların tıbbi kayıtları kullanılarak hastaların demografik ve klinik bilgileri toplanmıştır. Karmaşıklıktan kaçınmak için sadece bir gizli katman kullanılmış giriş katmanında 11 düğüm, gizli katmanda 5 ile 15 düğüm ve çıkış katmanında 1 düğüm ile YSA modeline uyacak 3 katmanlı bir MLP kullanılmıştır. Bağımlı değişken hasta durumu ikili değişken olarak tanımlanmıştır. Lojistik aktivasyon fonksiyonu ve önemli risk faktörlerini belirlemek için 0.1'den 0.5'e bozulmayla (decey) eğitim verileri için ileri besleme algoritması kullanılmıştır. Bu çalışmada YSA, Cox regresyondan daha üstün olan CRC hastalarının sağkalımını tahmin etmek için uygun bir yaklaşım olarak bulunmuştur.[9]

A.Biglarian vd. (2011) çalışmalarında, Cox regresyon ve yapay sinir ağı modellerini kullanarak İranlı mide kanseri hastalarının sağkalım oranlarını tahmin etmenin yanı sıra, bu yaklaşımların hastaların sağkalımını tahmin etmedeki yeteneklerini karşılaştırmışlardır. Sağkalım süresini tahmin etmek için, 2002 ve 2007 yılları arasında İran'daki Taleghani Hastanesinde ameliyat yaptıran 436 kayıtlı mide kanseri hastasından elde edilmiştir. Cox regresyon modelini kullanmanın yanı sıra, YSA modelinde, giriş katmanında 10 düğüm, gizli katmanda 7 düğüm ve çıkış katmanında bir düğüm bulunan üç katmanlı bir geri yayımlı sinir ağı kullanılmıştır. Her hastanın durumu bir ikili yanıt

değişkeni (ölü veya sansürlü) olduğundan, gizli ve çıkış katmanlarında aktivasyon fonksiyonu olarak sigmoid fonksiyon kullanılmıştır. Ağ eğitimi için geri yayılım öğrenme algoritması ve veri analizi için SPSS ve python yazılımları kullanılmıştır. Sonuç olarak, Yapay sinir ağının doğru tahmin oranı %83.1 ve Cox regresyon modeli için doğru tahmin oranı %75 olarak elde edilmiştir.[10]

Kiknadze vd.(2020) çalışmalarında yapay sinir ağı kullanarak sınıflandırma testi doğruluğunu, hassasiyet ve özgüllük değerlerini ölçerek sunmakta olan Wisconsin Meme kanseri teşhisi veri setini kullanmışlardır. Bu veri setinde 9 öznitelik bulunmaktadır. Bu sebepten giriş katmanında 9 nöron ve tümörler iyi huylu veya kötü huylu olarak ayrıldığı için çıkış katmanında 2 nöron ve ara katmanda 10 nöron kullanılmıştır. Uygulamanın başlangıcında rastgele yanlılık (bias) ve ağırlık değerleri oluşturulmuş ve uygulama süresinde bu değerler güncellenerek son değerler bulunmuştur. Uygulamada çıkış nöronlarında hatanın geri yayılması aracılığıyla değerlendirmede hata olasılığı en aza indirilmiştir. Aktivasyon fonksiyonu olarak sigmoid fonksiyon ve optimizasyon algoritması ise ADAM kullanılmıştır.[11]

Amir Hossein Hashemian vd.(2013)' çalışmalarında, Cox regresyon modelini böbrek nakli yapılan hastalarda böbrek sağkalımını öngörmeye ROC eğrisi yoluyla YSA modeliyle karşılaştırmışlardır. Bu çalışmada kullanılan YSA konfigürasyonu, düğüm sayısının 128 ve değişkenlerin 7 olduğu 3 katmandır. Araştırma bulguları YSA modelinin Cox regresyon modeline göre daha iyi tahmin yeteneğine sahip olduğunu ortaya koymuştur. Ayrıca, bu çalışma YSA modellerinin kabul edilebilir özgüllük ve duyarlılığa sahip olduğu sonucuna varmıştır. Eğri altındaki ROC alanları, çok katmanlı yapay sinir ağı modeli ve Cox regresyonu için sırasıyla %81.3 ve %71 olarak bulunmuştur.[12]

Akbar Biglarian vd.(2010)'da sağkalım analizinde Cox regresyon modeli (yarı parametrik model) ve üç katmanlı YSA modeli kullanılmıştır. YSA ve Cox regresyon modellerinin tahmin doğruluğu sırasıyla %81.51 ve %72.60 olarak bulunmuştur. Ayrıca, YSA ve Cox regresyon modellerinin AUROC'u sırasıyla 0.826 ve 0.754 elde edilmiştir. Bu çalışmada YSA'nın AUROC ve tahmin doğruluğu açısından Cox regresyon modelinden daha iyi olduğu sonucuna varılmıştır. Bu nedenle, sağkalım olasılığının tahmini için YSA modeli önerilmektedir.[14]

Chen vd. (2020)'de YSA modeli ile Cox regresyon modeli karşılaştırılmıştır. Bu çalışmada, YSA modelinin kalça kırığı sonrası postoperatif mortaliteyi tahmin etmede

daha doğru olduđu sonucuna varılmıştır. Çalışmada belirlenen postoperatif mortalite ile ilişkili tahmin belirleyicileri, kalça kırığı cerrahisi adaylarını iyileşme süreci ve sağlık sonuçları açısından eğitmek için de kullanılmıştır. Bu çalışmada kullanılan YSA modeli, 10 giriş nöronu, giriş katmanında 1 önyargı nöronu, 5 gizli nöron, gizli katmanda 1 önyargı nöronu (bias neuron) ve 2 çıkış nöronu içeren 3 katmanlı çok katmanlı yapay sinir ağıdır. En iyi gizli nöron sayısı, 5-35 aralığı arasından deneme yanılma yoluyla seçilmiştir [15]. Ayrıca, bu çalışmada, öğrenme sürecini (eğitim algoritması) gerçekleştirmek için Yarı- Newton yöntemi kullanılmıştır.

Spelt vd. (2013)'te kolorektal kanser metastazları için karaciğer rezeksiyonundan sonra sağkalımı tahmin etmek için yapay bir sinir ağı (ANN) modeli oluşturmayı amaçlamıştır. Kolorektal kanser metastazları için karaciğer rezeksiyonu geçiren hastalar de ilk klinik verilerden üretilen 1000 çift Cox ve YSA modelinin karşılaştırılmasından sonra, uzun vadeli sağkalımı tahmin etmek için Harrell C-indeksi YSA modellerinde Cox modellerine kıyasla (0.66) daha yüksek (0.72 ) bulunmuştur.[16]

Ansari vd. (2013)'te çalışmanın amacı, pankreas duktal adenokarsinomlu (PDAC) hastalarda sağkalımı tahmin etmek için bir YSA modeli geliştirmektir. Çalışma, radikal cerrahi geçirmiş pankreas kanseri hastalarında sağkalımı öngörmede YSA ve Cox modellerinin kalibrasyonunu ve geçerliliğini karşılaştırılmıştır. Yazarlar, YSA modelinin cerrahi rezeksiyon sonrası sağkalımı öngörmede daha doğru olduđu sonucuna varmışlardır. Harrell'in YSA modeli için uyum indeksi 0.79 ve Cox regresyonu için 0.67'olarak bulunmuştur.[16]

## 2. SAĞKALIM ANALİZİ

Sağkalım Analizi, aynı zamanda yaşam analizi veya İngilizce adı ile survival analizi olarak da adlandırılmaktadır. Bu analiz, ilgilenilen değişkenin olayın meydana gelmesine kadar sağkalım süresi olan verileri analiz etmek için bir takım istatistiksel teknik içeren, bir istatistik dalı olarak tanımlanır. Çeşitli faktörlerin etkisi ile birlikte inceleyerek o olay veya durum hakkında çeşitli sonuçlara ve tahminlere ulaşmamızı sağlayan bir analizdir. [18]

Sağkalım analizinin farklı terimleri bulunmaktadır. Sağlık alanında ölüm, enfeksiyon kapma veya hastalığın nüksetmesi vb. Mühendislik alanında güvenilirlik terimi, sosyal alanında ise olayın tarihi ile ilgili ifade edilmektedir. Önemli olan belirli bir zaman içerisinde gerçekleşen herhangi bir olay veya durumun olması, bir anlamda olaya bağlı bir ömür olmasıdır. Bu süre günler, haftalar, aylar, yıllar şeklinde olabilir. Olay ise, ölüm, bir hastalığın oluşumu, evlilik, boşanma, tedaviye yanıt vb olabilmektedir.[19]

### 2.1. Sağkalım Analizi Modelleri

Sağkalım Analizi Modelleri, belirli bir olayın meydana gelmesinden önce gelen zamanı ele alır ve bu alanda en çok uygulanan örneklerden biri, bu tez çalışmasında da olduğu gibi ölümden önceki zamandır. Ancak sağkalım analizinin uygulaması daha önce de belirtildiği gibi birçok farklı alan ve disiplinde kullanıldığı için bundan çok daha geniş alanı kapsamaktadır. Bir de zaman faktörü, çalışmada fenomeni analiz etmede temel bir faktör olarak saymaktadır. Ayrıca bu yöntemin temel özelliği, nicel, tanımlayıcı veya karışık olması açısından bu değişkenlerin niteliğine bakılmaksızın olaydan önceki zaman ile bir veya daha fazla bağımsız değişken arasındaki ilişkinin incelenmesidir.

Sağkalım analizinde kullanılan en önemli yöntemlerden biri, çok çeşitli alanlara uygulanan Cox Regresyon Modelidir. [20]

Sağkalım Analizi terimi, üç ana özelliği içeren verilerin analizi ile ilgili en yaygın kullanılan ve bilinen ismidir:[21]

1. Bağımlı değişken, belirli bir olay meydana gelene kadar sağkalım süresidir.
2. Sansür verilerinin varlığı.
3. Sağkalım süresini etkileyen açıklayıcı değişkenlerin varlığı ve bunların tanımlaması.

## 2.2. Sağkalım Analizinde Kullanılan Temel Fonksiyonlar.

Sağkalım süresi verileri, başarısızlık, ölüm, tepki, nüksetme gibi belirli bir hastalığın gelişimi, şartlı tahliye veya boşanma gibi belirli bir olaya kadar geçen süreyi ölçer. Bu zamanlar rastgele değişikliklere tabidir ve herhangi bir rastgele değişken gibi bir dağılım oluşturmaktadır. Sağkalım sürelerinin dağılımı genellikle üç temel fonksiyonla tanımlanır veya karakterize edilir:

- (1) Sağkalım fonksiyonu.
- (2) Tehlike fonksiyonu.
- (3) Başarısızlığın olasılık yoğunluk fonksiyonu.

Bu üç fonksiyon matematiksel olarak eşdeğerdir – bunlardan biri verilirse, diğer ikisi türetilir.

Uygulamada, verilerin farklı yönlerini göstermek için üç fonksiyon kullanılabilir. Sağkalım veri analizinde temel bir sorun, örneklenen verilerden bu üç fonksiyondan bir veya daha fazlasını tahmin etmek ve popülasyondaki sağkalım modeli hakkında çıkarımlar yapmaktır.

Sürekli bir rastgele değişken ( $T$ ) her zaman pozitif olduğunu ve ilgilenilen olay meydana gelene kadar sağkalım gereken süreyi gösterdiğini varsayalım, yani,  $\{t \geq 0\}$  ve  $t$ , ( $T$ ) değişkeninin aldığı belirli bir değeri göstermektedir. Olasılık yoğunluk fonksiyonu  $f(t)$  ve kümülatif dağılım fonksiyonu  $F(t)$  vardır, böylece sağkalım sürelerinin dağılımı temel sağkalım fonksiyonları ile tanımlanabilir.[22]

### 2.2.1. Sağkalım fonksiyonu

Ayrıca, güvenilirlik fonksiyonu olarak bilinir ve  $S(t)$  olarak gösterilir. Sağkalım fonksiyonu,  $t$ 'den sonra sağkalım olasılığı olarak tanımlanır ve aşağıdaki gibi matematiksel olarak ifade edilebilir.

$$S(t) = P(T \geq t), t > 0. \quad (2.1)$$

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^{\infty} f(x)dx, t > 0. \quad (2.2)$$

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}(1 - S(t)) = -S'(t) = \frac{dS(t)}{dt} \quad (2.3)$$

Dolayısıyla :

$S(t)$  : Sağkalım fonksiyonunu temsil eder.

$F(t)$  : Kümülatif dağılım fonksiyonunu temsil eder.

Ayrıca, dikkate değer.

$S(t) = 1$  olduğunda  $t = 0$  ,  $S(t) = 0$  olduğunda  $t = \infty$

Demek ki, sağkalım olasılığı, zaman sıfır olduğunda 1'e eşittir ve sonsuzda sıfıra eşittir. [4]

### 2.2.2. Tehlike fonksiyonu

Tehlike fonksiyonu, başarısızlık durumlarını incelerken önemli bir ölçüdür ve anlık başarısızlık oranını vermektedir. Gözlemin dönemin başına kadar devam ettiği varsayılarak çok kısa bir süre boyunca başarısızlık olarak tanımlanmaktadır. Tehlike fonksiyonu  $h(t)$ ,  $t$  zamanına kadar yaşayan bir gözlemin,  $(t+\Delta t)$  zamanına kadar yaşamının sona ermesi riskidir ve gözlemin ilgilenilen özellik bakımından başarısızlık eğiliminin bir ölçüsüdür.  $h(t)$  başarısızlık hızı (failure rate), ani ölüm hızı (instantaneous death rate) ya da ölümlülük gücü (force of mortality) olarak da ifade edilir.[23]

Tehlike fonksiyonu şu şekilde ifade edilebilir:

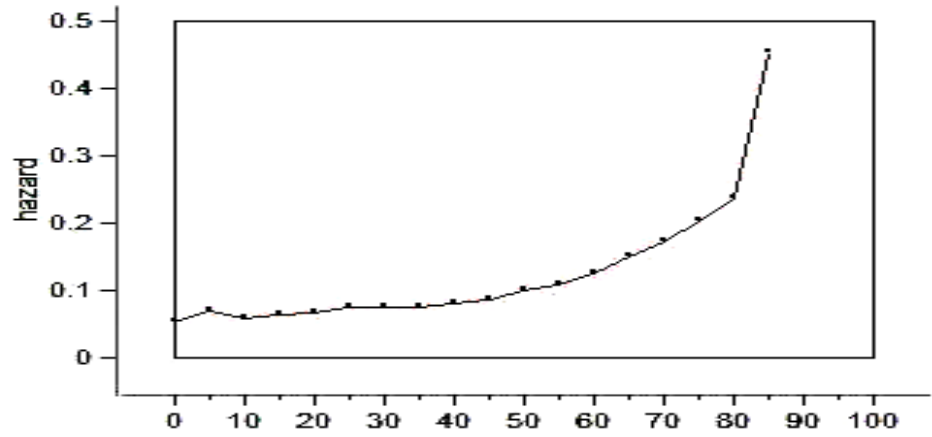
$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{p[t < T \leq t + \Delta t | T \geq t]}{\Delta t} \quad (2.4)$$

Tehlike Fonksiyonu sağkalım Fonksiyonu ve olasılık yoğunluğu Fonksiyonu arasındaki ilişki ise:

$$h(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \quad (2.5)$$

Tehlike Fonksiyonu aşağıdaki özelliklere sahiptir.[24]

1. Her zaman pozitif (negatif olmayan).
2. Üst sınırı yoktur, yani değeri  $(0, \infty)$  arasında kalır.:



Şekil 2.1. Tehlike fonksiyonu

Tehlike fonksiyonu türleri:

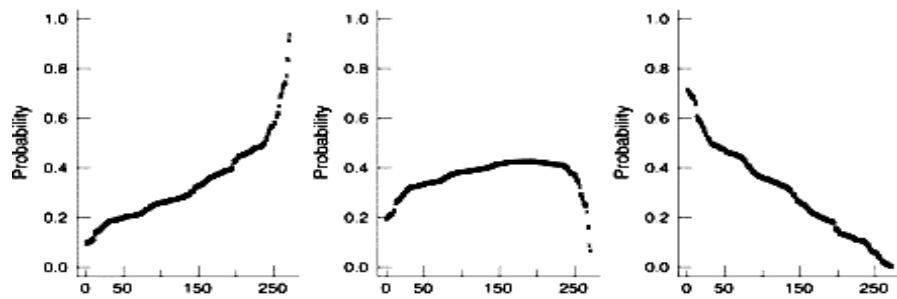
1- Artan tehlike fonksiyonu:

Zaman ilerledikçe ölüm olasılığının arttığı yerdur (doğrusu hastanın tedaviye yanıt vermemesidir).

2- Azalan tehlike fonksiyonu: Başlangıçta mortalitesi daha yüksek olan ve zaman ilerledikçe azalmaya başlayan (yani hasta tedaviye cevap verir ve iyileşir) hastaları temsil eder.

3- Sabit tehlike fonksiyonu: Hasta için ölüm olasılığı zaman içinde sabittir.

4- Tipik risk fonksiyonu: (u) harfinin şeklini alır. Önceki üç tipin birleşmesinin bir sonucudur ve bir kişinin doğumundan itibaren risk fonksiyonunu temsil etmektedir. Ölüm olasılığının doğumda büyük olduğu ve daha sonra gençlik döneminde kademeli olarak azalmaya ve stabilize olmaya başladığı ve daha sonra yaşlılıkta artışa dönmeye başladığı ve şekilde 2 gösterildiği gibidir.



Şekil 2.2. Farklı tehlike fonksiyonun biçimleri



Şekildeki gibi artan, sürekli ve azalan sırasıyla tehlike fonksiyonu anlamına gelir. Kümülatif tehlike fonksiyonu, tüm başarısızlık oranı değerlerinin zamandan (0'dan T'ye) entegrasyonu olan tehlike fonksiyonunun integrali alınarak elde edilebilir. Zaman geçtikçe tehlike birikimini ifade eder ve kümülatif risk fonksiyonu  $H(t)$  sembolü ile gösterilir ve matematiksel formülü aşağıdaki gibidir:[25]

$$H(t) = \int_0^t h(u) du \quad (2.6)$$

$$H(t) = -\ln S(t) \quad (2.7)$$

$t = 0$  eğer öyleyse, O zaman  $S(0) = 1$  ve  $H(0) = 0$ , Ayrıca  $t = \infty$  eğer o zaman

$S(\infty) = 0$ ,  $H(\infty) = \infty$ . ve Sıfır ile sonsuz arasında herhangi bir değeri alabilir.

Dolayısıyla;

$h(u)$ : Başarısızlık oranını temsil eder.

$H(t)$  Kümülatif tehlike fonksiyonu veya kümülatif başarısızlık oranı temsil eder.

### 2.2.3. Başarısızlığın olasılık yoğunluk fonksiyonu

Aynı hastalığa maruz kalan özdeş bireyler farklı zamanlarda ölür. Bazıları erken aşamalarda diğerleri ise uzun bir süre sonra ölür. Bu nedenle ölüm mekanizması rastgele olma eğilimindedir. Herhangi birim için başarısızlık olasılığı, arızanın olasılık yoğunluk fonksiyonu olarak adlandırılan ve TD sembolü ile gösterilen bir olasılık fonksiyonu olarak düşünülebilir ve bu fonksiyon, belirli bir kişinin rastgele davranışını temsil eder.

i. birey için yaşam süresi  $T$  rastgele değişken olmak üzere, olasılık yoğunluk fonksiyonu;

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t < T < t + \Delta t]}{\Delta t}; t \geq 0 \quad (2.8)$$

Biçimimde gösterilmektedir. Yaşam süresinin  $t$ 'ye eşit veya  $t$ 'den küçük olma olasılığı olan ve aşağıdaki özelliklere sahip birikimli dağılım fonksiyonudur.

1.  $\int_0^{\infty} f(t) dt = 1$
2.  $f(t) \geq 0$ , tüm  $t$  için.

### 2.3. Yaşam Süresi Verileri Ve Sansürleme.

Sağkalım analizindeki en önemli değişken sağkalım süresidir. Sağkalım süresi çoğu kez sansürlü olmaktadır. Bireyin sağkalım süresi hakkında her zaman tam bir bilgiye ulaşılamayabilir. Veri yapısındaki bu farklılık, sağkalım analizini diğer istatistiksel analiz yöntemlerinden ayıran en önemli özelliktir.

Olay meydana gelmesine kadar geçen süre verileri, bir olay meydana geldiğinde ana bitiş noktası olarak zamanı olan verilerin olmasıdır. Olayın meydana gelmesine kadar zaman verilerine bazı örnekler, elektriksel bir bileşenin arızalanmasına kadar geçen süredir. Söz konusu süre hastalıktan iyileşmek için gereken süredir ve bu çalışmalarda denekler bir sağkalım süresi veya bir sansür süresi sağlayabilir. Takip süresinin sonunda verilerdeki tüm birimler olayı yaşamamışsa, buna sansür denir. Sansür, ilgili olayı gözlemlemeksizin ya da varılmadan sona erdiği anlamına gelir.

Sansürlü veriler, yaşam süresi tam olarak bilinmeyen verilerdir. Çeşitli nedenlerden dolayı tam olarak izlenemeyen birimler izlendikleri süre kadarıyla sağkalım analizine dahil edilir.

Sansürlü veriyi ortaya çıkaran üç ana sebep vardır: [26]

- 1- Bir birim çalışma bitmeden olayı yaşamaz.
- 2- Çalışma süresi boyunca birim kaybetmektedir.
- 3- Kişinin ölüm nedeniyle (ölüm çalışmadaki olay değilse) veya çalışmanın nedeni dışında bir sebepten dolayı çalışmadan çekilmesidir.

Çeşitli sansür türleri vardır, ancak sıraladıklarımız en yaygın olanlarıdır.

#### 2.3.1. Birinci tür sansürleme

Birinci tür sansürlemede çalışmanın toplam süresi sabittir. Olayların sayısı ise (Olayın meydana geldiği kişi sayısı.) rastgele bir değişkendir. Bu tip sansürleme, çalışmayı durdurma zamanının belirli bir süre sonra belirlendiği sabit sansürleme olarak adlandırılır. 1.tür sansürleme aşağıdaki biçimlerden birini alır:

### **1. Sağdan Sansürleme**

Sağkalım verilerinde en yaygın durumdur ve bu durum olayın meydana gelmediği birimlerle ilgilidir. Çalışmanın sonunda bazı birimler hayatta kalır, yani birimin sağkalım süresinin çalışmanın bittiği noktayı aştığı ve bu birimlerin sağdan kaybolduğu söylenmektedir. Sağdan sansürleme nedenleri:

1. Olayın meydana gelmesinden önce çalışmanın sonlandırılmasıdır.
2. Herhangi bir nedenle birime ulaşamamaktır.
3. Bazı birimlerde olay yoktur.

### **2. Soldan Sansürleme:**

Çalışmaya dahil edilen bir birim olduğunu, ancak tehlikeye maruz kalma zamanının bilinmediğini, yalnızca olayın meydana gelme zamanı bilindiğini varsayalım. Örneğin, kanser hastaları ve AIDS hastalarının, hastalığın başlangıcı bilinmiyor ama o hastalığa bağlı ölüm zamanı biliniyor ve bu sansürün soldan sansürleme olduğu söylenilmektedir.

### **3. Aralıklı Sansürleme**

Bu tür, önceki iki tür arasında bir karışım olarak sayılır. Bu durumda olayın kesin zamanı bazı birimler için bilinmez, ama olayın meydana geldiği aralık zamanı ise bilinmektedir. Bu birimlerin Aralıklı Sansürleme olarak söylenmektedir.

#### **2.3.2. İkinci tür sansürleme**

Bu durumda, olayın meydana geldiği birim sayısı önceden bilinir (sabittir), toplam çalışma süresi ise önceden bilinmeyen rastgele bir değişkendir ve içerisinde, çalışmayı bitirme zamanı, belirli sayıda olay meydana geldikten sonra belirlenmektedir. Demek ki, çalışmanın bitirme süresini belirli sayıda olay meydana geldikten sonra belirlendiği anlamına gelmektedir.

### 3. COX REGRESYON MODELİ

Cox modeli, hastanın sağkalım süresi ile hastanın sağkalım süresini etkileyen bağımsız değişkenleri seti arasındaki ilişkiyi açıklayan istatistiksel bir modeldir. Aynı zamanda bu model açıklayıcı değişkenlerin üstel fonksiyonu olan parametrik bir kısım ve temel hazard fonksiyonu olan parametrik olmayan kısım içerdiği için yarı parametrik bir model olarak da adlandırılmaktadır.

Cox modeli, sağkalım analizi modellerinde en önemli ve en yaygın modellerden biri olarak sayılmaktadır. Çünkü bu model, belirli bir olayın meydana gelmesinden önce gelen zaman değişkeninin çalışmada ilgili olguyu analiz etmede önemli olduğu durumlarda kullanılmaktadır [27]. 1972'de İngiliz bilim adamı David Cox tarafından önerilen Cox Regresyon modelinin formülü aşağıdaki gibidir:[28]

$$h\left(\frac{t}{x}\right) = h_0(t) \exp \sum_{i=1}^p \beta_i x_i \quad (3.1)$$

$h(t/x)$  : Modelin koşullu risk fonksiyonunu temsil etmektedir.

$h_0(t)$  : Modelin parametrik olmayan kısmını temsil eden) Zamana bağlı Temel risk fonksiyonu (Baseline Hazard Function).

$\exp \sum_{i=1}^p \beta_i x_i$  : Modelin parametrik kısmını temsil eden zamana bağlı olmayan göreceli risktir. Yani, artan veya azalan riski olan bağımsız değişkenlerin etkisi sabittir ve zaman değişimi için değişmez.

$x = (x_1, x_2, \dots, x_p)'$  Açıklayıcı değişkenlerin vektör değerlerini göstermektedir.

$B=(\beta_1, \dots, \beta_p)$ :Bilinmeyen regresyon katsayılarının vektör değerlerini temsil etmektedir.

Cox Regresyon modelini uygulanmadan önce aşağıdaki koşulların sağlanması gerekir:

1. Bağımlı değişken iki bölümden (ikili bağımlı değişken + olayın meydana gelmesinden önce gelen zaman değişkeni) oluşmaktadır.
2. Bağımsız değişkenler, doğalarına bakılmaksızın niceliksel, tanımlayıcı veya karışım olabilir, ancak çalışma ile ilgili fenomen üzerinde bir etkiye sahip olmaları beklenir.

### 3.1. Cox Modelinin Varsayımları

1. Cox regresyon modelinin Temel varsayımı, riskin (göreceli) olmasıdır. Risk oranının çalışma içinde bir kişiden diğerine sabit olduğunu ve risk oranının zamana bağlı olmadığı varsayılmaktadır. Örneğin, varsayalım ki:

$x^* = (x_1^*, x_2^*, \dots, x_p^*)$  ve  $x = (x_1, x_2, \dots, x_p)$  İki gözlem için ortak değişkenlerinden iki vektörler olduklarını varsayalım, bu durumda risk oranı (HR) şu şekilde tanımlanır:

$$\hat{H}\hat{R} = \frac{\exp(\beta x^*)}{\exp(\beta x)} = \exp[\hat{\beta}(x^* - x)] = \exp\left[\sum_{i=1}^p \hat{\beta}_i(x_i^* - x_i)\right] \quad (3.2)$$

Yani, herhangi iki risk oranı arasındaki oran sabittir ve zamana bağlı değildir.[29]

2. Cox Regresyon Modeli, diğer sağkalım modellerinin aksine ilk veri dağılımını varsaymaz.
3. Açıklayıcı (bağımsız) değişkenlerin etkisini her zaman sabit ve zaman geçtikçe değişmemektedir.
4.  $h(t/x)$  açıklayıcı değişkenlerden bir fonksiyonu olan (t) zamanındaki risk fonksiyonunu temsil eder ve  $\ln h(t/x)$  ise, açıklayıcı değişkenlerinden doğrusal bir fonksiyonudur .
5. Açıklayıcı değişkenler, risk fonksiyonunun logaritması ile doğrusal olarak ilişkilidir.[30]

### 3.2. Cox Regresyon Dönüşümü

Dönüşüm kullanımı, Cox Regresyon modelinin yorumlanmasındaki zorlukların ve sorunların çözülmesine ciddi katkılar sağlayabilir. Bilindiği gibi, Cox regresyon modelindeki risk fonksiyonu şu şekildedir:

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_p x_p) \quad (3.3)$$

$$\ln h(t) = (\beta_1 x_1 + \dots + \beta_p x_p) \quad (3.4)$$

Risk fonksiyonunun logaritması, bağımsız değişkenlerin doğrusal bir kombinasyonudur. Cox regresyon modelinde logaritmik dönüşümünü kullanmanın amacı, ikili bağımlı değişkenlerle veriler için ilişkileri analiz ederken doğrusal regresyon uygulamasına izin vermek olduğu belirtilmektedir. Yani, bağımsız değişkenlerde

meydana gelen herhangi bir deęişiklięin, doęrusal regresyonda olduęu gibi baęımlı deęişkenin kendisinde deęil, yalnızca risk fonksiyonunun logaritmasında bir deęişikliğe yol açmaktadır.

### 3.3. Model Parametrelerinin Tahmini

Baęımlı deęişkenler ikili olduęunda model parametrelerini en küçük kareler yöntemiyle tahmin ederken ortaya çıkan istatistiksel problem, doęrusal regresyon analizinin hipotezlerinin saęlanmamasıdır. Bunlardan en önemlisi normallik ve varyans homojenliği varsayımlarının, yani baęımlı deęişkenin ikili doğası nedeniyle saęlanamamasıdır. Bu nedenle, bu tür baęımlı deęişkenler söz konusu olduęunda parametre tahmini en çok olabilirlik yöntemi (maximum likelihood) kullanılarak tahmin edilmektedir. Baęımlı deęişkenin ikili olduęu yarı parametrik Cox Regresyon modelinde ise, Cox regresyon modeli parametreleri kısmi olasılık (Partial likelihood) fonksiyonu kullanılarak tahmin edilmektedir.

### 3.4. Partial Likelihood Yöntemi

1975'te bilim adamı Cox tarafından önerilen bir yöntemdir ve modelin parametrelerini tahmin etmek için kısmi olabilirlik fonksiyonu kullanılır. Baęımsız deęişkenlerdeki deęişikliklerle tehlikelerdeki deęişiklikleri içeren modelin yalnızca ikinci kısmının tahminine izin vermektedir. Kısmi olabilirlik fonksiyonu aşağıdaki gibidir:

$$l_p(\beta) = \prod_{i=1}^k \left( \frac{\exp(BX_i)}{\sum_{j \in R(t_i)} \exp(BX_j)} \right) \quad (3.5)$$

B parametresinin kısmi olabilirlik fonksiyonunun logaritmasını alırsak aşağıdaki formülü elde edilmektedir:

$$\log(l_p(\beta)) = \sum_{i=1}^k \left[ (X_i B) - \log \sum_{j \in R(t_i)} \exp(BX_j) \right] \quad (3.6)$$

parametresi için ilk diferansiyeli alırsak aşağıdaki formülü elde edilmektedir:

$$\frac{\partial \log(l_p(\beta))}{\partial B_i} = \sum_{i=1}^k \left[ X_i - \frac{\exp(BX_i) X_{ij}}{\sum_{j \in R(t_i)} \exp(BX_j)} \right] \quad (3.7)$$

Denklemleri sıfıra eşitleyerek ve denklemleri çözmek için Newton Raphson yöntemini kullanarak, B parametresinin tahminlerini elde edilmektedir.

### 3.5. Cox Regresyon Modelinin Uyumunun Değerlendirmesi

Cox modelinin parametrelerinin tahmin edilmesinden ve göreceli risk varsayımının test edilmesinden sonraki adım, modelinin uyumunun değerlendirilmesi, yani bu modelin çalışma verilerine ne ölçüde uyduğunu test etmek olacaktır. Modelin uyumluluğunu doğruladıktan sonra, tahmin edilen model parametrelerinin anlamlılığını test edilmesinde kullanılacak olan yöntemlere sırasıyla yer verilecektir.

#### 3.5.1. Olabilirlik Oran Testi (Likelihood Ratio Test)

Önerilen modelin uyumluluğunu değerlendirme sürecindeki ilk adım, genellikle modeldeki değişkenlerin bir bütün olarak önemini, yani modeldeki bağımsız değişkenlerin istatistiksel olarak bağımlı değişkenlerle ilişkili olup olmadığının belirlenmesidir. Kesin modeli değerlendirmeye yardımcı olan en önemli standartlarından biri ise olabilirlik oran testi, genel modelin önemini kontrol etmek için kullanılan en önemli testlerden biridir. Başka bir deyişle uygun modelin seçilmesinde kullanılan önemli testlerden biridir. Çünkü modellerin çokluğu, incelenen verilere hangi modelin en uygun olduğunu bilmenin önemini zorunlu kılmaktadır. En uygun model, olabilirlik oran testi kullanılarak belirlenebilir. Burada bu test iki veri modeli tahmin edilerek yapılır, ardından iki modelin olabilirlik fonksiyonunun logaritması üzerinden ilk model ile ikinci model karşılaştırılır ve en düşük değere sahip model en iyisidir. Ancak farkın anlamlı olduğunun bilinmesi gerekir ve bu testin Ki-kare dağılımını izleyen Olabilirlik Oran Testi kullanılarak yapılır ( $\chi^2$ ) ve bu test için matematiksel formül,

$$G = 2[\log L(\hat{\beta}) - \log L(0)] \quad (3.8)$$

biçiminde tanımlanır

$(L(\hat{\beta}))$ : tam modeli (etkileşimli modeli).

$(L(0))$ : İndirgenmiş modeli (etkileşimsiz modeli).

Hesaplanan Ki-kare ( $\chi^2$ ) değeri tablo değeri ile karşılaştırılır. Hesaplanan Ki-kare ( $\chi^2$ ) değeri tablo değerinden büyükse, bağımsız değişkenlerin katsayısı sıfıra eşit değildir. Yani tam model bağımsız değişkenlerinin varlığı bağımlı değişkende istatistiksel olarak anlamlı bir etkiye sahiptir. Tam tersine, hesaplanan Ki-kare ( $\chi^2$ )

değeri tablo değerinden küçükse, bağımsız değişkenlerin katsayıları sıfıra eşittir. Yani bağımsız değişkenler bağımlı değişken üzerinde istatistiksel olarak anlamlı bir etkiye sahip değildir.

### 3.5.2. Wald Testi (Wald Statistics).

Wald testi, Cox regresyon katsayılarının her biri için istatistiksel önemi (katsayıların anlamlılığını) test etmek için kullanılan, her değişkenin önemini ayrı ayrı ölçmede önemli testlerden biri olarak kabul edilmektedir. Modeldeki kalması gereken önemli değişkenler nelerdir? Ve bunlardan hangileri anlamlı değildir ve modelden çıkarılması gerekir? Yani Wald testi ile Cox'un bağımsız değişken X ile ilişkili regresyon katsayısının sıfır hipotezinin testi gerçekleştirilmektedir.

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

biçimde yazılır.

$$Wald = \left( \frac{\hat{\beta}}{SE_{\hat{\beta}}} \right)^2 \approx \chi^2 \quad (3.9)$$

$\hat{\beta}$ : X değişkeninin Cox regresyon katsayısı.

SE: X değişkeni için Cox regresyon katsayısının standart hata değeri.

### 3.5.3. Akaike ölçütü (Akaike information criterion-AIC)

Akaike ölçütü (Akaike information criterion-AIC) belirli bir veri kümesi için kaliteli bir istatistiksel göreceli model ölçüsüdür. Yani, veri modelleri koleksiyonu verildiğinde, AIC her model kalitesini, diğer modellerin her birini göreceli olarak tahmin ediyor. Dolayısıyla, AIC model seçimi için bir yol sağlar.

$$AIC = 2k - 2 \ln(L) \quad (3.10)$$

K = modeldeki tahmin edilen parametrelerin sayısı

L = Tahmin edilen modelin maksimum Olabilirlik fonksiyonu.



#### 4. YAPAY SİNİR AĞLARI

Bu bölüm, yapay sinir ağlarının tanımını, tarihçesini, türlerini, bileşenlerini, ek olarak yapay sinir ağındaki işlem öğelerini, sinir ağlarının teorik yönünü ele alacaktır. Ayrıca, YSA, veri akışı yöntemlerine, içindeki katman sayılarına ve öğrenme yöntemlerine göre sınıflandırılacaktır. Son olarak, sinir ağlarının performansını değerlendirmek için yöntemler veya kriterler ele alınacaktır.

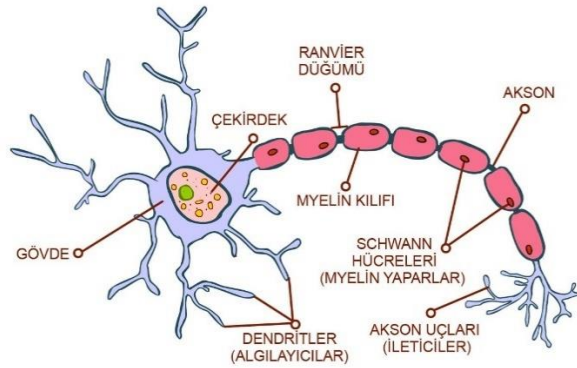
Yapay sinir ağları, biyolojik sinir ağlarını insan düşünme biçimine ve beynin bilgiyi nasıl işlediğine dayalı matematiksel modellerle simüle eden bir bilgi işleme sistemi olarak sayılır. Demek oluyor ki yapay sinir ağları, elektronik bilgi işleme modelleridir. YSA belirli problemleri çözmek için uyum içinde çalışan, ayrıca bilgiyi işlemek için biyolojik sinir sisteminden ilham alan birçok işlem elemanından oluşan modellerdir.

Yapay sinir ağları bilgiyi insan zihnini simüle etmek şeklinde işlediğinden ve yapay sinir ağlarının nasıl çalıştığını anlaşılabilmesi için biyolojik sinir ağlarının nasıl çalıştıklarının bilinmesi gerekir. Biyolojik sinir sisteminin yapay sinir ağlarının mimarisinin önerildiği temelde özelliklere sahip olduğu yapay sinir ağlarının anlaşılmasına yardımcı olmaktadır. Ayrıca, yapay nöron (işlem elemanı), biyolojik nörona benzerdir.

İnsan sinir sistemi iki bölümden oluşur: beyni ve omuriliği içeren merkezi bir sistem (Central nervous system CNS ) ve beyin sinirlerini, omurilik sinirlerini ve otonom sinirleri içeren bir alt sistem (peripheral nervous system). Sinir sisteminin oluşturma birimi hücredir (nöron). Nöronlar, gövde ve aksondan (Axon) oluşur ve hücre gövdesi hücre çekirdeğini içerir. Ayrıca yüzeyinden, elektrik sinyallerinin alınması veya iletilmesi ile ilgili dendritler ortaya çıkar. Sinir sisteminin temel bilgi işlem birimi hücredir (nöron). Uzantıların diğer birçok nörondan sinyal aldığı ve bu sinyaller, kimyasal işlemler olan sinapslar (Neurotransmitters ) aracılığıyla iletilen elektriksel impulslardır. Görevleri, yapay sinir ağlarındaki ağırlıkların çalışmasına benzer bir şekilde nörotransmitterler tarafından sinyalleri modüle etmek ve iletmektir.

Nöronun aksonu(Axon), hücre gövdesinden çıkan ve elektrik sinyallerini alıp diğer hücrelere yönlendirerek nörondan gelen elektrik sinyallerini ileten bir uzantıdır. Sinir sistemindeki nöronların gövdeleri çekirdek veya düğümler adı verilen gruplar halinde toplanır ve nöronların aksonları bir araya gelerek sinirleri oluşturur. Sinyal, ister başka

bir hücreye giriş veya çıkış sinyali olsun, aralıklı zaman adımlarına göre işlenir. Nörondaki hücre gövdesi, belirli yoğunluklara göre sitoplazma içerir ve bu yoğunlukların değerleri, hücrenin aktivitesini ve sinyali bir hücreden diğerine iletme yeteneğini belirlemektedir[31].



**Şekil 4.1.** Biyolojik nöronun genel yapısı [32]

Yapay sinir ağları (YSA), insan beyninden esinlenilerek geliştirilmiş, ağırlıklı bağıntılar aracılığıyla birbirine bağlanan ve her biri kendi bağına sahip işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır. Bir başka deyişle, biyolojik sinir ağlarını taklit eden bilgisayar programlarıdır. Demek oluyor ki yapay nöronlar, en azından kısmen, beyin nöronlarının gerçek davranış ve işlevlerinden ilham almıştır.

Yapay sinir ağları zaman zaman bağlantıcılık, paralel dağıtılmış işlem, sinirsel- işlem, doğal zekâ sistemleri ve makine öğrenme algoritmaları gibi isimlerle de anılmaktadır. Yapay sinir ağları bir programcının geleneksel yeteneklerini gerektirmeyen, kendi kendine öğrenme düzenekleridir. Bu ağlar öğrenmenin yanı sıra, ezberleme ve bilgiler arasında ilişkiler oluşturma yeteneğine de sahiptir[7].

Yapay sinir ağları günümüzde birçok probleme çözüm üretebilecek yeteneğe sahiptirler ve tanımlarının birkaç ortak noktası bulunmaktadır. Bunların en başında yapay sinir ağlarının birbirine hiyerarşik olarak bağlı ve paralel olarak çalışabilen yapay hücrelerden oluşmaları gelmektedir. Proses elemanları da denilen bu hücrelerin birbirine bağlandıkları ve her bağlantının bir değerinin olduğu kabul edilmektedir.

YSA'lar, birbirine bağlı bir yapay nöron grubu olarak birkaç katmandan oluşur. Ayrıca her katman, nöronların birbirleri üzerindeki etki miktarını gösteren bir ağırlığa

sahiptir. Genellikle, bir YSA modelinde girdi, gizli ve çıktı katmanları adı verilen üç katman bulunur. Yapay sinir ağı modeli basit olarak şekil (4.8)'de gösterilmiştir. Ayrıca, YSA, fikirlerini insan beynini simüle etmeye çalışmaktan ve sinirlerin iç içe geçmesinden alan makine öğrenimi alanındaki önemli algoritmalarından biridir.

Yapay Sinir Ağlarının doğuşu ve evrimi, sinir ağları fikrinin başlangıcı, Warren McCulloch ve Walter Pitts tarafından yayınlanan, zihnin nasıl çalıştığına dair matematiksel bir modelin ve hala kullanılan bir nöron modelinin açıklamasının yer aldığı bir makale ile 1943'e kadar uzanmaktadır. Söz konusu matematiksel model sinir ağlarının temel bileşenini olarak sayılmaktadır [33]. 1949'da Donald Hebb, birlikte kullanılan biyolojik nöronlar arasındaki bağlantıların her kullanımda nasıl güçlendiğini açıkladığı, öğrenmenin beyinde nasıl gerçekleştiğini açıklayan bir makale yayınladı [34]. Bu modellere dayanarak, en ünlüsü, 1943'te McCulloch ve Pitts tarafından geliştirilen modelin bir simülasyonu olan algılayıcı olarak isimlendirilen (Perceptron) birkaç sinir ağı önerilmiştir. Alıcı, girdileri alan, bunları ağırlıklı biçimde toplayan ve ardından sınırın değerine bağlı olarak 0 veya 1 değeri veren doğrusal bir sınıflandırıcıdır [35]. O zamanlar sinir ağlarında kusur olan şey, doğrusal olmayan sorunları sınıflandıramayan doğrusal bir sınıflandırıcı olmasıydı ve sinir ağlarındaki araştırmalar, sorun daha sonra çözülene kadar durgun bir aşamaya girdi.

Doksanların başında, sinir ağlarının modern gelişiminin başlangıcı olarak kabul edilen birkaç çalışma ortaya çıktı. Bu çalışmalar (Hines) 1993 çalışması, (Emmanuel)1993 çalışması, ve (Rene Bewailed) 1993 çalışmasıdır [36]. 1994 yılından itibaren günümüze kadar yapay sinir ağlarına ilgi artmıştır. Konferansların düzenlenmesi, özel dergilerin ortaya çıkması ve hükümetlerin sinir ağlarıyla ilgili bir takım projeye yatırım yapmasıyla zirveye ulaştı.

Yukarıdakilerden ve sinir ağlarının gelişim aşamaları izlenerek, sinir ağlarının tıp, mühendislik ve endüstriyel alanlar da dahil olmak üzere çeşitli alanlarda sinir ağlarının kullanımının başladığı modern çağına ulaşılmasına kadar birkaç aşamadan geçtiği açıktır. Hatta çoğu alanlarda klasik modellere göre doğruluk ve hız açısından üstünlük göstermişti.

#### 4.1. Yapay Sinir Ağının Bileşenleri.

Yapay sinir ağları, insan beynindeki biyolojik nöronlara benzeyen nöronlar veya sinir hücreleri adı verilen bir grup işlem biriminden oluşur ve bu birimler birbirine sinir ağı adı verilen ara bağlantılar şeklinde bağlanır. Böylece bağlantılar, sinyalleri bir nörondan diğerine geçirmek için çalışır ve her bağlantının kendisine eşlik eden sayısal bir ağırlığı vardır. Ayrıca her nöron, sinapsları aracılığıyla bir dizi girdiler alır ve yalnızca bir çıktılar sinyali üretir; burada çıkış sinyali, nöronlar aracılığıyla iletilir nöronun dış bağlantısı ve sırayla dış bağlantı, aynı sinyali ileten bir dizi Şubeye bölünmüştür. Bu bileşenlerin grubu, sinir ağlarının genel yapısını veya genel biçimini temsil etmektedir. Herhangi bir sinir ağı, bağlantılara ek olarak en az iki katman olmak üzere birkaç katmandan oluşmaktadır. Demek oluyor ki, bir sinir ağı aşağıdaki bileşenlerden oluşmaktadır[32][37].

##### 4.1.1. Giriş katmanı (Input Layer)

Nöronlar aracılığıyla ağına veri beslendiği katmandır. Bu katman, ağ yapısına bağlı olarak bir veya daha fazla (nöron) işlem biriminden oluşabilir. Aynı zamanda ağ girdileri hem veri şeklinde olabilir veya diğer işlem birimlerinden çıkışlar da olabilmektedir. Başka bir deyişle, girdi katmanı, modelin bağımsız değişkenlerinin sayısına eşit sayıda işlem elemanı içermektedir. Bu katman, farklı kaynaklarından veri alır, böylece her girdiğini eleman bir özelliği veya bağımsız bir değişkeni temsil etmektedir. [34]

Girdi katmanındaki işlem birimlerinin herhangi bir aritmetik işlemi yoktur, sadece bu katmandan verileri ağırlıklar aracılığıyla gizli katmandaki işlem birimlerine veya ağına gizli katman içermemesi durumunda çıkış katmanındaki işlem birimlerine aktarır.[38]

##### 4.1.2. Çıktı katmanı (Output Layer)

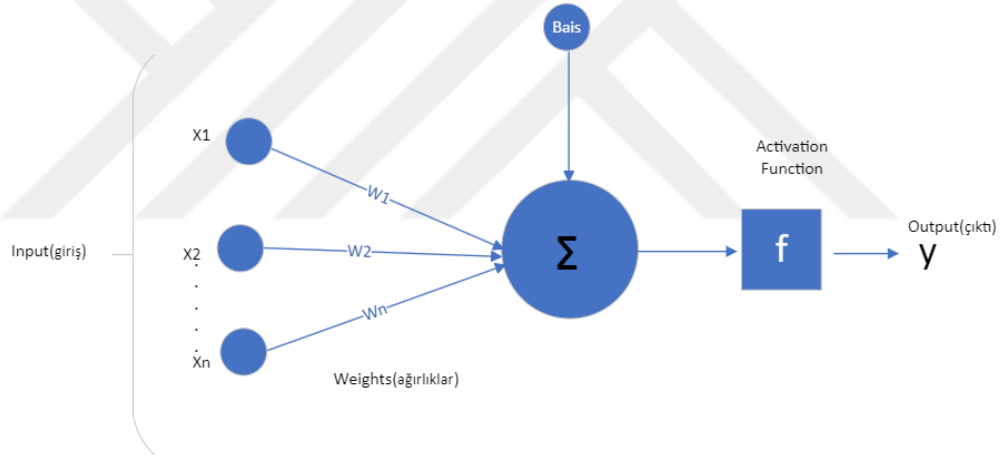
Bu katman, ağına nihai çıktısının çıktı olarak alındığı işlem birimlerinden oluşur. Bu katman, ağına mimarisine bağlı olarak bir veya daha fazla işlem birimi içerebilir. Bu katmandaki işlem birimlerinin, kendisine doğrudan giriş katmanından veya gizli katmandan gelen sinyalleri almaktadır. Gerekli işlemleri yaptıktan sonra nihai çıkışları ile bir sinyal gönderir veya istenen gibi veri işleme yapılmadığında bu çıkışları tekrar giriş olarak ağa döndürür.

#### 4.1.3. Gizli katman (Hidden Layer)

Bu katman, girdi katmanı ile çıktı katmanı arasında yer almaktadır. Tek bir yapay sinir ağı, bir veya daha fazla gizli katman içerebilir ve bazı ağların yapıları herhangi bir gizli katman içermeyebilir. Gizli katman, bağlantılar aracılığıyla kendisine gelen sinyalleri aldığı yerde, bunları işleyerek gerekli işlemi yapar ve ardından çıkış katmanına göndermektedir.[32]

#### 4.1.4. Bağlantılar veya ağırlıklar (Connections)

Bağlantılar veya ağırlıklar, farklı ağ katmanları arasında katmanları birbirine bağlayan veya her bir ara bağlantısına eşlik eden ağırlıklar aracılığıyla her katmandaki birimleri birbirine bağlayan bir iletişim bağlantısıdır. Bağlantıların ana görevi, katmanlar veya işlem birimleri arasında veri ya da sinyaller aktarmaktır. [37]



Şekil 4.2. Yapay nöronun genel yapısı

#### 4.2. Yapay Sinir Ağlarında İşlem Birimleri (Processing Elements)

İşlem birimleri veya nöronlar, sinir ağında bilgi işleme sürecini gerçekleştiren birimlerdir ve sinir ağının tüm katmanlarının temel bileşenlerini oluşturan ve bu birimler yapay sinir ağının genel şekli veya yapısını vermek için ara bağlantılarla farklı şekillerde bağlanır. İşleme, aritmetik işlemlerin gerçekleştirilmesinde ve verilerin işlenmesinde paralel işlem sistemini almakta, bir insanın problem çözmeyi düşünme şekline benzer (İnsan zihninin görevi), ve herhangi bir işlem birimi(nöron) aşağıdaki bileşenlerden oluşmaktadır:

- Ağırlık Katsayıları.
- Toplama Fonksiyonu.
- Transfer Fonksiyonu.
- Çıkış Fonksiyonu.

#### 4.2.1. Ağırlık katsayıları (Weight coefficients)

Bağlantılar veya ağırlıklar, farklı ağ katmanları arasında katmanları birbirine bağlayan veya her bir ara bağlantısına eşlik eden ağırlıklar aracılığıyla her katmandaki birimleri birbirine bağlayan bir iletişim bağlantısıdır. Bağlantıların ana görevi, katmanlar veya işlem birimleri arasında veri ya da sinyaller aktarmaktır.

#### 4.2.2. Toplama fonksiyonu (summation function)

İşlem birimi tarafından gerçekleştirilen ilk işlem, toplama fonksiyonu kullanılarak birime gelen ağırlıklı girdilerin hesaplanmasıdır. Burada bu fonksiyon, işlem birimine gelen tüm girdilerin ortalama ağırlıklarını hesaplanma, bu girilen her değeri, beraberindeki ağırlığıyla çarparak ardından çarpmanın sonucu toplamının bulunmasıyla yapılmaktadır. Toplama Fonksiyonu matematiksel formülü aşağıdaki gibidir:[39]

$$S_j = \sum_{i=1}^n X_i W_{ij} \quad (4.1)$$

$S_j$ : Her işlem birimi J için toplama işleminin sonucu.

$X_i$ : i biriminden gelen ve j birimine giren bir giriş değeri.

$W_{ij}$ : Önceki katmandaki işlem birimi j'yi birim i'ye bağlayan bir ağırlık.

#### 4.2.3. Transfer fonksiyonu (Transfer function)

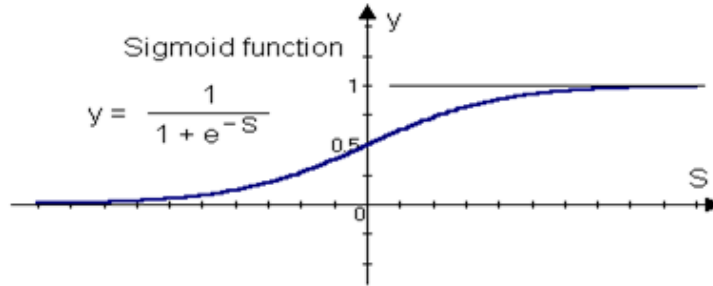
Toplama işleminin sonucunu istenen ağırlık çıkışlarından biri olması gereken değerlerden birine dönüştürmek olan toplama fonksiyonunun çalışmasından sonra işlem birimindeki ikinci işlemdir. Bu, toplama sonucunun eşik adı verilen (Threshold ) ve ( $\emptyset$ ) sembolü ile gösterilen bir değerle karşılaştırılmasıyla yapılır. Bunu belirlemek için, karşılaştırmadan önce toplama özel bir aktivasyon fonksiyon uygulanır ve buna dönüşüm (aktivasyon) fonksiyonu olarak adlandırılmaktadır.

Aktivasyon fonksiyonu, girdi ve çıktı arasındaki doğrusal ya da doğrusal olmayan eşleştirmeyi sağlayan fonksiyondur. Aktivasyon fonksiyonunun farklı biçimleri vardır, bunlardan en önemlilerine aşağıda yer verilecektir.[40][41][42].

#### 4.2.3.1. Sigmoid aktivasyon fonksiyonu (Sigmoid function)

Bu fonksiyon, çıktıyı (0,1) arasına sıkıştırılmış bir değere çevirir. Sigmoid aktivasyon fonksiyonu ya da ikili lojistik aktivasyon fonksiyonu veya çıktıları (-1, +1) arasındaki değerlere dönüştürmek denir ve iki kutuplu lojistik aktivasyon olarak adlandırılır. Sigmoid fonksiyonu veya lojistik fonksiyonu, geri yayılım ağı için eğitim algoritmalarında en çok kullanılan fonksiyonlarından biridir ve bu fonksiyon aşağıdaki matematiksel formülü almaktadır[42].

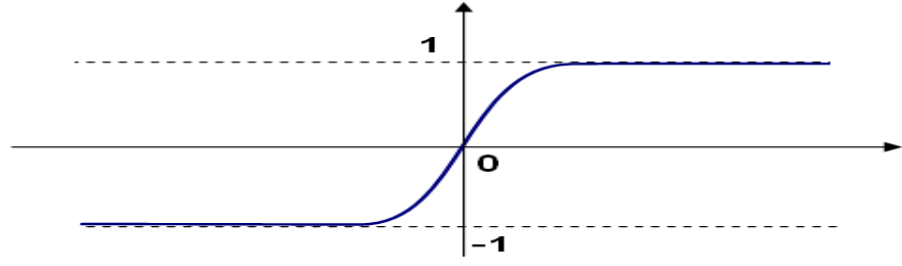
$$F(X) = \frac{1}{1 + e^{-s}} \quad (4.2)$$



Şekil 4.3. İkili Sigmoid aktivasyon fonksiyonu

Kutupsal İkili sigmoid aktivasyon fonksiyon istenilen çıktı değerleri aralığı  $[+1 \ -1]$  arasında olduğunda bir sinir ağında aktivasyon fonksiyonu olarak kullanılır ve aşağıdaki gibi matematiksel formülü almaktadır.

$$\begin{aligned} g(x) &= 2f(x) - 1 = \frac{2}{1 + \exp(-x)} - 1 \\ &= \frac{1 - \exp(-x)}{1 + \exp(-x)} \end{aligned} \quad (4.3)$$

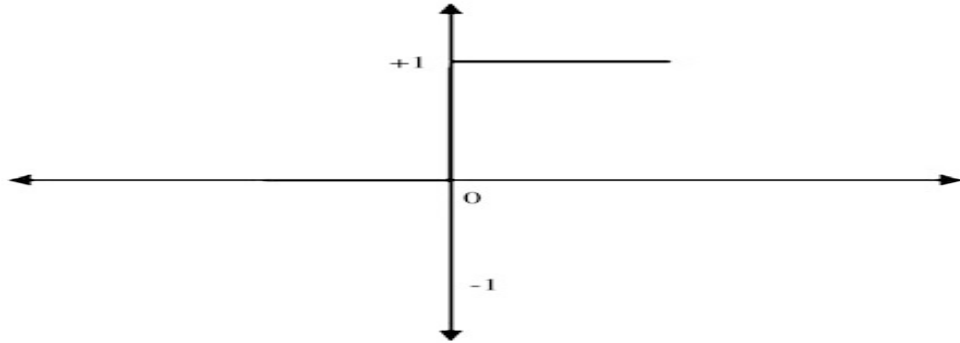


Şekil 4.4. kutupsal ikili sigmoid aktivasyon fonksiyonu

#### 4.2.3.2. Adım fonksiyonu (Step function)

Adım fonksiyonuna ikili fonksiyon da denir. Bu fonksiyon sınır çizgisi fonksiyonlarından biri (Hard limiter) olarak bilinir ve bu fonksiyon, ikili sistem gibi işlem biriminden çıkış değerini 0 ile 1 arasında sınırlandırır, bu nedenle ikili aktivasyon fonksiyonu olarak veya ikili adım işlevi adlandırılır, ve matematiksel formu şekildeki gibidir:

$$f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (4.4)$$

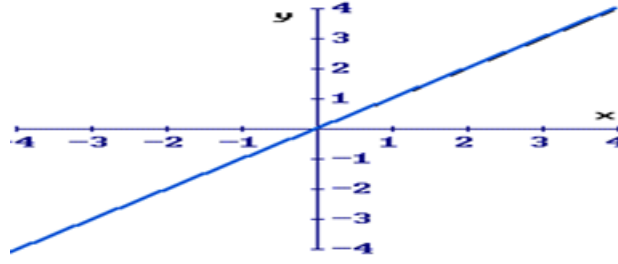


Şekil 4.5. Adım fonksiyonu

#### 4.2.3.3. Doğrusal fonksiyon (linear function)

Doğrusal aktivasyon fonksiyonu, işlem biriminin ağırlıklı girişlerine eşit çıktılar sağlar. Genellikle işlem birimleri, doğrusal filtrelerde kullanılan nöronlarda doğrusal fonksiyonu kullanır ve bu fonksiyon aşağıdaki matematiksel formu alır[42].



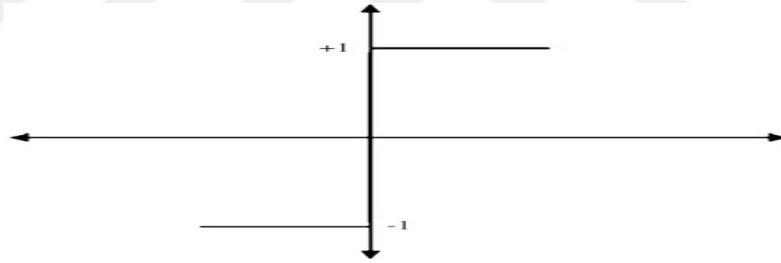


Şekil 4.6. Doğrusal fonksiyonu

#### 4.2.3.4. Sinyal fonksiyonu (Sign function)

Bu fonksiyon, çelik sınırı (hard limit) işlevlerinden biri olarak kabul edilir ve genellikle sınıflandırma ve örüntü ayırmacılıkta işlemlerinde kullanılan sinir ağlarının işlem birimlerinde kullanılır ve bu fonksiyon,  $\theta$  olarak adlandırılan belirli bir değeri kullanır[7].

$$f(x) = \begin{cases} +1, & \text{if } x \geq 0 \\ -1, & \text{if } x < 0 \end{cases} \quad (4.5)$$



Şekil 4.7. Sinyal fonksiyonu

#### 4.2.4. Çıkış fonksiyonu

Toplama fonksiyonu ağırlıklı toplama işlemini gerçekleştirdikten sonra, transfer fonksiyonu toplamının sonucunu belirli bir aralıkla sınırlı bir değere dönüştürür. Daha sonra çıktı fonksiyonu yapay sinir ağının nihai çıktısını verir ve çıktılar çoğu durumda dönüşüm fonksiyonunun çıktısına eşit olabilmektedir. Ancak bazı ağlarda, işlem biriminin transfer fonksiyonunun sonucunu değiştirdiği için, komşu işlem birimlerinin birbirleriyle rekabet etmesiyle yapılır. Rekabet genellikle en büyük aktivasyona sahip işlem birimleri arasında olur. Bu rekabet hangi işlem biriminin aktif olacağını veya

hangilerinin çıktısı gerçekleştireceğini belirlemektedir. Aynı zamanda, öğrenme ve eğitim sürecine dahil olan birimlerin belirlenmesine de yardımcı olmaktadır.[43]

Yapay sinir ağındaki işlem biriminin görevleri aşağıdaki gibi özetlenebilir:

- ❖ Gizli katmandan veya giriş katmanından sinyallerin veya girdilerin alınması.
- ❖ Her sinyalin birim bağlantı hattındaki ağırlıkla çarpıldığı ve ağırlıklar aracılığıyla birime giren sinyalin modifikasyonu.
- ❖ Toplama fonksiyonunu kullanarak diğer birimlerden gelen girdin ağırlıklarını biriktirmesi.
- ❖ Bu birimden çıkış sinyali belirlenene kadar ağırlıklı giriş sinyallerinin toplamına belirli bir aktivasyon fonksiyonu uygulanması.
- ❖ Bu birim tarafından kaynaklanan çıktı, diğer birkaç işlem birimine yayınlanabilir ya da ağın nihai çıktısı olabilmektedir.

#### **4.3. Yapay Sinir Ağlarının Sınıflandırması**

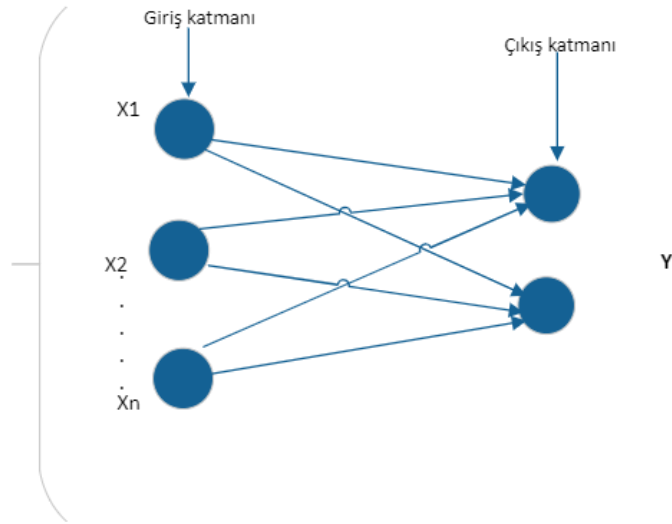
Yapay sinir ağları birkaç temele göre sınıflandırılabilir. Bu temellerden en önemlisi, ağları oluşturan katman sayısına öğrenme veya eğitim yöntemine göre, ayrıca veri akışının yöntemine ve çözülecek problemin türüne göre sınıflandırılması mümkündür.

##### **4.3.1. Katman sayısına göre sınıflandırma**

Yapay sinir ağları, katmanlar sayısına göre aşağıdaki şekilde sınıflandırılabilir.

###### **4.3.1.1. Tek katmanlı Ağlar**

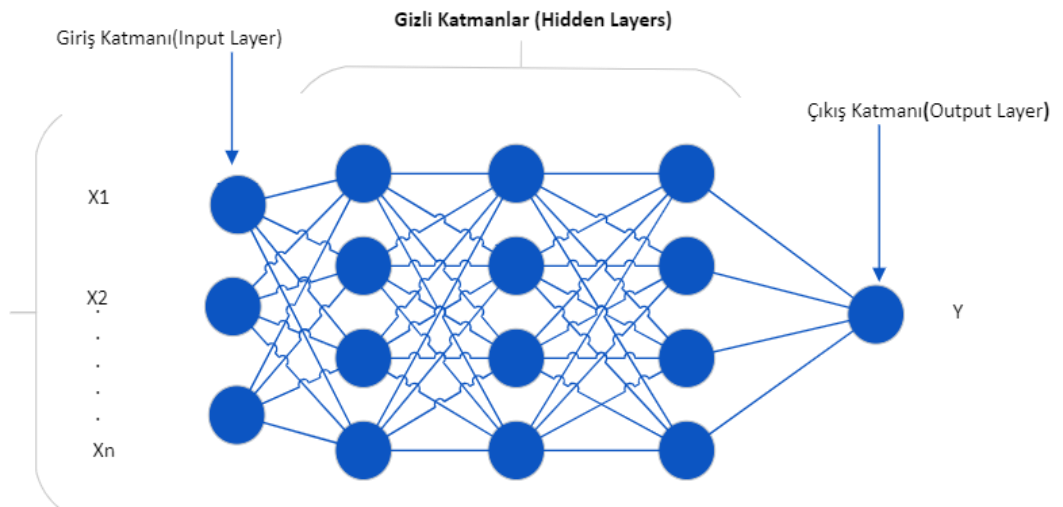
Tek katmanlı ağlar, ağı yapılarının en basit türlerinden biri olarak kabul edilir ve genellikle, ağı girdilerini çıktılarına doğrudan bağlayan tek bir işleme öğeleri katmanından oluşmaktadır. Tüm hesaplamalar çıktı katmanında gerçekleştirilir ve genellikle ağdaki katman sayısı, herhangi bir matematiksel işlem yapmadığından giriş katmanı dikkate alınmadan hesaplanmaktadır.



**Şekil 4.8.** Tek katmanlı ağı

#### 4.3.1.2. Çok katmanlı ağlar

Bu ağı yapıları, ara bağlantılar (ağırlıklar) tarafından birbirine bağlanan birden fazla işlem öğeleri katmanı içerir. Burada çok katmanlı ağı, girdi katmanı ve çıktı katmanı olmak üzere en az iki katmandan oluştuğunda girdi ve çıktı katmanları arasında gizli katman bulunur ve ağı, ağıda kullanılan uygulamanın türüne bağlı olarak birden fazla gizli katman içerebilmektedir.[38]



**Şekil 4.9.** Çok katmanlı ağı

#### 4.3.2. Öğrenme ve eğitim yöntemleri göre sınıflandırma

Yapay sinir ağları, öğrenme ve eğitim yöntemleri göre aşağıdaki türlere sınıflandırabilir:

##### 4.3.2.1. Gözetimli öğrenme (Supervised learning )

Danışmanlı öğrenme yönetiminde, yapay sinir ağı kullanılmadan önce eğitilmelidir[44]. Bu yöntemin temeli, eğitim verilerinin ağına bir çift olarak, doğrusu girilen şekil (input) ve istenilen şekil şeklinde (target) sunulması fikrine dayanmaktadır. Başka bir ifadeyle, eğitimin bir çift gerçek girdi ve istenilen çıktının eklenmesi ve istenilen çıktı ile gerçek çıktı arasındaki farka göre ağırlıkların ayarlanmasıyla gerçekleştirildiği ve iki değer arasındaki (input and target) fark sıfıra yakın olana kadar eğitim devam etmektedir. Gerçek ve istenen çıktılar arasındaki fark aşağıdaki formüle göre belirlenmektedir. Gerçek çıktı (Real output) - Hedef çıktı (Target output) = hata (error)[45]

##### 4.3.2.2. Gözetimsiz öğrenme (Unsupervised learning )

Bu öğrenme yöntemine "kendi kendine öğrenme yöntemi" olarak ta adlandırılmaktadır. Danışmansız veya kendi kendine öğrenme yönetiminde bir öğrenme sürecini yönetmek için bir öğretmen yoktur. Bu yöntemde göre eğitim, ağ üzerinde istenen hedefi görüntülemekten yalnızca girdilerin bir vektörüdür. Başka bir deyişle, yapay sinir ağlarının, kendilerine sunulan şekil ve örüntülerin özelliklerini keşfetme ve danışmanlı öğrenme yönteminin aksine, bu şekillerin içsel bir temsilini önceden bilgi sahibi olmadan ve ne üretmeleri gerektiğine dair örnekler göstermeden geliştirme yetenekleri temelinde eğitim yöntemleridir .[46]

##### 4.3.2.3. Hibrit öğrenme (Hybrid learning )

Bu yöntem, istenilen çıktıların gerçek değerlerinin şekli danışmansız öğrenmede olduğu gibi ağa açıklanamaması, danışmanlı öğrenme ve danışmansız öğrenme olmak üzere iki yöntemin karışımına dayanmaktadır. Başka bir deyişle, sadece çıktısının ağına sonucu doğru olup olmadığı danışmanlı öğrenme yönteminde olduğu gibi belirtilmektedir. Demek oluyor ki bu yöntemde ağ, deneme yanılma kuralına göre

öğrenir, ağ doğru tepkiler gösterdiğinde bu tepkiler pekiştirilir ve istenmeyen çıktı ya da kötü performansa yol açan tepkiler zayıflandırılır.

#### **4.3.3. Veri akışının doğasına göre sınıflandırma**

Ağlar, ağ katmanları arasında veri akışının veya yayılmasının doğasına ve bunların ara bağlantılarına göre aşağıdaki şekilde sınıflandırılabilir:

##### **4.3.3.1. İleri beslemeli ( Feed forward neural networks)**

İleri beslemeli sinir ağı, en basit, en küçük boyutlu ve en yaygın kullanılan türlerden biri olarak kabul edilir ve bu tür en az iki katmandan oluşur. İleri beslemeli sinir ağı bazen bir veya birden fazla gizli katman içerebilir bazen ise herhangi bir gizli katman içermez. Veri, giriş katmanından çıkış katmanına sadece tek yönde ileri, varsa gizli katmandan geçerek ve orada hiç bir akar geri yön yoktur. Bu ağın türü genellikle ses tanıma, görüntü tanıma ve kendi kendine sürüş için kullanılmaktadır.

##### **4.3.3.2. Geri beslemeli (Feedback neural networks)**

Ağlar, en iyi sonuçları vermek amacıyla tekrar girdiler olmak üzere çıktıları tekrar geri yol olmaya çalışmaktadır. Ağı eğitmek için bir geri yayılım algoritmasının kullanıldığı çok katmanlı sinir ağlarıdır. Bu ağ, eğitim sürecinin, ağın eğitim sürecinde kullanılan girdi verilerine doğru yanıt verme yeteneği arasında bir denge durumuna ulaşmayı amaçladığı danışmanlı bir öğrenme yöntemiyle eğitilir ve eğitim verilerine benzer girdi verilerine iyi bir yanıt verme yeteneğidir. Bu ağlarından bir örnek olarak, Hopfield ağıdır.[47]

#### **4.4. Sinir Ağlarında Öğrenme Algoritmaları**

Yapay sinir ağında en önemli hususlardan biri uygun bir öğrenme algoritmasının kullanılmasıdır. Uygun algoritmanın seçilmesi ağın türüne bağlıdır çünkü birçok öğrenme algoritması vardır ancak en çok kullanılan algoritmalar şunlardır:

##### **4.4.1. Geri yayılım algoritması (Back propagation algorithm)**

Sinir ağının geri yayılım metodolojisine göre eğitilmesinin amacı, sinir ağının çıktısı ile model verileri arasında en az hatayı veren optimal ağırlıkları elde etmektir. Bu

ağırlıkların, sinir ağına daha önce öğrenmemiş yeni veriler için tahminleri hesaplamakta kullanılır ve bu algoritma iki şekilde uygulanmaktadır:

### 1. İleri yayılım (Forward Propagation)

İleri yayılım ağı, girdilerini alt düzeyde katmandaki birimlerden alan ve birimlerden oluşan katmanlı bir yapıya sahiptir. İleri yayılma aşamasında, girdiler sinyali gizli katmanın her bir düğümüne yayılır, daha sonra bu sinyal için gizli katmanın her bir düğümü için aktivasyon değeri hesaplanır, sonradan bu düğümler sinyallerini çıktılar katmanının her bir düğümüne gönderir ve ardından Aktivasyon değeri, verilen girdi örneğine ağına yanıtını oluşturmak için çıktılar katmanının her bir düğümü için hesaplanır. Eğitim veya öğrenme aşamasında, çıktı katmanındaki her düğüm, o düğüm için hata değerini belirlemek için hesaplanan aktivasyonu gerçek çıktı değeriyle karşılaştırır, yani ağ ileri beslemeli bir sisteme göre çalışır ve bu aşamadaki bağlantıların ağırlıklarında herhangi bir değişiklik yoktur. Ve çok katmanlı yapay sinir ağını öğretir, girdilerinden çıktılara kadar veri işleme yapılır ve gizli katmanın (m) için (i) düğümündeki girdiler şu ilişki ile tanımlanabilir:

$$n_i^m = b_i^m + \sum_{j=1}^{s^{m-1}} w_{ij}^m a_j^{m-1} \quad (4.6)$$

$$n_i^m = b_i^m + w_{i1}^m a_1^{m-1} + w_{i2}^m a_2^{m-1} + \dots + w_{is^{m-1}}^m a_{s^{m-1}}^{m-1} \quad (4.7)$$

$S^{m-1}$ : Gizli katmandaki (m-1) düğüm sayısı .

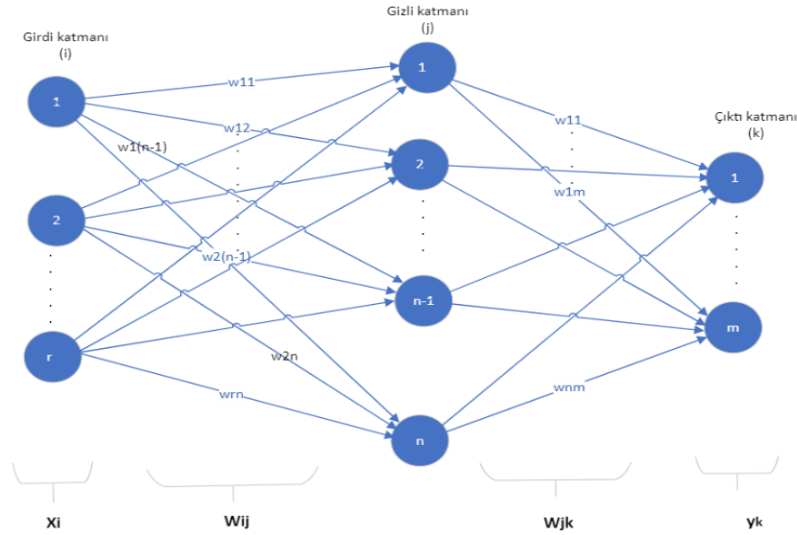
Bir gizli katman ve  $S^1$ e eşit sayıda düğüm varlığında, bu katmanın çıktıları aşağıdaki gibidir:

$$a^1 = f_i^1(n_i^1), i = 1, 2, \dots, s^1 \quad (4.8)$$

Önceki denklemin formülü, ikinci katmana (çıktı katmanı) bir girdidir, ardından modelin formu şöyledir:

$$a^2 = f^2 \left\{ b^2 + \sum_{i=1}^{s^1} w_i^2 f_i^1(b_i^1 + \sum_{j=1}^R w_{ij}^1 a_j^0) \right\} \quad (4.9)$$

$f^1(.)$   $f^2(.)$  : Her ikisi de sırasıyla birinci gizli katman ve ikinci çıktı katmanı için modelde kullanılan aktivasyon fonksiyonlarıdır.



Şekil 4.10. İleri yayılım ağı (Forward propagation)

## 2. Geri yayılım (Back Propagation)

Ağırlıkların güncellenmesi veya değiştirilmesi aşamasında sinyalin çıktı katmanından girdi katmanına ters yönde yeniden yayılmasını sağladığı için ağırlıkların ayarlamak için kullanılan bir algoritmadır ve bu işlem ağ çıktıları istenen sonuçlarla aynı olana kadar tekrarlanır. Geri yayılım, çok katmanlı sinir ağlarını öğrenmek için önemli bir algoritma olup ve yeterlilik ölçüsünü, sinir ağı modelinin ağırlıkları tahmin edilerek azaltılan, Hata kareler Ortalaması (MSE). Hataları azaltmak, sinir ağlarının ve diğer klasik istatistiksel yöntemlerin hedeflerinden biridir. İleri yayılım yöntemi kullanılarak modifiye edilmiş ağırlıkların hesaplanması tamamlandıktan sonra, ileri yayılım yönteminde ulaşılan hata ve ağırlıklara bağlı olarak ağırlıklarının modifikasyonu dikkate alınarak son katmandan giriş katmanına dönerek duyarlılık hesaplanır. Bu süreçten, geri yayılımın anlamı gelmiştir, çünkü katman  $(m + 1)$ 'den  $(m)$  katmana dönerek ağırlıkların ayarlanmasının hesaplanmasındaki ilişkinin tekrarını açıklanmaktadır.

Geri yayılım algoritmasının kullanımının aşağıdaki gibi üç temel aşamayı içerdiği söylenebilir:

- Hatanın ileriye yayılma adımı (Error forward propagation step).

- Hatanın geri yayılım adımı (Error back propagation step).
- Ağ ağırlıkları sentetik adım ( Network weights synthetic step).

Geri yayılım algoritmasının adımları:

1- Başlangıç değerlerinin belirlenmesi.

Ağırlıkların değerlerini, öğrenme parametresini ve örneğin seviyelerinin belirlenmesi.

2- Aktivasyon süreci:

Geri yayılım ağı, girdi  $X_i(t)$  ve istenen çıktı  $Y_d(t)$  uygulanarak etkinleştirir .

3- Gizli katmandaki işlem biriminin gerçek çıktıları şu şekilde hesaplanır:

$$y_j(t) = \text{sigmoid}[\sum_{i=1}^n x_i(t)w_{ij}(t) - \theta_j] \quad (4.10)$$

n: Gizli katmandaki nöron j'ye girdi sayısı.

4- Çıktı katmanındaki işlem birimlerinin gerçek çıktıları aşağıdaki gibi hesaplanır:

$$y_k(t) = \text{sigmoid}[\sum_{j=1}^m x_{jk}(t)w_{jk}(t) - \theta_k] \quad (4.11)$$

m: Çıktı katmanındaki k nöronu için girdi sayısı.

5- İşlem birimleriyle ilişkili hataları geriye doğru dağıtarak ağırlıkları yenilenmesi ve çıktı katmanındaki ve gizli katmandaki işlem birimleri için hataların eğimi aşağıdaki gibi hesaplanır.

I. çıktı katmanı

$$S_k(t) = y_k(t)[1 - y_k(t)]e_k(t) \quad (4.12)$$

$$e_k(t) = y_{d,k}(t) - y_k(t) \quad (4.13)$$

$Y_{d,k}(t)$ : istenen çıktılar.

$Y_k(t)$ : Ağın gerçek çıktıları.

$e_k(t)$ : hata.

Ağırlıkların düzeltilmeleri aşağıdaki gibi hesaplanır:

$$\Delta w_{jk}(t) = \alpha y_j(t)\delta_k(t) \quad (4.14)$$

Çıktı birimlerindeki ağırlıkları aşağıdaki gibi yenilenmesi veya değiştirir.

$$w_{jk}(t + 1) = w_{jk}(t) + \Delta w_{jk}(t) \quad (4.15)$$



## II. Gizli katman

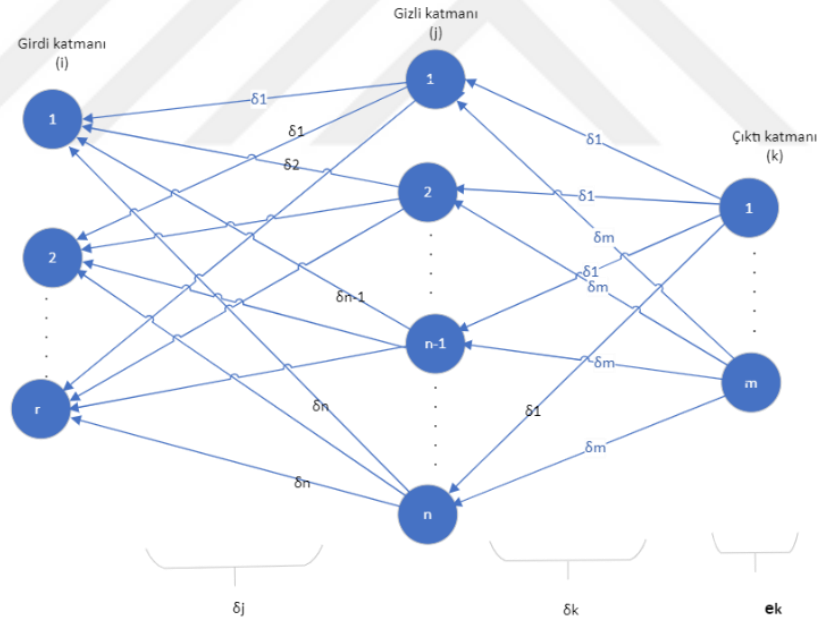
$$s_i(t) = y_i(t)[1 - y_j(t)] \sum_{k=1}^n s_k(t)w_{jk}(t) \quad (4.16)$$

$$\Delta w_{ij}(t) = \alpha x_i \delta_j(t) \quad (4.17)$$

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \quad (4.18)$$

## III. Tekrarlama:

Yinelemeyi T bir tamsayı miktar ile artırır ve ikinci adıma geri dönün ve seçilen hata kriteri gerçekleşmesine kadar işlemi tekrarlanmaktadır. Bu kriterlerin en çok kullanılan, hataların karelerinin toplamını minimize etme kriteridir. Bu kriter, eğitim süresi boyunca bir yolun tamamının hatalarının karelerinin toplamı yeterince küçük olduğunda, ağın yakınsak olarak kabul edildiği (converged) belirtirmektedir.



Şekil 4.11. Geri yayılım ağı (Back propagation)

## 5. GERÇEK VERİ UYGULAMASI VE BULGULAR

Bu bölümde, çalışma verileri incelenecek ve çalışma verilerine tanımlayıcı analiz uygulanacaktır. Ardından, hastaların sağkalım süresini etkileyen en önemli faktörler belirlenip, bu değişkenlerin anlamlılığını ölçmek için Cox regresyon modeli kullanarak veriler analiz edilecektir. Cox regresyon modeli uygulanarak, lösemi hastaları için sağkalım süresini anlamlı ölçüde etkileyen değişkenlerin belirlenmesi, sağkalım verileri için yaşam tablolarının hazırlanması, ayrıca, aynı veriler üzerinde yapay sinir ağıları yöntemi uygulanarak her iki yöntem içinde sonuçlar karşılaştırılacaktır.

### 5.1. Çalışma Verileri

Çalışmada kullanılan veriler, Yemen Cumhuriyetindeki Ulusal Kansere mücadele Vakfı ve Ulusal Kanser mücadele Merkezi 2017 yılı ocak ayı ve 2022 yılı Şubat ayı arasında 1168 lösemi hastasına ait bilgilerdir. Veriler, iki bağımlı değişken ve sekiz bağımsız değişken (risk faktörleri) olmak üzere toplam 10 değişken için hastalara ait bilgileri içermektedir. Bu değişkenler aşağıdaki gibi gösterilmiştir.

#### 5.1.1. Bağımlı Değişkenler

Yaşam (Sağkalım) süresi yada başarısızlık süresi, çalışmada takibe başlanmasından ilgilenilen sonucu gösteren olayın gerçekleşmesine kadar geçen zamandır. Durum, çalışma ilgilenilen sonuçtur. Başarısızlık olarak da adlandırılmaktadır. Ölüm, nüksetme, iyileşme gibi olay olarak adlandırılmaktadır. Sağkalım süresi tam olarak bilinmeyen vakalar ise, sansürlü (censored) olarak adlandırılır.

#### 5.1.2. Bağımsız Değişkenler (Risk Faktörleri)

Risk faktörleri sekiz ana faktör olarak belirlenmiştir: cinsiyet, yaş, kan grubu, medeni durum, il, hastalığın türü, kullanılan tedavi metodu ve tedavi düzenlilik.

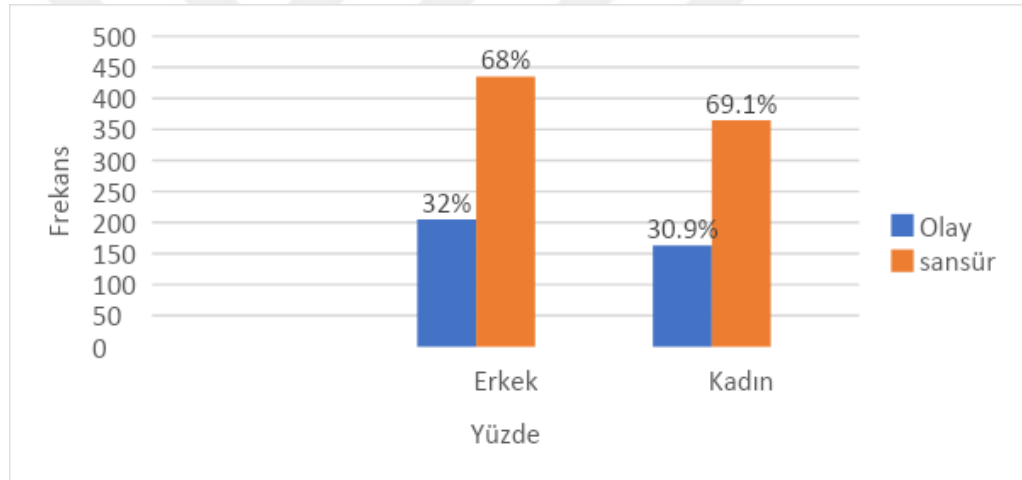
## 5.2. Tanımlayıcı İstatistikler

Tanımlayıcı istatistiklerde yer alan bağımlı değişkenlere göre aşağıdaki gibi veri analiz edilecektir.

### 1- Cinsiyete Göre Hastaların Dağılımı

**Tablo 5.1.** Cinsiyet ve dönem sonundaki durum çapraz tablosu

| Cinsiyet | Olay    |       | Sansür  |       | Toplam  |       |
|----------|---------|-------|---------|-------|---------|-------|
|          | Frekans | Yüzde | Frekans | Yüzde | Frekans | Yüzde |
| Erkek    | 205     | %32   | 435     | %68   | 640     | %54.8 |
| Kadın    | 163     | %30.9 | 365     | %69.1 | 528     | %45.2 |
| Toplam   | 368     | %31.5 | 800     | %68.5 | 1168    | %100  |



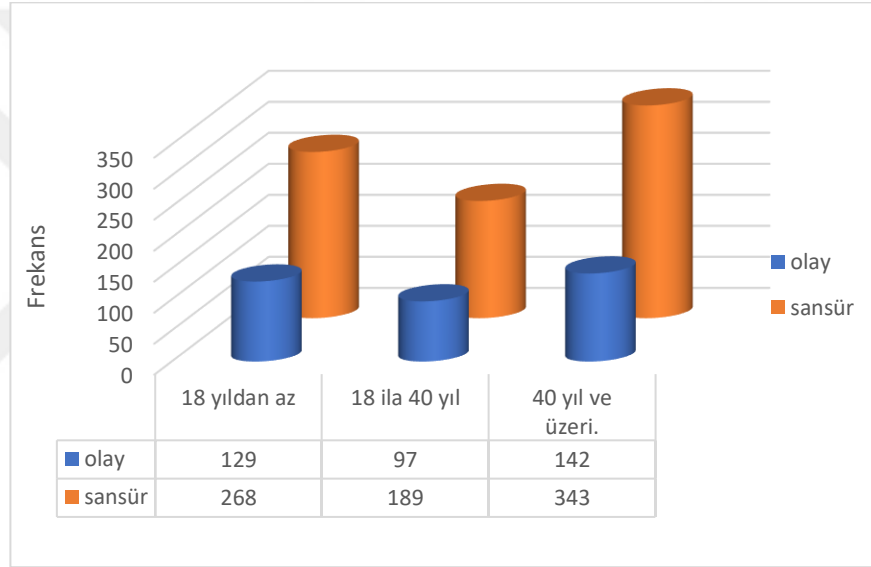
**Şekil 5.1.** Cinsiyet ve dönem sonundaki durum çapraz tablosu

Tablo 5.1 ve Şekil 5.1'e, cinsiyete göre analiz edilecek hastaların %54.8'sini erkek hastalar oluştururken, %45.2'sini kadın hastalar oluşturmaktadır. Yukardaki Tabloda, cinsiyet dağılımına göre bakıldığında erkeklerdeki ölüm oranı %32, kadınlarda ölüm oranı %30.9 olarak bulunmuştur. Erkeklerin ölüm oranı kadınlarından fazladır.

## 2- Yaş Grubuna Göre Hastaların Dağılımı

**Tablo 5.2.** Yaş ve dönem sonundaki durum çapraz tablosu

| Yaş     | Olay    |       | Sansür  |       | Toplam  |       |
|---------|---------|-------|---------|-------|---------|-------|
|         | Frekans | Yüzde | Frekans | Yüzde | Frekans | Yüzde |
| 0-18    | 129     | %32.5 | 268     | %67.5 | 397     | %34   |
| 18 - 40 | 97      | %33.9 | 189     | %66.1 | 286     | %24.5 |
| + 40.   | 142     | %29.3 | 343     | %70.7 | 485     | %41.5 |
| Toplam  | 368     | %31.5 | 800     | %68.5 | 1168    | %100  |



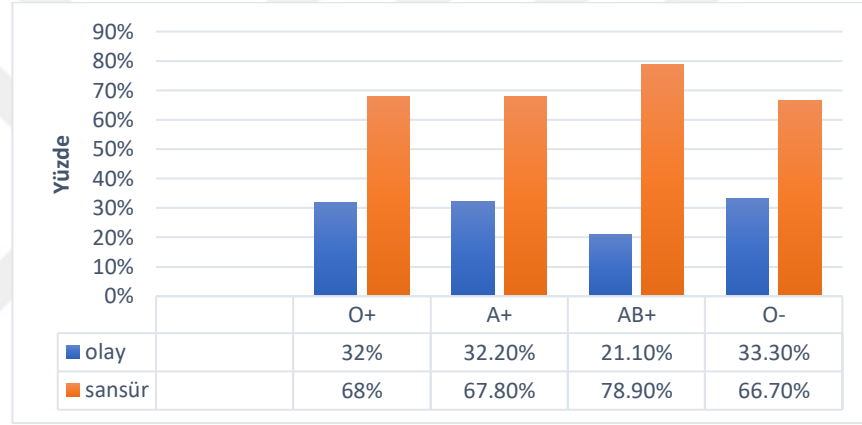
**Şekil 5.2.** Yaş ve dönem sonundaki durum çapraz tablosu

Tablo 5.2 ve Şekil 5.2'ye göre yaş grupları ele alınmıştır ve 40 yaş ve üzeri yaş aralığında hastaların %41.5'nin yer aldığı görülmektedir. Daha sonra 18 yaştan az yaş aralığı ve son olarak 18 ile 40 yaş aralığı olarak sıralanmaktadır. Diğer taraftan 18-40 yaş grubuna bakıldığında, bu yaş grubunda en fazla ölüm oranının (%33.9) gerçekleştiği görülmektedir.

### 3- Kan Grubuna Göre Hastaların Dağılımı

**Tablo 5.3.** Kan grubu ve dönem sonundaki durum çapraz tablosu

| Kan grubu | Olay    |       | Sansür  |       | Toplam  |       |
|-----------|---------|-------|---------|-------|---------|-------|
|           | Frekans | Yüzde | Frekans | Yüzde | Frekans | Yüzde |
| O+        | 239     | %32   | 509     | %68   | 748     | %64   |
| A+        | 109     | %32.2 | 230     | %67.8 | 339     | %29   |
| AB+       | 12      | %21.1 | 45      | %78.9 | 57      | %4.9  |
| O-        | 8       | %33.3 | 16      | %66.7 | 24      | %2.1  |
| Toplam    | 368     | %31.5 | 800     | %68.5 | 1168    | %100  |



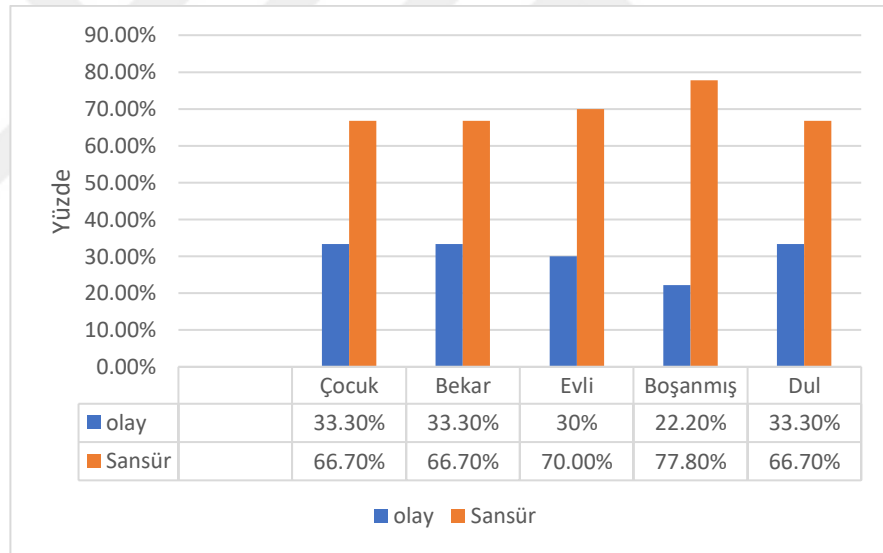
**Şekil 5.3.** Kan grubu ve dönem sonundaki durum çapraz tablosu

Tablo 5.3 ve Şekil 5.3'e göre kan grupları dağılımına bakıldığında hastaların en fazla kan grubu O+ kan grubu ve ikinci sırada ise A+ kan grubudur. Daha sonra sırasıyla AB+ ve son olarak O- kan grubu görülmektedir. Ancak olay (ölüm) oranına bakıldığında en fazla ölüm oranının (%33.3) olduğunu ve O- kan grubunda olduğu görülmektedir.

#### 4- Medeni Durumu Grubuna Göre Hastaların Dağılımı

**Tablo 5.4.** Medeni durumu ve dönem sonundaki durum çapraz tablosu

| Medeni durumu | Olay    |       | Sansür  |       | Toplam  |       |
|---------------|---------|-------|---------|-------|---------|-------|
|               | Frekans | Yüzde | Frekans | Yüzde | Frekans | Yüzde |
| Çocuk         | 120     | %33.3 | 240     | %66.7 | 360     | %30.8 |
| Bekar         | 49      | %33.3 | 98      | %66.7 | 147     | %12.6 |
| Evli          | 186     | %30   | 433     | %70.0 | 619     | %53.0 |
| Boşanmış      | 2       | %22.2 | 7       | %77.8 | 9       | %0.8  |
| Dul           | 11      | %33.3 | 22      | %66.7 | 33      | %2.8  |
| Toplam        | 368     | %31.5 | 800     | %68.5 | 1168    | %100  |



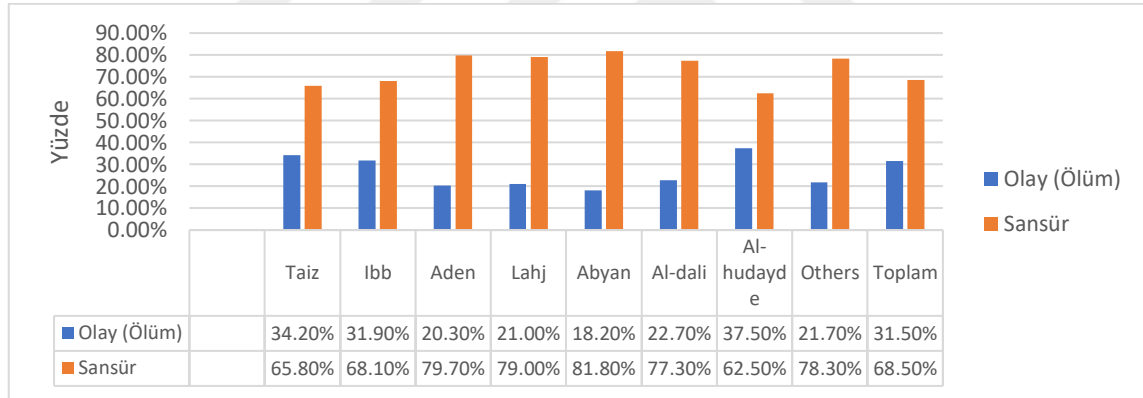
**Şekil 5.4.** Medeni durumu ve dönem sonundaki durum çapraz tablosu

Tablo 5.4 ve Şekil 5.4'te bakıldığında hastaların medeni durumu ve dönem sonundaki durum dağılımına göre %53'ü Evli iken %43.4, si ise Çocuk ve Bekar,%3.6 ise diğer kategoriler arasındadır. Öte yandan, medeni durumuna göre hastalar arasında ölüm oranların %30 ile %33 arasında yakın oranlarda olduğu görülmüştür.

## 5- İkamet Ettiği Şehrine Göre Hastaların Dağılımı

**Tablo 5.5.** İkamet ettiği şehir ve dönem sonundaki durum çapraz tablosu

| İkamet ettiği şehir | Olay    |       | Sansür  |       | Toplam  |       |
|---------------------|---------|-------|---------|-------|---------|-------|
|                     | Frekans | Yüzde | Frekans | Yüzde | Frekans | Yüzde |
| Taiz                | 281     | %34.2 | 541     | %65.8 | 822     | %70.4 |
| Ibb                 | 36      | %31.9 | 77      | %68.1 | 113     | %9.7  |
| Aden                | 14      | %20.3 | 55      | %79.7 | 69      | %5.9  |
| Lahj                | 17      | %21.0 | 64      | %79.0 | 81      | %6.9  |
| Abyan               | 4       | %18.2 | 18      | %81.8 | 22      | %1.9  |
| Al-dali             | 5       | %22.7 | 17      | %77.3 | 22      | %1.9  |
| ALhudayde           | 6       | %37.5 | 10      | %62.5 | 16      | %1.4  |
| Others              | 5       | %21.7 | 18      | %78.3 | 23      | %2.0  |
| Toplam              | 368     | %31.5 | 800     | %68.5 | 1168    | %100  |



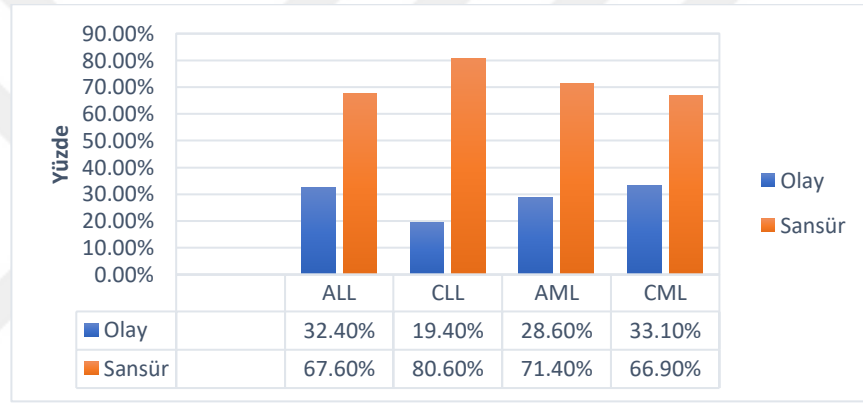
**Şekil 5.5.** İkamet ettiği şehir ve Dönem sonundaki durum Çapraz tablosu

Tablo (5.5) ve Şekil (5.5)'e göre, hastaların yarısından fazlası (%70.4) Taiz ilinde ikamet etmektedir, ardından (%9.7) IBB ilinde ikamet edenler, Lahj ilinde ve diğer illerde oranlar birbirine yakındır. Tablodan da görüldüğü gibi olay (ölüm) oranı toplamda (%31.5) ulaştığı ve hastaların toplam Sansür yüzdesi (%68.5) olarak görülmektedir. Hasta sayılarına göre olay (ölüm) oranı, her bir il için karşılaştırıldığında, ilk sırada Al-hudayde ili (%37.5), ardından Taiz (%34.2), İbb (%31.9) illerinde ve geri kalan illerde ise ölüm oranları birbirine yakın bulunmaktadır.

## 6- Hastalığın Türüne Göre Hastaların Dağılımı

**Tablo 5.6.** Hastalığın türü ve dönem sonundaki durum çapraz tablosu

| Hastalığın türü | Olay    |       | Sansür  |       | Toplam  |       |
|-----------------|---------|-------|---------|-------|---------|-------|
|                 | Frekans | Yüzde | Frekans | Yüzde | Frekans | Yüzde |
| ALL             | 249     | %32.4 | 520     | %67.6 | 769     | %65.8 |
| CLL             | 6       | %19.4 | 25      | %80.6 | 31      | %2.7  |
| AML             | 56      | %28.6 | 140     | %71.4 | 196     | %16.8 |
| CML             | 57      | %33.1 | 115     | %66.9 | 172     | %14.7 |
| Toplam          | 368     | %31.5 | 800     | %68.5 | 1168    | %100  |



**Şekil 5.6.** Hastalığın türü ve dönem sonundaki durum çapraz tablosu

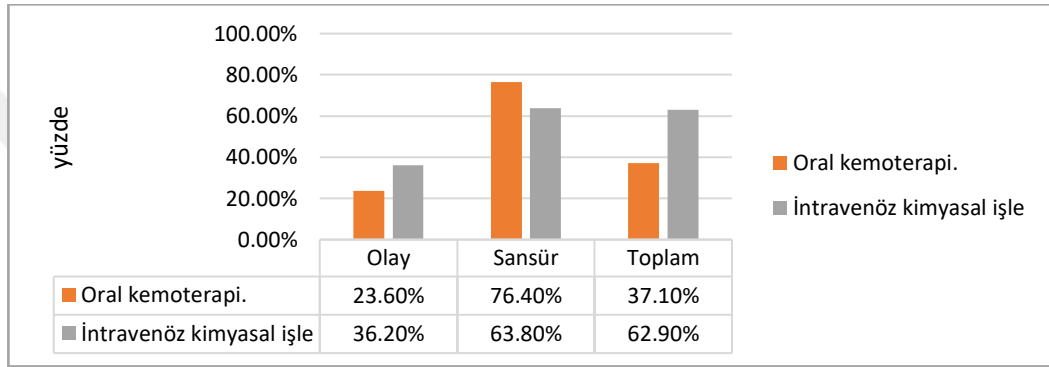
Tablo (5.6) ve Şekil (5.6)'ya göre, en sık görülen tip %65.8 ile Akut lenfositik lösemi(Acute Lymphoblastic Leukemia) (ALL) iken, bunu %16.8 ile Akut miyeloid lösemi (Acute Myeloid Leukemia) (AML) hastaları izlemektedir. Kronik miyeloid lösemi (Chronic Myeloid Leukemia) (CML) hastaların yüzdesi (%14.7) ve Kronik lenfositik lösemi (Chronic Lymphocytic Leukemia)(CLL) hastaların yüzdesi (%2.7)'dir. Son olarak, tüm hastalarda en yüksek ölüm oranı CML hastaları iken (%33.1), akut lenfositik lösemi (ALL) ölüm oranı (%32.4), ardından AML (%28.6) ve son olarak CLL' den (%19.4) olarak görülmektedir.



## 7- Kullanılan Tedavi Metoduna Göre Hastaların Dağılımı

**Tablo 5.7.** *Kullanılan tedavi metodu ve dönem sonundaki durum çapraz tablosu*

| Tedavi metodu            | Olay    |       | Sansür  |       | Toplam  |       |
|--------------------------|---------|-------|---------|-------|---------|-------|
|                          | Frekans | Yüzde | Frekans | Yüzde | Frekans | Yüzde |
| Oral kemoterapi.         | 102     | %23.6 | 331     | %76.4 | 433     | %37.1 |
| İntravenöz kimyasal işle | 266     | %36.2 | 469     | %63.8 | 735     | %62.9 |
| Toplam                   | 368     | %31.5 | 800     | %68.5 | 1168    | %100  |



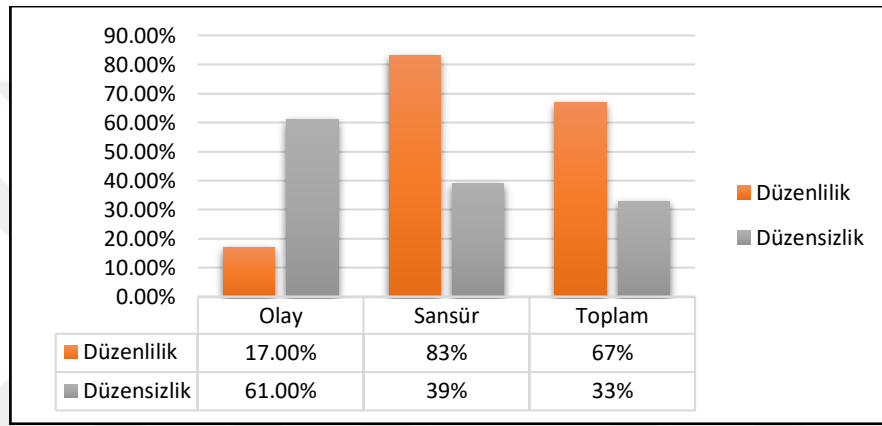
**Şekil 5.7.** *Kullanılan tedavi metodu ve dönem sonundaki durum.*

Tablo (5.7) ve Şekil (5.7)'ye göre intravenöz kemoterapi kullanılan tedavi metodu alan hastaların oranı (%62.9) iken, geri kalanı hastalar (%37.1) oral kemoterapi alanlardır. Aynı zamanda, intravenöz kemoterapi alanlarının ölüm oranı (%36.2) iken, oral kemoterapi alanların ölüm oranı ise (%23.6) olarak görülmektedir.

## 8- Tedavi Düzenliliğine Göre Hastaların Dağılımı

**Tablo 5.8.** Tedavi düzenlilik ve dönem sonundaki durum çapraz tablosu

| Tedavi düzenlilik | Olay    |       | Sansür  |       | Toplam  |       |
|-------------------|---------|-------|---------|-------|---------|-------|
|                   | Frekans | Yüzde | Frekans | Yüzde | Frekans | Yüzde |
| Düzenlilik        | 133     | %17.0 | 650     | %83   | 783     | %67   |
| Düzensizlik       | 235     | %61.0 | 150     | %39   | 385     | %33   |
| Toplam            | 368     | %31.5 | 800     | %68.5 | 1168    | %100  |



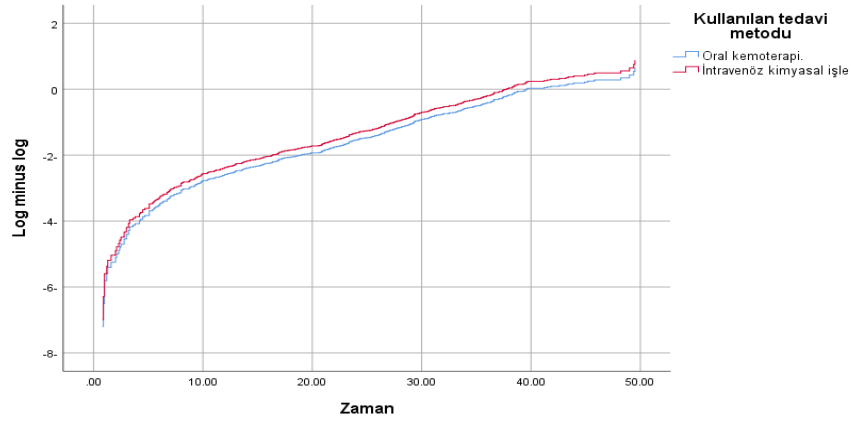
**Şekil 5.8.** Tedavi düzenliliği

Tablo (5.8) ve Şekil (5.8)'e göre bu çalışmadaki hastaların üçte ikisinden fazlası (%67) ile düzenli bir şekilde tedavi görürken, (%33)'nün düzensiz tedavi görmektedir. Ayrıca, düzenli olarak tedavi olan hastaların ölüm oranı (%17) iken, düzensiz şekilde tedavi gören hastaların ölüm oranı (%61)'dir.

### 5.3. Göreceli Risk Varsayım Testi

Cox regresyon modelini tahmin etmedeki ilk adım risk oranının (risk görecelisi) zaman içinde bir kişiden diğerine sabit olduğunu varsayan göreceli risk varsayımını test etmektir. Bu bağlamda, göreceli risk varsayımı iki şekilde test edilebilmektedir.

### 5.3.1. Grafik Yöntemi



Şekil 5.9. Grafik kullanarak göreceli risk varsayımı testi

Göreceli risk varsayım, ikili değişkenlerden biri zaman içinde Kaplan-meier yöntemi kullanılarak ve çizilerek test edilir. Grafik, risk oranlarının iki bölüm için benzer bir davranışa sahip olduğunu göstermektedir. Şekil (5.9)'da görüldüğü gibi hastanın sağkalım süresi boyunca ve göreceli varsayımın doğrulandığı gözlenmektedir. Bu durumda, Cox modeline göre bu çalışmanın verilerinde göreceli risk varsayımının sağlandığını göstermektedir.

### 5.3.2. Schoenfeld Artıklar (residuals) Testi

Schoenfeld artıkları testi, artıklar ve sağkalım arasındaki bağımsızlığı test etmek de ve Cox modelindeki göreceli tehlike varsayımını test etmek için de kullanılmaktadır. Cox orantılı tehlikeler modelindeki temel varsayımlardan biri sayılmaktadır. Göreceli riskin varsayımı, Schoenfeld artıkları test edilmesiyle ve Schoenfeld artıkları ile sağkalım süreleri arasındaki korelasyonun sıfır olduğunu belirten sıfır hipotezine göre sayısal yöntemle de doğrulanır. Doğrusu, bu hipotezin reddedilmesi, göreceli risk koşulunu karşılamada başarısızlık olarak değerlendirilir:

$$H_0: \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = 0$$

**Tablo 5.9.** Schoenfeld artıkları görel risk testinin sonuçları

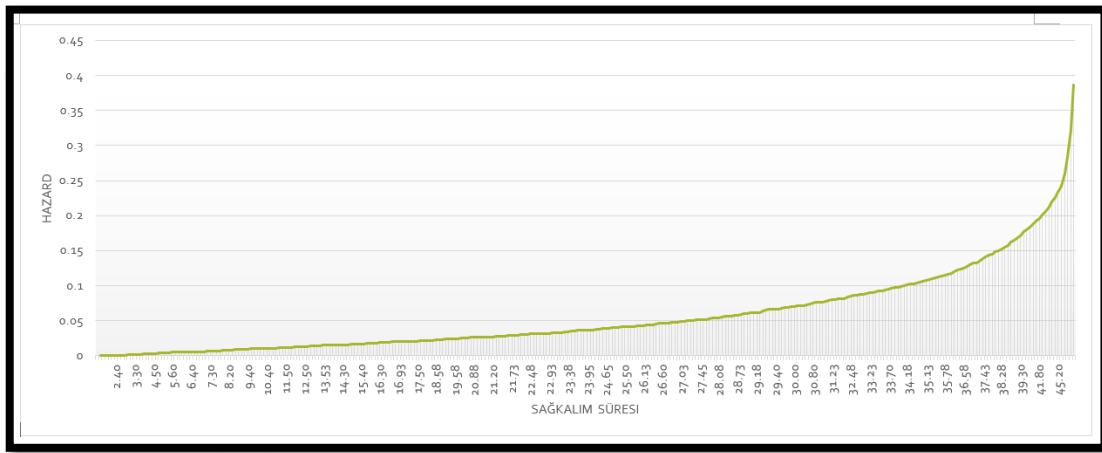
| Değişken                                      | Ki-kare değeri | D.F | Sig          |
|-----------------------------------------------|----------------|-----|--------------|
| Kan grubu için kısmi artıklar.                | 0.829          | 1   | 0.362        |
| Cinsiyet için kısmi artıklar.                 | 0.022          | 1   | 0.882        |
| Medeni durumu için kısmi artıklar.            | 0.593          | 1   | 0.441        |
| İkamet ettiği şehir için kısmi artıklar.      | 0.142          | 1   | 0.706        |
| Hastalığın türü için kısmi artıklar.          | 0.193          | 1   | 0.660        |
| Kullanılan tedavi metodu için kısmi artıklar. | 0.000          | 1   | 0.991        |
| Tedavi düzenlilik için kısmi artıklar..       | 3.630          | 1   | 0.057        |
| Yaş için kısmi artıklar.                      | 0.612          | 1   | 0.434        |
| Genel Model                                   | 5.209          | 8   | <b>0.735</b> |

Tablo (5.9)'a bakıldığında, 8 değişken içeren genel model için hesaplanan Ki- Kare değeri 8 serbestlik derecesinde (5.209) olarak elde edilmiştir ve Ki Kare tablo değerinden (15.51) daha küçük olduğu için, anlamlı olmadığı görülmektedir. Bu nedenle Schoenfeld hataları ve sağkalım süreleri arasındaki tüm korelasyon katsayılarının sıfır olduğunu varsayan yokluk hipotezi kabul edilmektedir. Aynı şekilde, her bir değişken ayrı ayrı incelendiğinde, her bir değişken için (1) serbestlik derecesinde hesaplanan Ki- Kare değerinin Ki-Kare tablo değerinden (3.84) küçük olduğu ve ayrıca, tüm değişkenler için ( $p>0.05$ ) olduğu bulunmaktadır. Bu, Schoenfeld hataları ile her bir değişken için ayrı ayrı sağkalım süreleri arasındaki korelasyonun sıfıra eşit olduğu anlamına gelmektedir. Ve bu sonuç, tüm değişkenler için nispi riskin varsayılmasının başarıldığı anlamına gelmektedir. Bu sonuçlar da Şekil (5.9)'da verilen grafik yöntemiyle de tutarlı sonuçları vermektedir.

#### 5.4. Cox Modelinin Temel Fonksiyonlarının Tahmin Edilmesi

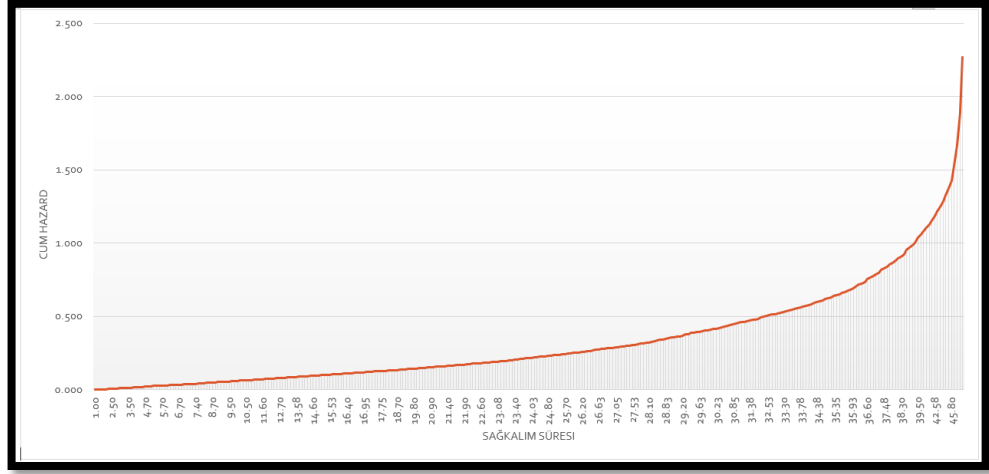
Göreceli risk varsayımı gerçekleştirildikten sonra, Cox regresyon modelinde kullanılan temel fonksiyonlar, temel riskin fonksiyonu (Baseline Hazard)  $h_0(t)$ , temel sağkalım fonksiyonu (Baseline Survival)  $S_0(t)$  ve temel toplam risk fonksiyonu (Baseline Cum Hazard)  $H_0(t)$  tahmin edilmiştir. 50 ay boyunca kontrol edilen tüm hastalar ait veriler SPSS programı kullanarak analiz edilmiş ve Şekil (5.10)'de Lösemili hastalara ait Temel Hazard Fonksiyonunda gösterildiği gibi sonuçlara ulaşılmıştır. Cox regresyon modelinin fonksiyonlarını içeren ilk aylardaki risk oranının sıfır olduğu, yani ilk aylarda lösemi

hastaları için temel riskin neredeyse olmadığı anlamına gelmektedir. Daha sonra bu risk üçüncü aydan itibaren ortaya çıkmaya başlamakta, yirmi dördüncü aya ulaşana kadar giderek artmakta ve ardından, başlangıçta olduğundan çok daha fazla artmaya başlamaktadır. 34. ayda temel risk oranı %10'a ulaşmıştır. Bununla beraber, kırk beşinci ayda risk oranının %24'e daha giderek artmakta, çalışmanın son aylarında risk daha hızla artarken, temel risk oranının %37'ye ulaştığı, kırk dokuzuncu ayda ise riskin daha da arttığı görülmektedir. Şekil 5.10'dan da görüldüğü gibi, dikey eksen risk olasılıklarını temsil ederken, yatay eksen ise sağkalım süresini temsil etmektedir.



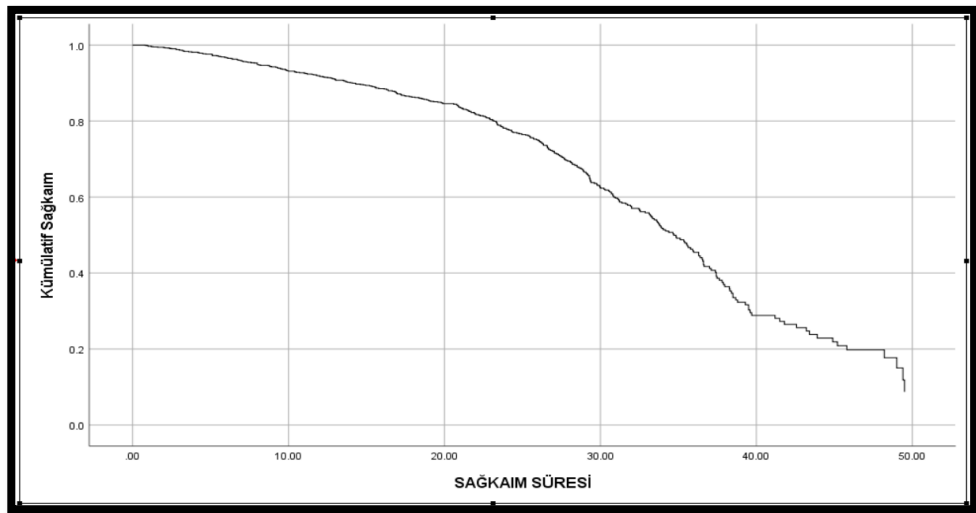
**Şekil 5.10.** Lösemili hastalara temel hazard fonksiyonu

Kümülatif risk fonksiyonu (Cum Hazard Fonksiyonu) Şekil (5.11)'de gösterildiği gibi, zamanın başlangıcında yavaş bir artış göstermektedir ve bu artış yaklaşık olarak (0.50) ulaşmıştır. 32.5 aya kadar ise düşük bir seviyede devam etmektedir. Daha sonra, (38) ayda kümülatif risk fonksiyonu (.985)'e ulaşmıştır. Otuz dokuzuncu ayda de ise, (1.003) ulaşınca kadar yükselmeye devam etmiştir. Ayrıca, (39) aydan fazla hayatta kalanlar diğerlerine göre ölüm riskiyle daha fazla karşı karşıya kalırken, risk (2.27)'ye ulaşana kadar büyük oranlarda artmıştır .



Şekil 5.11. Cum hazard fonksiyonu

Şekil 5.12'e bakıldığında, temel sağkalım fonksiyonu (Survival Function) zamanın ilerlemesiyle yavaş yavaş azalmaktadır. Hastaların sağkalım olasılığı eğrisi, (29) aydaki sağkalım olasılığı (0.70)'e ulaşmıştır. Bundan sonra, (40) aydan fazla kalan sağkalım az sayıda hasta olduğu için eğri daha az düzgün hale geldiğinden, sağkalım olasılığındaki düşüş oldukça hızlı bir şekilde başlanmaktadır. Medyan sağkalım süresi (36) aya, ki bu da lösemili hastaların yarısının hayatta kalamadığı (ölümle karşı karşıya kaldığı) noktadır, o zaman eğri önemli ölçüde düşmeye devam etmektedir. Burada sağkalım olasılığı, 49.ayda,(0.03) eşittir.



Şekil 5.12. Sağkalım fonksiyonu (survival)eğrisi

### 5.5. Cox Regresyon Modelinin Parametrelerinin Tahmin Edilmesi

Göreceli riski varsayımı sağlandıktan sonra, temel fonksiyonların tahmini ve bağımlı değişken için parametre tahmini gerçekleştirilecektir. Bağımlı değişkenin ikili olduğu yarı parametrik cox regresyon modelinde, cox modeli parametreleri kısmi olasılık (Partial likelihood) kullanılarak tahmin edilmektedir. Bu çalışmada, Lösemili hastalar için risk fonksiyonunun, tüm bağımsız değişkenler ile ilişkili olduğu ve sağkalım süresi üzerinde etkili olduğu varsayılmaktadır.

**Tablo 5.10.** Cox regresyon modelinin parametre tahminleri

| Değişkenler              | B     | Wald    | SE    | D.F | Sig   |
|--------------------------|-------|---------|-------|-----|-------|
| Kan Grubu                | 0.051 | 0.401   | 0.081 | 1   | 0.527 |
| Cinsiyet                 | 0.016 | 0.022   | 0.106 | 1   | 0.882 |
| Medeni Durumu            | 0.173 | 4.115   | 0.085 | 1   | 0.042 |
| İkamet ettiği şehir (il) | 0.102 | 8.363   | 0.035 | 1   | 0.004 |
| Hastalığın Türü.         | 0.193 | 9.404   | 0.063 | 1   | 0.002 |
| K.Tedavi Metodu          | 0.453 | 7.859   | 0.162 | 1   | 0.005 |
| Tedavi Düzenlilik.       | 1.127 | 106.600 | 0.109 | 1   | 0.005 |
| Yaş                      | -.216 | 4.599   | 0.101 | 1   | 0.032 |

Wald istatistiği bağımsız değişkenlere ilişkin katsayıların anlamlı olup olmadığını belirlemektedir. Tablo 5.10’da Wald istatistiğinin anlamlılık (Sig) değerine göre ( $0.527 > 0.05$ ) olduğundan dolayı kan gruplarının hastaların sağkalım süresine etkisi anlamlı değildir. Aynı şekilde cinsiyet değişkeninde anlamlılık değeri ( $0.882 > 0.05$ ) olduğundan dolayı anlamlı risk etkisinin olmadığı bulunmuştur.

Medeni Durumu, İkamet ettiği şehir (il), Hastalığın türü, Kullanılan tedavi metodu, Tedavi düzenlilik ve Yaş değişkenlerinde anlamlılık değerleri sırasıyla önemlilik düzeyinde anlamlı bulunmuştur.

Diğer bir deyişle, Kan grubu ve Cinsiyet değişkenleri dışında tüm diğer bağımsız değişkenlerin, (Medeni Durumu, İkamet ettiği şehir (il), Hastalığın türü, Kullanılan tedavi

metodu, Tedavi düzenlilik ve Yaş), (0.05) anlam düzeyinde sağkalım süresi üzerinde anlamlı ve önemi bir etkiye sahip olduğu görülmektedir.

### 5.6. Modele Dahil Edilen Değişkenlerin Anlamlılığını Testi

Modelin bağımsız değişkenlerinin parametrelerini tahmin ettikten sonra, değişkenler teker teker incelenir, genel modele anlamlı katkıda bulunmayanlar sırayla elenir ve bu genel modelin aşamalı olarak geriye (Stepwise Backward) doğru silinmesiyle yapılmaktadır. (Wald) testi ile sağkalım süresinde anlamlı ve etkili değişkenlere ulaşana kadar en az anlamlı değişken aşamalı olarak silinir. Hesaplanan (Wald) değerleri kullanarak her değişken için (1) serbestlik derecesinde, ve (0.05) anlamlılık değerinde tablo değerleri ile aşağıdaki gibi karşılaştırılmaktadır:

**Tablo 5.11.** Denklemdaki değişkenler

| Adım    | Değişkenler      | B      | SE    | Wald    | df | Sig.  | Exp(B) |
|---------|------------------|--------|-------|---------|----|-------|--------|
| 1. Adım | Kan grubu        | 0.051  | 0.081 | 0.401   | 1  | 0.527 | 1.052  |
|         | Cinsiyet         | 0.016  | 0.106 | 0.022   | 1  | 0.882 | 1.016  |
|         | M.durumu         | 0.173  | 0.085 | 4.115   | 1  | 0.042 | 1.188  |
|         | K.şehir (İl)     | 0.102  | 0.035 | 8.363   | 1  | 0.004 | 1.107  |
|         | H.türü.          | 0.193  | 0.063 | 9.404   | 1  | 0.002 | 1.213  |
|         | Tedavi metodu    | 0.453  | 0.162 | 7.859   | 1  | 0.005 | 1.573  |
|         | T. düzenlilik.   | 1.127  | 0.109 | 106.600 | 1  | 0.000 | 3.085  |
|         | Yaş              | -.216- | 0.101 | 4.599   | 1  | 0.032 | 0.805  |
| 2. Adım | Kan grubu        | 0.051  | 0.081 | .400    | 1  | 0.527 | 1.052  |
|         | M.durumu         | 0.174  | 0.085 | 4.185   | 1  | 0.041 | 1.190  |
|         | K.şehir (İl)     | 0.101  | 0.035 | 8.344   | 1  | 0.004 | 1.107  |
|         | H.türü.          | 0.194  | 0.063 | 9.461   | 1  | 0.002 | 1.214  |
|         | Tedavi metodu    | 0.455  | 0.161 | 7.954   | 1  | 0.005 | 1.576  |
|         | T.düzenlilik.    | 1.126  | 0.109 | 106.751 | 1  | 0.000 | 3.082  |
|         | Yaş              | -.217- | 0.101 | 4.651   | 1  | 0.031 | .805   |
| 3. Adım | M. durumu        | 0.171  | 0.085 | 4.062   | 1  | 0.044 | 1.187  |
|         | K.şehir (İl)     | 0.102  | 0.035 | 8.399   | 1  | 0.004 | 1.107  |
|         | Hastalığın türü. | 0.193  | 0.063 | 9.385   | 1  | 0.002 | 1.213  |
|         | Tedavi Metodu    | 0.456  | 0.162 | 7.972   | 1  | 0.005 | 1.578  |
|         | T.düzenlilik.    | 1.122  | 0.109 | 106.402 | 1  | 0.000 | 3.072  |
|         | Yaş              | -.214- | 0.101 | 4.516   | 1  | 0.034 | 0.807  |



İlk adımda ve Tablo (5.11)'e göre, hesaplanan Wald değerleri, (0.05) anlamlılık değeri, Ki-Kare tablo değerleriyle karşılaştırmıştır. Bu durumda Medeni Durum, İkamet ettiği şehir (il), Hastalığın türü, Kullanılan tedavi metodu, Tedavi düzenlilik ve Yaş değişkenleri anlamlı değişkenler olarak gözlenmiştir. Ancak diğer değişkenlerin anlamlı ve önemli bir etkiye sahip olmadığı görülmektedir.

Dolayısıyla, bu adım da cinsiyet değişkeninin sağkalım süresi üzerine anlamlı bir etkisi olmadığından be sebeple bu adımda modelden silinmiştir.

İkinci adımda; Medeni durum, İkamet ettiği şehir (il), Hastalığın türü, Kullanılan tedavi metodu, Tedavide düzenlilik ve Yaş değişkenleri sağkalım süresini etkileyen önemli değişkenler olduğun görülmektedir. Modelden silinecek olan, tek değişken, kan grubu değişkenidir.

Üçüncü ve son adımda; Medeni Durumu, İkamet ettiği şehir (il), Hastalığın türü, Kullanılan tedavi metodu, Tedavide düzenlilik ve Yaş değişkenleri anlamlı etkiye sahip değişkenler oldukları için modelde yer alacaklardır. Yani, bu değişkenler sağkalım süresi üzerinde etkilidir ve modelden çıkarılan diğer değişkenler ise, sağkalım süresi üzerinde hiçbir etkisi yoktur. Böylece, bütün olarak modelden indirgenen model, son adımda önemini koruyan değişkenler Medeni Durumu, İkamet ettiği şehir (İl), Hastalığın Türü, Kullanılan Tedavi Metodu, Tedavide Düzenlilik ve Yaş değişkenlerini içermektedir. Böylece, Cox regresyon modelinin indirgenmiş olan modeli aşağıdaki gibidir.

- Risk Fonksiyonu (Hazard Function)

$$h(t) = h_0(t) \exp(0.171M.d + 0.102K.S + 0.193H.t + 0.456T.m + 1.122T.d - 0.214Yaş)$$

- Sağkalım Fonksiyonu(Survival Function)

$$S(t) = [S_0(t)]^{\exp(0.171M.d+0.102K.S+0.193H.t+0.456T.m+1.122T.d-0.214Yaş)}$$

Ayrıca, Tablo (5.11)'e göre, son sütundaki değerler (Exp) tahmin edilen risk oranını temsil etmektedir. Değer 1'den büyükken riskin yüksek, 1'den küçük ise, riskin düşük olduğu anlamına gelmektedir.

Yukarıdakilere ek olarak, anlamlı bulunan değişkenler (Medeni Durumu, İkamet ettiği şehir (İl), Hastalığın Türü, Kullanılan Tedavi Metodu, Tedavide Düzenlilik ve Yaş) içeren üçüncü ve son adım için Tablo (5.11)'deki risk oranı açıklanacaktır. Hastalık türü değişkeni için, (Exp(B)) değeri (1.213) olduğundan, geri kalan değişkenler sabit kaldığı

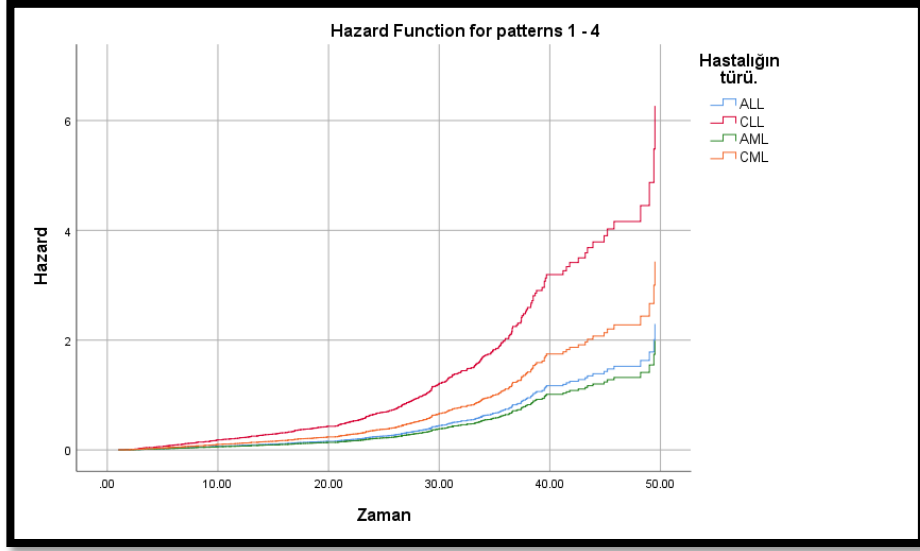
sürece, hastalık türündeki her bir değişiklik için risk oranının (%21,3)' arttığı anlamına gelmektedir. Sağkalım progresyonunun hastalığın tipine göre farklılık gösterip göstermediğini bulmak için, referans hücre olarak en düşük seviye kategorisi olan akut lenfoblastik löseminin (ALL) seçildiği ve diğer kategorilerle karşılaştırıldığı referans hücre kodlama yöntemini, aşağıdaki gibi kullanılacaktır:

**Tablo 5.12.** Referans hücre kullanılarak hastalık türü için hazard oranı testi

| Hastalığın Türü | B      | WALD  | Sig.  | EXP(B) |
|-----------------|--------|-------|-------|--------|
| CLL             | 1.005  | 5.789 | 0.016 | 2.731  |
| AML             | -.143- | 0.926 | 0.336 | 0.867  |
| CML             | 0.402  | 7.367 | 0.007 | 1.495  |

Tablo (5.12) ve Şekil (5.13)'göre, (AML) hastalığı türü için, Wald testinin değeri anlamlı bulunmamıştır. Yani ( $p>0.05$ ) olması nedeniyle, (AML) ve (ALL) hastalığının türleri arasında hastalar için Hazard oranı açısından anlamlı bir fark yoktur. Hem (CLL) hem de (CML) hastalığı olan hastalar için ise Wald testi değerinin anlamlı olduğu, yani (CLL) ve (CML) tipleri ile hastalar arasında, (ALL) tipi olan hastalarla karşılaştırıldığında risk oranları arasında önemli farklılıklar olduğu gösterilmektedir.

Exp(B) değerine bakıldığında, tip (AML)'i olan hastalar için tahmini risk oranının (0.867) olup, tip (AML)'i olan hastalarda ölüm oranı, tip (ALL) olan hastalara göre (%12.4) daha düşüktür. Ayrıca, tip (CML)'i hastaları, tip (ALL)'e hastalarına göre (%49.5) kat daha fazla ölürken, tip (CLL) olan hastaların ise, tip (ALL)'e hastalarına göre, (1.731) kat daha fazla bulunmuştur. Bu sonuçlar, modeldeki iki değişkenin önemini (CML ve CLL), yani üç hastalık türünden ikisinin, (ALL) hastalık türünden anlamlı ölçüde farklı bir risk oranına sahip olduğunu göstermektedir.



Şekil 5.13. Referans hücre kullanılarak hastalık türü için hazard oranı

### 5.7. En İyi Modelin Belirlenmesi

Modele dahil edilen değişkenlerin anlamlılığı test edilirken, birkaç model elde edilir, tüm modeller arasında risk fonksiyonunu tahmin etmede en iyi model, Ki-kare dağılımına göre, AIC ve Olabilirlik Oranı testi ile şu şekilde belirlenebilir:

Tablo 5.13. En iyi modeli seçmek için Olabilirlik Oranı testi sonuçları

| Adım.No                     | -2 Log Likelihood | Hesaplanan Ki-Kare ( $\chi^2$ ) | Tabo'dai Ki-kare ( $\chi^2$ ) | D.f | Sig    |
|-----------------------------|-------------------|---------------------------------|-------------------------------|-----|--------|
| -2 Log Likelihood= 4395.454 |                   |                                 |                               |     |        |
| 1.Adım                      | 4256.558          | 138.896                         | 15.507                        | 8   | 0.0004 |
| 2.Adım                      | 4256.580          | 138.874                         | 14.067                        | 7   | 0.000  |
| 3.Adım                      | 4256.972          | 138.481                         | 12.592                        | 6   | 0.000  |

Tablo (5.13)'e göre tüm değişkenleri içeren ilk model için, ( $0.0004 < 0.005$ ) olduğundan diğer tüm modellerden daha az anlamlı olduğunu ve hesaplanan ki-kare değerinin ki-kare tablo değerinden daha büyük olduğu söylenebilir. Aynı şekilde ikinci ve üçüncü modeller için de ( $p < 0.05$ ) ve hesaplanan ki-kare değerinde ki-kare tablo değerinden büyük olarak bulunmuştur. Bununla birlikte (Medeni Durumu, İkamet ettiği şehir (il), Hastalığın türü, Kullanılan tedavi metodu Tedavi düzenlilik ve Yaş)

değişkenleri içeren üçüncü model için anlamlılık (0.0001) olduğundan diğer tüm modeller arasında en anlamlı model olduğu görülmektedir.

**Tablo 5.14.** AIC ve Olabilirlik oran testi

| Adım.No | AIC      | Olabilirlik Oran Testi |
|---------|----------|------------------------|
| 1.Adım  | 8806.908 | 277.792                |
| 2.Adım  | 8804.908 | 277.748                |
| 3.Adım  | 8802.908 | 276.964                |

Tablo (5.14)'e göre AIC ve Olabilirlik Oran Testi için en düşük değer 3.Adım modelinde bulunmuştur. Böylece, Medeni Durumu, İkamet ettiği şehir (il), Hastalığın türü, Kullanılan tedavi metodu, Tedavide düzenlilik ve Yaş değişkenlerini içeren üçüncü modelin diğer tüm modeller arasında en iyi model olduğu görülmektedir.

## 5.8. Sağkalıma Süreleri Analizi

Yaşam tablosunun denklemlerine dayanarak, sağkalım sürelerinin her biri üçer aylık eşit sürelerle bölünerek ve SPSS programı kullanılarak aşağıdaki sonuçlar elde edilmiştir:

**Tablo 5.15.** Lösemili hastalar için sağkalım verilerinin yaşam tablosu

| Zaman Aralığı | Giren sayısı | Gözlemlenen sayısını | Riske maruz kalan sayı | Olayların Sayısı | Ölenlerin yüzdesini | Sağkalım oranı | Dönemin sonunda sağkalım olasılığını | Tehlike oranı |
|---------------|--------------|----------------------|------------------------|------------------|---------------------|----------------|--------------------------------------|---------------|
| 0             | 1168         | 31                   | 1152.5                 | 11               | 0.01                | 0.99           | 0.99                                 | 0.00          |
| 3             | 1126         | 63                   | 1094.5                 | 26               | 0.02                | 0.98           | 0.97                                 | 0.01          |
| 6             | 1037         | 78                   | 998.00                 | 26               | 0.03                | 0.97           | 0.94                                 | 0.01          |
| 9             | 933          | 56                   | 905.00                 | 22               | 0.02                | 0.98           | 0.92                                 | 0.01          |
| 12            | 855          | 49                   | 830.50                 | 23               | 0.03                | 0.97           | 0.89                                 | 0.01          |
| 15            | 783          | 31                   | 767.50                 | 26               | 0.03                | 0.97           | 0.86                                 | 0.01          |
| 18            | 726          | 67                   | 692.50                 | 21               | 0.03                | 0.97           | 0.84                                 | 0.01          |
| 21            | 638          | 99                   | 588.50                 | 39               | 0.07                | 0.93           | 0.78                                 | 0.02          |
| 24            | 500          | 88                   | 456.00                 | 33               | 0.07                | 0.93           | 0.72                                 | 0.03          |
| 27            | 379          | 75                   | 341.50                 | 43               | 0.13                | 0.87           | 0.63                                 | 0.04          |
| 30            | 261          | 41                   | 240.50                 | 25               | 0.10                | 0.90           | 0.57                                 | 0.04          |
| 33            | 195          | 55                   | 167.50                 | 29               | 0.17                | 0.83           | 0.47                                 | 0.06          |
| 36            | 111          | 32                   | 95.000                 | 25               | 0.26                | 0.74           | 0.35                                 | 0.10          |
| 39            | 54           | 10                   | 49.000                 | 8                | 0.16                | 0.84           | 0.29                                 | 0.06          |
| 42            | 36           | 7                    | 32.500                 | 5                | 0.15                | 0.85           | 0.24                                 | 0.06          |
| 45            | 24           | 11                   | 18.500                 | 2                | 0.11                | 0.89           | 0.22                                 | 0.04          |
| 48            | 11           | 7                    | 7.500                  | 4                | 0.53                | 0.47           | 0.10                                 | 0.00          |

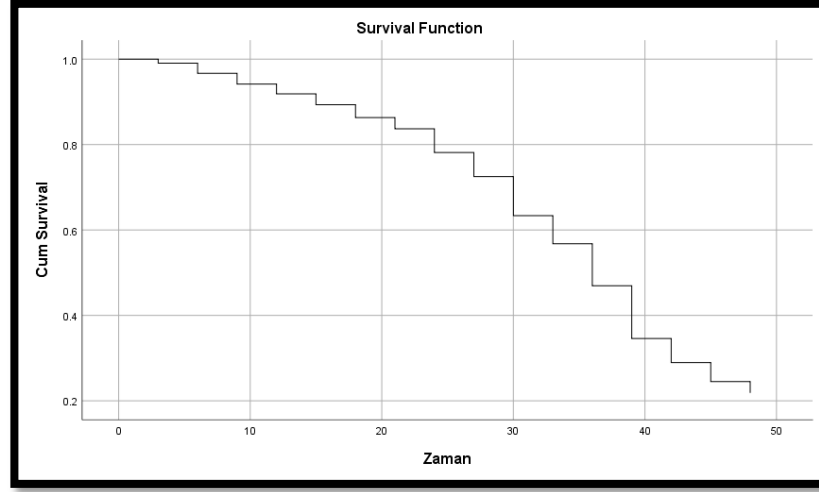
The median survival time is 35.06

Tablo (5.15), 9 sütundan oluşmakta, ve her satır sağkalım bir süre olarak sayılarak, her periyot ilk periyodun başlangıç zamanı sıfırdan başlayarak, üç aya bölünmektedir. Üçüncü aya kadar uzanan süreyle, ikinci süreye ise, üçüncü aydan başlayarak altı aydan az olan ve böylece geri kalanı, aynı şekilde devam edilen süreler 49 aya kadar olan süreyi kapsamaktadır.

İkinci sütun, her döneme giren kişi sayısını temsil eder ve cari sürenin başına kadar hala sağkalım olanların sayısını gösterir. Burada, sürenin başında girenlerin sayısı 1168 olup, ikinci dönemdeyse altı aydan az, yani  $n_2 = 1168 - (31 + 11) = 1126$ , sayıları 1126 kişidir. Üçüncü sütun, gözlemlenen kişi sayısını temsil ederken, dördüncü sütun, gerçek örnek büyüklüğünü, yani zaman süresine girerken, kaç vakanın gözlemlendiğini tahmin etmektedir. Örneğin üçüncü ayın başında giren hasta sayısı şu şekildeki gibi hesaplanır  $(1126 - \left(\frac{63}{2}\right) = 1094.5)$ . Beşinci sütun ise, zaman süresinde olayın (ölüm) meydana geldiği hasta sayısını temsil etmektedir, örneğin, üçüncü ayın başında, olayın (ölüm) meydana geldiği hastaları sayısının 26 olduğunu, altıncı sütun ise ölenlerin yüzdesini temsil etmektedir.

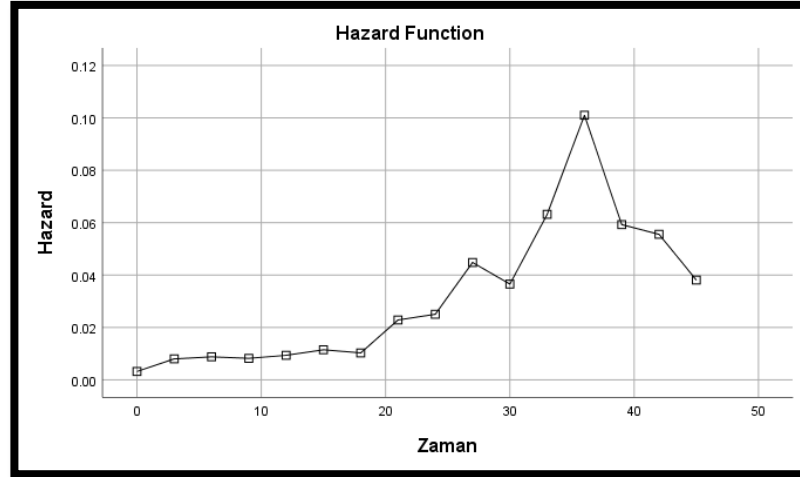
Yedinci sütun ise, sağkalım sağ kalan yüzdesidir örneğin, on ikinci aya giren 855 vaka varken, on iki ayı tamamlayan hastaların sonraki dönemde sağkalım olasılığı  $(1 - 0,03 = 0,97)$ 'e eşittir. Sekizinci sütun her dönemin sonunda sağkalım olasılığını temsil ederken, örneğin, hastaların on beşinci ayın sonuna kadar kümülatif sağkalım yüzdesi (%86)'dir. Dokuzuncu sütun ise, hastanın süresinin başından sonuna kadar sağkalım kaldığını varsayarak, en büyük riskin 36.ay sağkalım zamanında (0.10)'e ulaştığı, çok küçük bir süre boyunca başarısızlık olasılığı olarak tanımlanan risk fonksiyonunu temsil etmektedir.

Ek olarak, medyan sağkalım süresinin (35.06) ay olduğu ve bu zamanda sağkalım olasılığı (0.50) olarak elde edilmiştir. Bu durum, incelenen vakaların yarısının hayatta sağ kalacağını beklediği zamanı ifade etmektedir.



Şekil 5.14. Yaşam tabloları yöntemine göre sağkalım fonksiyonu eğrisi

Şekil (5.14)'göre tüm dönemler için kümülatif sağkalım fonksiyonu eğrisini göstermekte olup, her üç aylık, bir dönem sayılmakta olup, dikey eksen kümülatif sağkalım fonksiyonunu ve yatay eksen ise ay olarak zamanı temsil etmektedir. Hastaların sağkalım olasılığının sürekli azalması, (21) ay sağkalım olasılıkları (0.78)'e ulaşıldığında bulunmuştur. Buna ek olarak, yirmi dört ay Sağkalım olasılıkları (0.72)'e iken, (45) ay sağkalım olasılıkları ise (0.22) gözlenmektedir.



Şekil 5.15. Yaşam tabloları yöntemine göre hazard fonksiyonu eğrisi

Şekil (5.15)'e göre risk fonksiyonu eğrisini göstermektedir. Her üç ay, bir dönem sayılmakta olup, dikey eksen risk fonksiyonu ve yatay eksen ise ay olarak zamanı temsil

etmektedir. Dönem başında risk düşük göstermekteyken, yirminci ayda ise artmaya başlamakta, otuz aylık dönemde azalmaya dönerken, otuz altıncı aya kadar ise tekrar yükselmektedir. Daha sonra otuz dokuzuncu ayda düşüşe dönmekte ve en yüksek risk oranı (33-36) aylık dönemde bulunduğundan dönemdeki risk oranı (0.10)'a ulaşmakta olup, diğer tüm dönemler arasında en yüksek oran olarak sayılmaktadır.

### 5.9. Yapay Sinir Ağlarını Kullanarak Veri Analizi

Bu bölümde çalışma verileri yapay sinir ağları kullanılarak analiz edilecektir. Sağkalım verilerini analizinde yapay sinir ağlarını kullanmak için, veri işlemeden başlayarak, en iyi yapay sinir ağı modelini elde etmek için verilerin analiz edildiği son ağ modelinin formülasyonu ile biten birkaç sıralı adım vardır. Python ve Spss paket programlarıyla, bu bağlamda, aşağıdaki gibi veriler yapay sinir ağları kullanılarak iki aşamada analiz edilecektir.

- ***Birinci aşama***

Bu aşamada, veriler, Zaman ve hasta durumu olarak (Olay, Sansür) iki bağımlı değişkene dayalı olarak analiz edilecektir. Bağımsız değişkenler ise, her iki aşamada sekiz değişkende kalacaktır.

- ***Veri seti ve değişkenlerin belirlenmesi***

Yapay sinir ağları kullanılarak 1168 hastalara ait veriler 2 bağımlı değişken (ağın çıkışları) olarak ve 8 bağımsız değişken ise (ağın girişleri) belirlenmiştir. Bu değişkenler (yapay sinir ağın temel bileşenleri) aşağıdaki gibi gösterilecektir.

- ***Bağımlı değişkenler***

Yaşam (sağkalım) süresi yada başarısızlık süresi, çalışmada takibe başlanmasından ilgilenilen sonucu gösteren olayın gerçekleşmesine kadar geçen zamandır.

Durum, çalışma ilgilenilen sonuçtur. Başarısızlık olarak da adlandırılmaktadır. Ölüm, nüksetme, iyileşme gibi olay olarak adlandırılmaktadır. Sağkalım süresi tam olarak bilinmeyen vakalar ise, sansürlü (censored) olarak adlandırılır.

- **Bağımsız değişkenler (girdiler)**

Risk faktörleri sekiz ana faktör olarak Cinsiyet, Yaş, Kan grubu, Medeni durum, İl, Hastalığın türü, Kullanılan tedavi metodu ve Tedavi düzenlilik olarak belirlenmiştir.

- **Verilerin gruplara bölünmesi**

Bu adımda, eğitim sürecini başlatmak için gerekli bir aşama olarak çalışma verileri iki gruba bölünecektir.

- **Eğitim grubu (Training)**

Bu aşama, toplam çalışma verilerinden sırasıyla %90, %80 ve %70'inin her aşamada tahsis edileceği üç alt aşamaya ayrılmıştır. Bundan sonra, bulunacak üç oranın eğitiminin sonuçları arasında bir karşılaştırma yapılacaktır. Sinir ağının oluşturulmasında kullanılacak hata kareler ortalaması (MSE) kriterine göre en iyi eğitim oranı seçilecektir.

- **Test Grubu (Testing)**

Test gruplarında da aynı şekilde eğitim yüzdelere sırasıyla %10, %20 ve %30 olmak üzere üç tamamlayıcı yüzde tahsis edilecektir. Aşağıdaki tablo, daha önce bahsedilen eğitim ve test yüzdelerini göstermektedir.

**Tablo 5.16.** Eğitim ve test verilerinin bölünmüş oranları

| Gruplar  | 1.Grup |      | 2.Grup |      | 3.Grup |      |
|----------|--------|------|--------|------|--------|------|
|          | N      | Oran | N      | Oran | N      | Oran |
| Training | 1068   | %90  | 936    | %80  | 821    | %70  |
| Testing  | 100    | %10  | 232    | %20  | 347    | %30  |
| Toplam   | 1168   | %100 | 1168   | %100 | 1168   | %100 |

- **Eğitim (Training)**

Yapay sinir ağında eğitim aşaması en önemli aşamalardan biri olarak sayılmaktadır. Bu aşamada veriler, en az standart hataya göre elde edilecek iyi modeli seçmeye hazırlık eğitiminden kaynaklanan modelleri bilmek amacıyla sinir ağını eğitmek için kullanılacaktır. Ayrıca yapay sinir ağı, iki eğitim algoritması (scaled conjugate gradient



backpropagation (Gradient descent) kullanılarak 270 kez, bununla beraber Sigmoid, SoftMax ve TANH aktivasyon fonksiyonları kullanarak eğitilmiştir.

Aşağıdaki tablolar, yapay sinir ağlarının eğitiminden sonucu olarak bulunan en önemli eğitim sonuçlarını göstermektedir ve bu çalışma için yapay sinir ağı modelinin oluşturulmasında kullanılmak üzere en iyileri seçilecektir.

▪ **Modeller arasında karşılaştırma.**

Farklı modeller için eğitim sürecinden sonra, en düşük Hata Kareler Ortalaması ve en az da yanlış tahmin oranına göre en iyi modeli seçmek için üç eğitim oranı ve bunların eğitim algoritması yöntemleri arasında bir karşılaştırma yapılacaktır.

**Tablo 5.17.** Modeller arasında karşılaştırma sonuçları

| NO | Ağ Yapısı<br>Giriş, Gizli,<br>Çıktılar | Gizli<br>Katman<br>Sayısı | Eğitim<br>Algoritması | Eğitim Grubu. |       | Test Grubu.    |        |                            |
|----|----------------------------------------|---------------------------|-----------------------|---------------|-------|----------------|--------|----------------------------|
|    |                                        |                           |                       | Eğ.<br>Oranı  | MSE   | Test.<br>Oranı | MSE    | Yanlış<br>Tahmin<br>Oranı. |
| 1  | 8, 4, 3                                | 1                         | SCD                   | %90           | 0.219 | %10            | 0.203  | %10                        |
| 2  | 8, 5, 3                                |                           | GDX                   | %90           | 0.140 | %10            | 0.1180 | %9                         |
| 3  | 8,2,3,3                                |                           | SCD                   | %90           | 0.219 | %10            | 0.197  | %10                        |
| 4  | 8,6,5,3                                |                           | GDX                   | %90           | 0.136 | %10            | 0.114  | %9                         |
| 5  | 8,5,5,3                                | 2                         | SCG                   | %80           | 0.187 | %20            | 0.213  | %14.6                      |
| 6  | 8,8,5,3                                |                           | GDX                   | %80           | 0.130 | %20            | 0.148  | %14.2                      |
| 7  | 8,2,10,3                               |                           | SCG                   | %70           | 0.202 | %30            | 0.216  | %15.3                      |
| 8  | 8,6,5,3                                |                           | GDX                   | %70           | 0.128 | %30            | 0.145  | %13.8                      |

Tablo (5.17), Bir önceki aşamadaki eğitim sürecinin bir özetini içermektedir. Hata kareler ortalaması ve yanlış tahmin oranının standartlarına göre, en iyi modeli seçmenin başlangıç noktası olarak, üç eğitim oranlarında 90, 80 ve 70'te her bir eğitim algoritmasına en iyi model seçilecektir. Daha önce belirtilen standartlara göre, bu çalışma için en iyi sinir ağı yapısının modeli, en az Hata Kareler Ortalaması(MSE) ve en az yanlış tahmin oranı içeren 8,6,5,3 olduğu bulunmuştur.

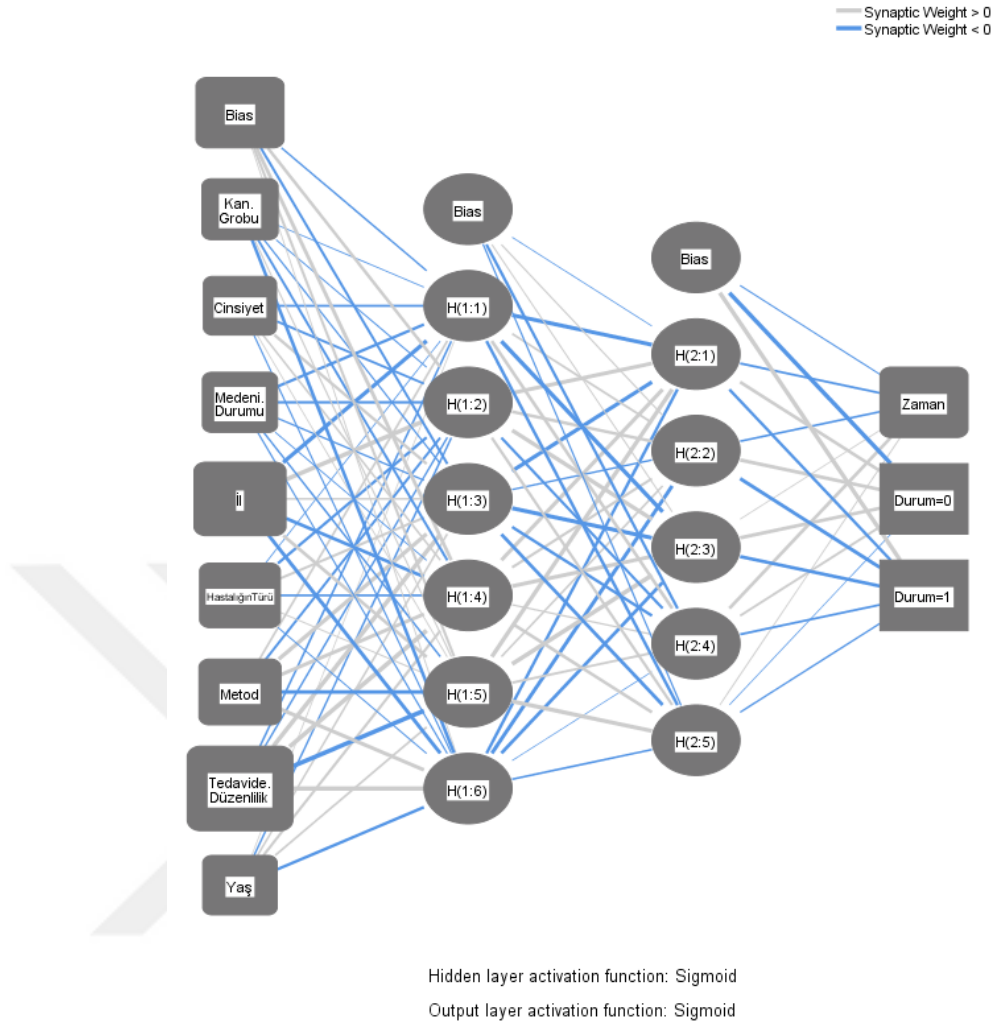
▪ **Önerilen yapay sinir ağı modelini oluşturmak**

Bu adım, önceki adımların nihai bir sonucu olarak gelmektedir. Başka bir deyişle bu aşamadan önce modelleri eğitmek, karşılaştırmak ve eğitim sürecinden kaynaklanan en iyi modeli kullanarak sağkalım modellerinin (Zaman, Durum) bağımlı değişkenlerini tahmin etmek için en iyi modeli seçmek söz konusuydu. Bu adımda önerilen en iyi yapay sinir ağı modeli aşağıdaki gibi oluşturacaktır.

**Tablo 5.18.** Ağ bilgileri

|                        |                                       |   |                          |
|------------------------|---------------------------------------|---|--------------------------|
| <b>Giriş Katmanı</b>   | <b>Değişkenler.</b>                   | 1 | Kan grubu                |
|                        |                                       | 2 | Cinsiyet                 |
|                        |                                       | 3 | Medeni durumu            |
|                        |                                       | 4 | Konut şehir              |
|                        |                                       | 5 | Hastalığın türü          |
|                        |                                       | 6 | Kullanılan tedavi metodu |
|                        |                                       | 7 | Tedavi düzenlilik        |
|                        |                                       | 8 | Yaş                      |
| <b>Gizli Katmanlar</b> | Birim Sayısı.                         |   | 8                        |
|                        | Gizli Katman Sayısı                   |   | 2                        |
|                        | Birinci Gizli Katmandaki Birim Sayısı |   | 6                        |
|                        | İkinci Gizli Katmandaki Birim Sayısı  |   | 5                        |
|                        | Aktivasyon Fonksiyonu                 |   | Sigmoid                  |
| <b>Çıkış Katmanı</b>   | Bağımlı değişkenler                   | 1 | Zaman                    |
|                        |                                       | 2 | Dönem sonundaki durum.   |
|                        | Birim Sayısı.                         |   | 3                        |
|                        | Rescaling Method for Scale Dependents |   | Normalized               |
|                        | Aktivasyon Fonksiyonu                 |   | Sigmoid                  |
|                        | Hata Fonksiyonu                       |   | Kareler Toplamı          |

Tablo 5.18 ve Şekil 5.16, her katmandaki nöron sayısını, giriş katmanı için sekiz bağımsız değişkeni (Kan grubu, Cinsiyet, Medeni durumu, Konut şehir, Hastalığın türü, Kullanılan tedavi metodu, Tedavi düzenlilik, Yaş) göstermektedir. Ayrıca, çıkış katmanı hariç her katmandaki yanlılık (Bias) ek olarak, birinci gizli katmanda 6 nöron, ikinci gizli katmanda ise 5 nöron bulunmaktadır. Çıkış katmanı ise, ikili bağımsız değişken de dahil olmak üzere iki bağımsız değişken olarak üç nöron yer almaktadır. Aktivasyon fonksiyonu olarak gizli ve çıkış katmanları için sigmoid fonksiyonu, Hata Fonksiyonu olarak ise hatanın karelerinin toplamı kullanılmıştır.



Şekil 5.16. Birinci aşamadaki seçilen sinir ağı mimarisi

#### ▪ **Modelin kalitesinin geçerliğinin ölçülmesi**

Bu adımda, önerilen modelin kalitesi bir dizi istatistiksel kriter kullanılarak ölçülecektir. Sinir ağı modelinin oluşumunda yer alan bu kriterlerden en önemlisi, Hata Kareler Ortalaması(MSE) buna ek olarak, yanlış tahmin yüzdesi ve değişkenlerin göreceli önemini, ve sınıflandırma tablosudur.

**Tablo 5.19. Model özeti**

|               |                                                          |         |
|---------------|----------------------------------------------------------|---------|
| <b>Eğitim</b> | Sum of Squares Error                                     | 145.596 |
|               | Average Overall Relative Error                           | 0.567   |
|               | Percent Incorrect Predictions for Categorical Dependents | 13.9%   |
|               | Relative Error for Scale Dependents                      | 0.858   |
|               | Dönem sonundaki durum.                                   |         |
| <b>Test</b>   | Sum of Squares Error                                     | 11.412  |
|               | Average Overall Relative Error                           | 0.435   |
|               | Percent Incorrect Predictions for Categorical Dependents | 9.0%    |
|               | Relative Error for Scale Dependents                      | 0.885   |
|               | Dönem sonundaki durum.                                   |         |

a. Error computations are based on the testing sample.

Tablo (5.19), eğitim ve test verisi için hata kareler toplamı sırasıyla 145.596 ve 11.412 olduğu ve test grubu için yanlış tahmin yüzdesinin %9 olduğu model bilgilerinin bir özetini sunmaktadır. Başka bir deyişle, eğitim grubu için Hata Kareler Ortalaması (MSE) 0.114 olarak bulunmuştur.

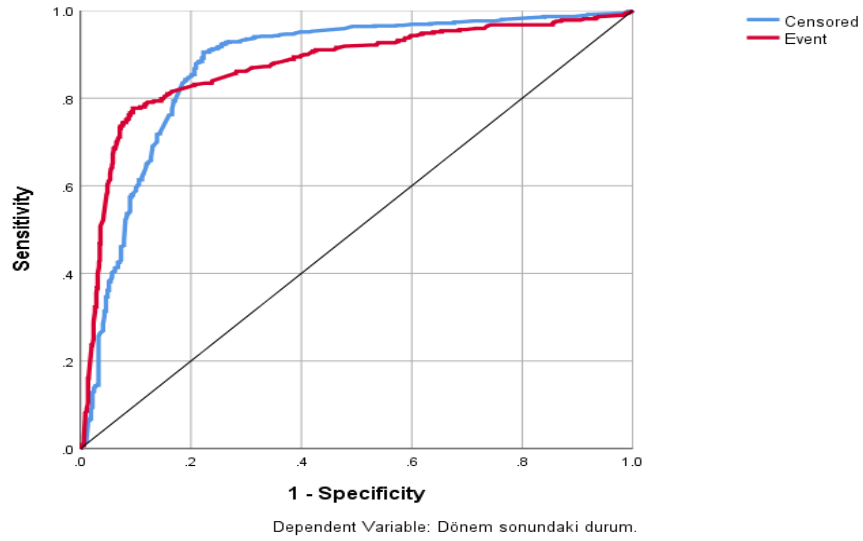
**Tablo 5.20. Sınıflandırma tablosu**

| Örnek  | Observed        | Censored | Event | Tahmin.     |
|--------|-----------------|----------|-------|-------------|
|        |                 |          |       | Doğru Oranı |
| Eğitim | Sansürlü        | 677      | 60    | %91.9       |
|        | Olay            | 88       | 243   | %73.4       |
|        | Overall Percent | %71.6    | %28.4 | %86.1       |
| Test   | Sansürlü        | 63       | 0     | %100        |
|        | Olay            | 9        | 28    | %75.7       |
|        | Overall Percent | %72.0    | %28   | %91.        |

Bağımlı değişken: Dönem sonundaki durum.

Tablo (5.20), Eğitim aşamasında Durum değişkeni Sansürlü, Olay (Censored, Event) için doğru sınıflandırma yüzdelерinin sırasıyla %92 ve %73 olduğu, test aşamasında ise oranın %100 ve %76 olduğu göstermektedir. Genel olarak tablo sonuçlarına göre, eğitim ve test aşamalarında doğru sınıflandırma yüzdeleri sırasıyla %86

ve %91 idi. Burada elde edilen sonuçlar önerilen yapay sinir ağı modelin sağkalım verileri analizinde doğru sınıflandırma gücünü göstermektedir.

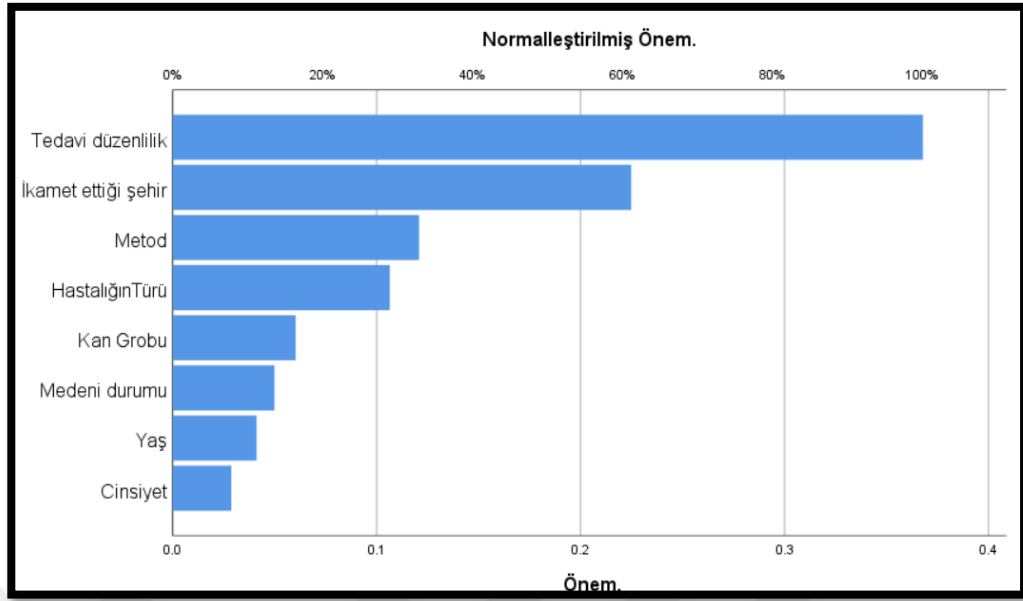


Şekil 5.17. ROC eğrisi

Tablo 5.21. Eğri altında kalan alanı

|                       |          |       |
|-----------------------|----------|-------|
| Dönem sonundaki durum | Sansürlü | 0.874 |
|                       | Olay     | 0.874 |

ROC eğrisinin altında kalan alan, testin olay ile olay olmayan (Censored) hastaları ayırmadaki doğruluk oranını belirlemektedir. Eğriyi karşılaştırmada kullanılan referans değer olan 0.5'ten (%50) saptığının açık olduğu durumlarda, ROC eğrisine göre, yapay sinir ağı modeli gözlemlenen verilerin durumlarını sınıflandırmada şans faktöründen daha iyi performans göstermektedir. Tablo (5.21)'da gösterildiği gibi, eğri altındaki alan yüzdesi %87.4 idi ve bu, referans değere kıyasla çok iyi bir yüzdedir. Demek ki, doğru pozitif oranı (Sensitivity) ile yaklaştıkça kullanılan modelin doğru sınıflandırılmasının kalitesini göstermektedir.



Şekil 5.18. Bağımsız değişkenin önemi

Tablo 5.22. Bağımsız değişkenin önemi

| değişkenler              | Önem | Normalleştirilmiş Önem |
|--------------------------|------|------------------------|
| Kan grubu                | .060 | 16.4%                  |
| Cinsiyet                 | .029 | 7.8%                   |
| Medeni durumu            | .050 | 13.6%                  |
| İkamet ettiği şehir      | .225 | 61.1%                  |
| Hastalığın türü.         | .106 | 28.9%                  |
| Kullanılan tedavi metodu | .121 | 32.8%                  |
| Tedavi düzenlilik.       | .368 | 100.0%                 |
| Yaş                      | .041 | 11.2%                  |

Tablo (5.22)'Yapay sinir ağ modelindeki her bağımsız değişkenin etkisini görel ve normalleştirilmiş önem açısından vermektedir. Şekil(5.18)'deki grafik ayrıca değişkenlerin önemini, modelin her bir girdi değişkeninin değişimine ne kadar duyarlı olduğunu göstermektedir. Ayrıca Tablo (5.22) gösterildiği gibi tüm değişkenlerden en fazla önemin ilk sırada tedavi düzenliliği değişkeninde (%100) olduğu bulunmaktadır. Ardından ikamet edilen şehir değişkeni ve tedavi yöntemi vb. aynı şekilde cinsiyet değişkeni, değişkenlerden en az önemli olarak ve son sırada yer almaktadır.

▪ **İkinci aşama**

Bu aşamada yapay sinir ağları uygulanacak ancak bu durumda bağımlı değişken risk fonksiyonu olurken, bağımsız değişkenler ise ilk aşamadaki ile aynı sekiz değişken (cinsiyet, yaş, kan grubu, medeni durum, il, hastalığın türü, kullanılan tedavi metodu ve tedavi düzenlilik) olacaktır. Doğrusu ağın girdileri birinci adımda olduğu gibi üçüncü bölümde ele alınan Cox Regresyon modelinde olduğu gibi Hazard fonksiyonunu etkileyen sekiz bağımsız değişken olacaktır. Gizli katmanlar eğitim sürecinden çıkan en iyi mimariye göre belirlenecek, ağ çıktıları ise tahmin edilecek Hazard fonksiyonu değişkenleri olacaktır.

### 5.9.1. Eğitim (Training)

Eğitim sürecinde, Hata Kareler Ortalaması (Mean Squares Error) en küçük veren, en iyi ağ mimarisini elde etmek için ağı On iki bin yüz elli (12150) kez eğitilmiştir. Python programı kullanılarak ikinci aşamada gerçekleştirilmiştir. Eğitim sürecini Tablo 5.23'de gösterilen üç eğitim oranında gerçekleştirdikten sonra sonuçlar aşağıdaki gibi elde edilmiştir.

### 5.9.2. Modeller arasında karşılaştırma.

On iki bin yüz elli kere farklı modeller için eğitim sürecinden sonra, ADAM ve SGD Algoritmasını kullanarak en az Hata Kareler Ortalamasına göre en iyi altı model seçilecek, en iyi olanından birini seçmek için daha önce seçilen üç eğitim oranı birinciler arasında bir karşılaştırma yapılacaktır.

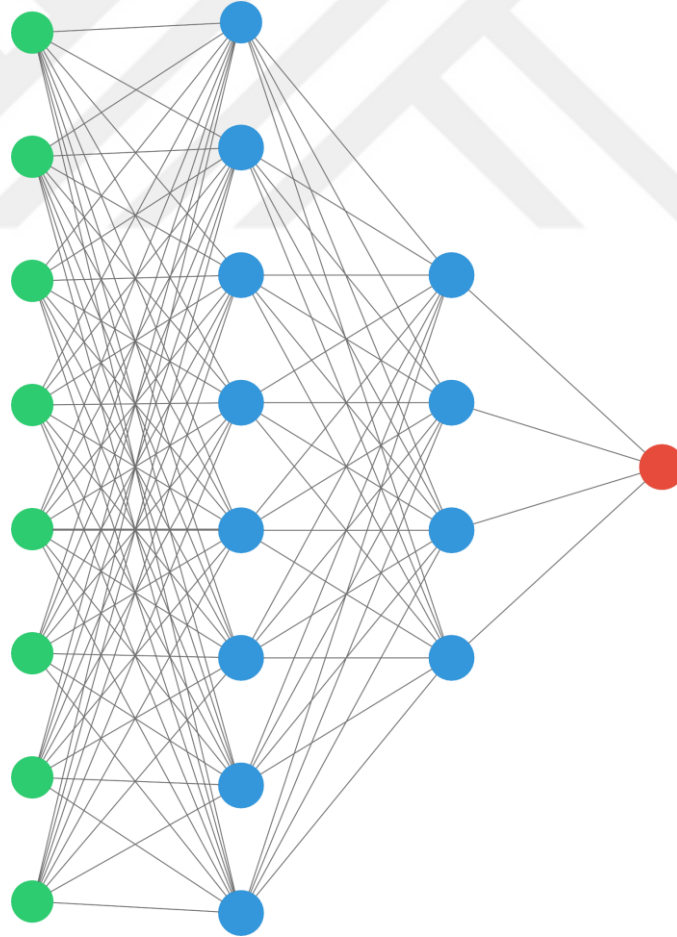
**Tablo 5.23.** Modeller arasında karşılaştırma sonuçları

| NO | Ağ Yapısı | Gizli Katman Sayısı | Eğitim Algoritması | Eğ. Oranı | Test. Oranı | MSE   | MAE   |
|----|-----------|---------------------|--------------------|-----------|-------------|-------|-------|
| 1  | 8,8,4,1   | 2                   | Adam               | %90       | %10         | 0.064 | 0.181 |
| 2  | 8,8,8,1   |                     | Adam               | %80       | %20         | 0.071 | 0.187 |
| 3  | 8,10,2,1  |                     | Adam               | %70       | %30         | 0.102 | 0.216 |
| 4  | 8,2,6,1   |                     | SGD                | %90       | %10         | 0.065 | 0.203 |
| 5  | 8,10,10,1 |                     | SGD                | %80       | 20          | 0.073 | 0.197 |
| 6  | 8,10,10,1 |                     | SGD                | %70       | %30         | 0.104 | 0.220 |

Tablo (5.23) bir önceki aşamadaki eğitim sürecinin bir özetini içermektedir. (MSE) ve (MAE) göre, en iyi modeli seçmenin başlangıç noktası olarak, üç eğitim oranlarında 90, 80 ve 70’te ADAM ve SGD algoritmasına göre en iyi model seçilecektir. Daha önce belirtilen standartlara göre, bu çalışma için en iyi sinir ağı yapısının modeli, en az Hata Kareler Ortalaması (MSE) ve en az (MAE) içeren 8,8,4,1 modeli olduğu bulunmuştur.

### 5.9.3. Önerilen yapay sinir ağı modelini Oluşturmak

Bu adım da, 1.aşamadaki gibi, önceki adımların nihai bir sonucu olarak ta gelmektedir. Başka bir deyişle bu aşamadan önce modelleri eğitmek, karşılaştırmak ve eğitim sürecinden kaynaklanan en iyi modeli kullanarak bağımlı Hazard değişken tahmin etmek için en iyi modeli seçmek gerçekleştirilmiştir. Bu adımda önerilen en iyi yapay sinir ağı modeli aşağıdaki gibi oluşturacaktır.



Şekil 5.19. Seçilen sinir ağı mimarisi.



### 5.10. Cox Modeli İle YSA'nın Sonuçlarının Karşılaştırılması.

Veriler her iki yöntem kullanılarak ayrı olarak analiz edildikten sonra, her iki yöntemden en iyi model seçilmiştir. Sonrasında ise en önemli adım sayılan MSE ve MAE karşılaştırma kriterlerine göre, en iyi yöntem seçilecektir.

**Tablo 5.24.** *Cox Modeli ile ysa'nın sonuçlarının karşılaştırılması*

| Metot | Cox   | YSA   |
|-------|-------|-------|
| MSE   | 0.138 | 0.064 |
| MAE   | 0.189 | 0.180 |

Tablo 5.24'e göre, yapay sinir ağlarının sağkalım verilerini analiz etmede Cox regresyon modelinden daha iyi performans gösterdiği bulunmaktadır. Yapay sinir ağları modelinin MSE ve MAE değerleri sırasıyla (0.064), (0.180) iken Cox modelinin MSE ve MAE değerleri sırasıyla (0.138), (0.189) olarak bulunmuştur.

## 6. SONUÇ VE TARTIŞMA

Bu tez çalışmasında, Sağkalım Analizinde Cox Regresyon Modeli ve Yapay Sinir Ağları modelleri kullanılarak, yapay sinir ağı modelleri ile Cox regresyon modeli karşılaştırılmıştır. Tezin uygulama kısmında ise en yaygın hastalıklardan biri olan kanserle başa çıkmada Yemen Cumhuriyeti'ndeki lösemi hastalarının sağkalım sürelerini etkileyen en önemli faktörler (risk faktörleri) belirlenmeye çalışılmıştır. Çalışmada kullanılan veriler ise, Yemen Cumhuriyetindeki Ulusal Kanserle mücadele Vakfı ve Ulusal Kanser mücadele Merkezi 2017 yılı ocak ayı ve 2022 yılı Şubat ayı arasında 1168 lösemi hastasına ait bilgilerdir. İlgili verilere Cox regresyon modeli ve yapay sinir ağları modelleri uygulanarak elde edilen sonuçlar değerlendirildiğinde:

Yapılan analizler sonucunda, ilk aylardaki ölüm risk oranı sıfıra oldukça yakın olarak bulunmuştur, bu durum ilk aylarda lösemi hastaları için temel riskin neredeyse yok olduğunu göstermektedir. Ardından risk üçüncü aydan itibaren ortaya çıkmaya başlamakta, yirmi dördüncü aya ulaşana kadar giderek artmakta ve ardından yine başlangıçta olduğundan çok daha fazla artmaya başlamaktadır. 36.ayda ise temel risk oranı %10'a ulaşmıştır. Bununla beraber, kırk beşinci ayda risk oranının %24'e doğru giderek artmakta ve çalışmanın son aylarında ise risk daha hızla artarken, temel risk oranı %37'ye ulaşmıştır. Kırk dokuzuncu ayda ise risk daha da artmıştır.

Temel sağkalım fonksiyonu ise, zamanın ilerlemesiyle azar azar azalmaktadır. Hastaların sağkalım olasılığı eğrisi, (29) aydaki sağkalım olasılığı (0.70)'e ulaşmıştır. Bundan sonra, (40) aydan fazla kalan sağkalım az sayıda hasta olduğu için eğri daha az düzgün hale geldiğinden, sağkalım olasılığındaki düşüş oldukça hızlı bir şekilde başlanmaktadır. Sağkalım süresinin medyanı (36) aydır. Bu da lösemili hastaların yarısının hayatta kalamadığı (ölümle karşı karşıya kaldığı) noktadır. Bu durumda eğri önemli ölçüde düşmeye devam etmektedir. Burada sağkalım olasılığı, 49.ayda, (0.03)'e eşittir.

Cox modeli, sağkalım sürelerinin belirli bir dağılımına bağlı değildir. Bu nedenle, (Medeni Durumu, İkamet ettiği şehir (İl), Hastalığın Türü, Kullanılan Tedavi Metodu, Tedavide Düzenlilik ve Hasta Yaşı) bağımsız değişkenlerin üzerindeki etkisinin göstermek için kullanılabilir. Ayrıca, model aynı anda birden fazla değişkenin etkisinin değerlendirilmesine ve lösemili kişilerin sağkalım süresi üzerindeki etkilerinin gösterilmesine olanak tanıyan her bir değişken için katsayı tahminleri sağlanmaktadır.

Cox modelinde sağkalımı etkileyen risk değişkenleri; Medeni durumu, İkamet ettiği şehir (il), Hastalığın türü, Kullanılan tedavi metodu, Tedavide Düzenlilik ve Yaş değişkenleri olup, sağkalım süresi üzerinde anlamlı bir etkiye sahip olduğu bulunmuştur. Geri kalan, cinsiyet ve kan grupları değişkenlerinin ise, sağkalım süresi üzerinde anlamlı bir etkisinin olmadığı görülmüştür.

Medyan sağkalım süresi (35.06) ay olarak bulunmuş ve bu zamanda sağkalım olasılığı ise (0.50) olarak elde edilmiştir. Elde edilen bu sonuç, incelenen vakaların yarısının sağkalım kalacağını beklediği zamanı ifade etmektedir.

Tip (AML)'i olan hastalar için tahmini risk oranı (0.876) olup, tip (AML)'i olan hastalarda ölüm oranı, tip (ALL) olan hastalara göre (%12.4) daha düşüktür. Ayrıca, tip (CML)'i hastaları, tip (ALL)'e hastalarına göre (%49.5) kat daha fazla ölürken, tip (CLL) olan hastaların ise, tip (ALL)'e hastalarına göre, (1.731) kat daha fazla olduğu bulunmuştur.

Yapay sinir ağları kullanarak, zaman ve hasta durumu olmak üzere iki bağımlı değişken ile ilk aşamada en iyi model olarak 8,6,5,3 modeli seçilmiştir. Bu durum göstermektedirki, 8,6,5,3 modeli, olayı ve bu olay meydana gelmeden önceki zamanı tahmin edilebilmekte ve Cox Regresyon modeli kullanıldığında bu sonuç elde edilemediği için bu yapay sinir ağı modelinde bir avantaj olarak sayılmaktadır.

Yapay sinir ağı kullanarak hazard fonksiyonunu tahmin etmek içinse on iki bin yüz elli (12150) model arasından en iyi model olarak 8,8,4,1 modeli seçilmiştir.

Bu çalışmada elde edilen sonuçlar, Yapay sinir ağlarının sağkalım verilerini analiz etmede Cox regresyon modelinden daha iyi performans gösterdiğini ortaya koymaktadır. Yapay sinir ağları modelinin Hata Kareler Ortalaması ve Ortalama mutlak hatası sırasıyla 0.064, 0.180 iken Cox modelinin Hata Kareler Ortalaması ve Ortalama mutlak hatası sırasıyla 0.138, 0.189 olarak elde edilmiştir.

Bu tez çalışmasında, yapay sinir ağlarının performansının Cox regresyon modelinden daha iyi performans gösterdiği yönünde elde edilen sonuçlar literatürde daha önce yapılan çalışmalarla tutarlı sonuçlar ortaya koymaktadır. Öte yandan, bu çalışmanın literatürde yapılan diğer çalışmalardan farkı ise, önceki çalışmalardan daha büyük bir örneklem büyüklüğüne sahip olan lösemi hastaları ile ilgilenmesidir. Ayrıca, önceki çalışmalarda sınıflandırma doğruluğu %72 ile %83 arasında değişirken, bu çalışmada sınıflandırma doğruluğu %87.4 olarak elde edilmiştir. Ek olarak bu çalışma, hazard

fonksiyonunun tahminini içerirken, önceki çalışmalar sadece sınıflandırma sürecini kullanmakla sınırlı kalmıştır.

Bu tez çalışmasının sonuçları göz önünde bulundurularak lösemiden kaynaklanan ölümlerin nedenlerinin araştırılabileceği önerisinde bulunulabilir. Ayrıca Sağkalım analizi çalışmalarının yapılmasını, kanser ve diğer hastalıklara maruz kalan hastaların sağkalım süresini etkileyen faktörlerin, sosyo-ekonomik olayların belirlenmesini, bununla birlikte sonuçlardaki esneklikleri ve doğrulukları nedeniyle yapay sinir ağıları kullanarak uzun süreler ve büyük örnekler alınmasında önerilebilir.



## KAYNAKÇA

- [1] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, Using multivariate statistics, vol. 5. *Pearson Boston, MA*, 2007.
- [2] S. T. Liu, SAS ® Survival Analysis Techniques for Medical Research , vol. 46, no. 2. 2004.
- [3] P. Walstra, J. T. M. Wouters, and T. J. Geurts, “Survival Analysis A Practical Approach,” *Dairy Science & Technology, CRC Taylor & Francis Group*, p. 267, 2005.
- [4] E. T. LEE and J. WENYUWANG, Statistical Methods for Survival Data Analysis, Third Edit. Canada. *A john wiley & sons, inc., publication*, 2003.
- [5] A. W. Marshall and I. Olkin, Life distributions, vol. 13. *Springer*, 2007.
- [6] J. P. Klein, Handbook of Survival Analysis. 2016.
- [7] Ç. Elmas, Yapay zeka uygulamaları:(yapay sinir ağı, bulanık mantık, genetik algoritma). *Seçkin Yayıncılık*, 2007.
- [8] E. Öztemel, “Yapay sinir ağları [Artificial neural networks],” *İstanbul: Papatya Yayıncılık*, 2012.
- [9] S. Sabouri, H. Esmaily, S. Shahidsales, and M. Emadi, “Survival prediction in patients with colorectal cancer using artificial neural network and cox regression,” *International Journal of Cancer Management*, vol. 13, no. 1, 2020, doi: 10.5812/ijcm.81161.
- [10] A. Biglarian, E. Hajizadeh, A. Kazemnejad, and M. R. Zali, “Application of artificial neural network in predicting the survival rate of gastric cancer patients,” *Iranian Journal of Public Health*, vol. 40, no. 2, pp. 80–86, 2011.
- [11] M. Kiknadze, İ. A. Üniversitesi, M. Bölümü, and A. Gürhanli, “Araştırma Makalesi / Research Article YapaSinir Ağı Kullanarak MemeKanseri Hastalığını Tahmini,” *journal of engineering systems and architecture cilt*, vol. 4, no. 2, pp. 255–272, 2020.
- [12] B. B. 3\* Amir Hossein Hashemian 1, 2, M. R. 2, A. Bardideh4, and E. Zand-Karimi, “Comparison of Artificial Neural Networks and Cox Regression Models in Prediction of Kidney Transplant Survival,” *International journal of Advanced Biological and Biomedical Research*, vol. 1, no. 10, pp. 1204–12012, 2013.
- [13] A. Biglarian, E. Hajizadeh, and A. Kazemnejad, “Comparison of artificial neural network and Cox regression models in survival prediction of gastric cancer patients,” *Koomesh*, vol. 11, no. 3, pp. 205–211, 2010.

- [14] A. Biglarian, ( E. H., and Anoshirvan Kazemnejad, “Comparison of artificial neural network and Cox regression models in survival prediction of gastric cancer patients,” *koomeshjournal.semums.ac.ir/article-1-741-en.html*, vol. 11, no. 3, pp. 215–220, 2010, [Online]. Available: <http://koomeshjournal.semums.ac.ir/article-1-741-en.html>.
- [15] C. Y. Chen, Y. F. Chen, H. Y. Chen, C. T. Hung, and H. Y. Shi, “Artificial neural network and cox regression models for predicting mortality after hip fracture surgery: A population-based comparison,” *Medicina (Lithuania)*, vol. 56, no. 5, pp. 1–10, 2020, doi: 10.3390/medicina56050243.
- [16] L. Spelt, J. Nilsson, R. Andersson, and B. Andersson, “Artificial neural networks e A method for prediction of survival following liver resection for colorectal cancer metastases,” 2013, doi: 10.1016/j.ejso.2013.02.024.
- [17] D. Ansari, J. Nilsson, R. Andersson, S. Regnér, B. Tingstedt, and B. Andersson, “Artificial neural networks predict survival from pancreatic cancer after radical surgery,” *American Journal of Surgery*, vol. 205, no. 1, pp. 1–7, 2013, doi: 10.1016/j.amjsurg.2012.05.032.
- [18] E. L. Kaplan and P. Meier, “Nonparametric estimation from incomplete observations,” *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.
- [19] D. R. Cox, “Regression Models and Life-Tables,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, Jan. 1972, doi: 10.1111/j.2517-6161.1972.tb00899.x.
- [20] J. Fox and S. Weisberg, “Cox Proportional-Hazards Regression for Survival Data in R The Cox Proportional-Hazards Model,” no. February, pp. 1–20, 2011.
- [21] P. D. Allison, Survival analysis using SAS: a practical guide, *Second Edi. USA: Sas Institute*, 2010.
- [22] J. D. Kalbfleisch and R. L. Prentice, The statistical analysis of failure time data, vol. 360. *John Wiley & Sons*, 2011.
- [23] E. Jarosova, I. Mala, and M. Esser, “Analysis of factors influencing time of unemployment,” *Aplimat*, sp, pp. 525–530, 2004.
- [24] D. Collett, Modelling survival data in medical research. *CRC press*, 2015.
- [25] E. T. Lee and J. Wang, Statistical methods for survival data analysis, vol. 476. *John Wiley & Sons*, 2003.
- [26] D. G., Kleinbaum, and M. Klein, Survival Analysis A Self-Learning Text, *Third Edit., vol. 33, no. 4. Spring*, 2012.

- [27] J. O'Quigley, Proportional hazards regression, vol. 542. Springer, 2008.
- [28] J. F. Lawless, Statistical models and methods for lifetime data, vol. 362. John Wiley & Sons, 2011.
- [29] J. P. Klein and M. L. Moeschberger, Survival analysis: techniques for censored and truncated data. *Springer Science & Business Media*, 2006.
- [30] M. F. A. Azzoum, "Statistical Models to Predict Ischemic Stroke in Gaza Strip," *Batch2*, 2012.
- [31] Z. Şen, Yapay sinir ağları. Su Vakfı, 2004.
- [32] A. Thakur, "Fundamentals of Neural Networks," *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, no. VIII, pp. 407–426, 2021, doi: 10.22214/ijraset.2021.37362.
- [33] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [34] D. O. Hebb, "The first stage of perception: growth of the assembly," *The Organization of Behavior*, vol. 4, pp. 60–78, 1949.
- [35] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.
- [36] B. Kröse, B. Krose, P. van der Smagt, and P. Smagt, "An introduction to neural networks," 1993.
- [37] R.-M. Ion, D. Munteanu, and G.-C. Cocina, "Concept of artificial neural network (ANN) and its application in cerebral aneurism with multi walls carbon nanotubes (MWCNT)," 2009.
- [38] B. D. Ripley and R. M. Ripley, "Neural networks as statistical methods in survival analysis," *Clinical Applications of Artificial Neural Networks*, pp. 237–255, 2009, doi: 10.1017/cbo9780511543494.011.
- [39] M. Buscema, "A brief overview and introduction to artificial neural networks," *Substance use & misuse*, vol. 37, no. 8–10, pp. 1093–1148, 2002.
- [40] J. M. Zurada, Introduction to artificial neural systems, vol. 8. West St. Paul, 1992.
- [41] P. D. Wasserman, *Neural computing: theory and practice*. Van Nostrand Reinhold Co., 1989.
- [42] L. V Fausett, Fundamentals of neural networks: *architectures, algorithms and applications*. Pearson Education India, 2006.

- [43] S. Sharma, S. Sharma, and A. Athaiya, “Activation functions in neural networks,” *Towards Data Sci*, vol. 6, no. 12, pp. 310–316, 2017.
- [44] D. Eyl, M. Enst, L. Anab, V. E. End, and S. Tez, “Petrokimya sektöründe talep tahmininde yapay sinir ağlarının kullanılması “ *petkim a.ş. örneği* “,” 2012.
- [45] Ç. Elmas, “Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama),” *Seçkin Yayıncılık, Ankara*, pp. 22–37, 2003.
- [46] H. Cartwright Editor, Artificial Neural Networks Artificial Neural Networks *Third Edition*, vol. 2190. .
- [47] Beckenkamp, Fábio Ghignatti. A component architecture for artificial neural network systems. *PhD Thesis*, 2002..





## ÖZGEÇMİŞ

**ORCID NO** : 0000-0003-0663-0305

**Ad Soyad** : Elias Abdullah AL-SAMAI

**Yabancı Dil** : Arapça, Türkçe, İngilizce

### Eğitim ve Mesleki Geçmişi:

- 2019-2023, Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, İstatistik Anabilim Dalı,Doktora.
- 2017-2018, Anadolu Ün. TÖMER, C1 Türkçe dili.
- 2015-2017, Araştırma Görevlisi, Taiz Ün. İdari Bilimsel Fakültesi İstatistik Bölümü
- 2012-2014, Aden Ün. İdari Bilemler Fakültesi, Uygulamalı İstatistik, Yüksek Lisans
- 2012-2014 Araştırma Görevlisi, Aden Ün. İdari Bilimler Fakültesi, İstatistik Bölümü.
- 2009-2012, Araştırma Görevlisi, Taiz Ün. İdari Bilimsel Fakültesi İstatistik Bölümü
- 2005-2009, Aden Ün. İdari Bilimler Fakültesi, Uygulamalı İstatistik, Lisans
- 2004-2006, Aden Ün. Bilgisayar Koleji, Ön lisans (iki yıllık) - Bilgisayar Programcılığı.

### Yayınları ve/veya Bilimsel/Sanatsal Faaliyetleri:

**AL-SAMAI EA, ŞENTÜRK S,(2023).** The Use Of Cox Regression Model In The Survival Analysis For Leukemia Patients In The Republic Of Yemen. Eskişehir Technical University Journal of Science and Technology A - Applied Sciences and Engineering. September 2023;24(3):192-206. doi:10.18038/estubtda.1335496

**Al-Ansari, A, A and Al-samai,(2021)** The Use of Simple linear Regression and Box-Cox Regression for analyze the impact of exports on GDP in the Republic of Yemen for the period (1990 - 2015). International Conference of Technology, Science, and Administration- 2021 (ICTSA-2021).

**Ateq et al,Alsamai(2021),** A summary of public opinion trends regarding the current situation and the role of the parties to the conflict. Yemeni indicator, Mokha center for strategic studies, İstanbul-Türkiye.

**Al-samai, Elias (2014)**, Yüksek Lisans Tezi, Aden Ün. Use of Box – Jenkins Models and Exponential Smoothing for Forecasting By Consumption of Electrical Power in Adengovernorate.

**Alsamai, E(2013)**, Evaluation of the conditions of prisoners in the central prisons (Ibb - Lahj - Al Dhalea - Aden) - Rehabilitation and Legal Aid Foundation - Aden. 2013

**Ödülleri:**

- 2010, Fakülte birincisi, Cumhurbaşkanı Ödülü, Sana’a-Yemen
- 2009, Aden Ün. Bölüm birincisi, İdari Bilimler Fakültesi Ödülü Aden- Yemen.

**Mesleki Birlik/Dernek/Kuruluş Üyelikleri:**

2022, İSNAT Öğrenciyi Destekleme ve Dayanışma Derneği, İstanbul-Türkiye .