BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİMDALI
BİLGİSAYAR MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS PROGRAMI

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KANSER HASTALIĞI TAKİBİ

HAZIRLAYAN

CANER BOZKURT

YÜKSEK LİSANS TEZİ

ANKARA – 2022

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİMDALI
BİLGİSAYAR MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS PROGRAMI

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KANSER HASTALIĞI TAKİBİ

HAZIRLAYAN

CANER BOZKURT

YÜKSEK LİSANS TEZİ

TEZ DANIŞMANI

Dr. Öğr. Üyesi TUNÇ AŞUROĞLU

ANKARA – 2022

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı çerçevesinde Caner Bozkurt tarafından hazırlanan bu çalışma, aşağıdaki jüri tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 11 / 08 / 2022

Tez Adı: Makine Öğrenmesi Yöntemleri İle Kanser Hastalığı Takibi

Tez Jüri Üyeleri (Unvanı, Adı - Soyadı, Kurumu)	**İmza**
Doç. Dr. Mehmet Serdar Güzel, Ankara Üniversitesi	……..
Dr. Öğr. Üyesi Çağatay Berke Erdaş, Başkent Üniversitesi	……..
Dr. Öğr. Üyesi Tunç Aşuroğlu, Başkent Üniversitesi	……..

ONAY

Prof. Dr. Faruk Elaldı
Fen Bilimleri Enstitüsü Müdürü
Tarih: … / … / ……..

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
YÜKSEK LİSANS ÇALIŞMASI ORİJİNALLİK RAPORU

Tarih: 26 / 07 / 2022

Öğrencinin Adı, Soyadı : Caner Bozkurt

Öğrencinin Numarası : 22020163

Anabilim Dalı : Bilgisayar Mühendisliği

Programı : Bilgisayar Mühendisliği Tezli Yüksek Lisans

Danışmanın Unvanı/Adı, Soyadı : Dr. Öğr. Üyesi Tunç Aşuroğlu

Tez Başlığı : Makine Öğrenmesi Yöntemleri İle Kanser Hastalığı Takibi

Yukarıda başlığı belirtilen Yüksek Lisans tez çalışmamın; Giriş, Ana Bölümler ve Sonuç Bölümünden oluşan, toplam 121 sayfalık kısmına ilişkin, 26 / 07 / 2022 tarihinde tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %3 'tür. Uygulanan filtrelemeler:

1. Kaynakça hariç

2. Alıntılar hariç

3. Beş (5) kelimeden daha az örtüşme içeren metin kısımları hariç

"Başkent Üniversitesi Enstitüleri Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Usul ve Esaslarını" inceledim ve bu uygulama esaslarında belirtilen azami benzerlik oranlarına tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Öğrenci İmzası:…….

ONAY

… / … / 20…

Dr. Öğr. Üyesi Tunç Aşuroğlu

TEŞEKKÜR

Çalışma sürecini ve benden beklenenleri açıklayan, karşılaşılan güçlüklerin aşılmasında her zaman yardımcı olan ve çalışmanın sonuca ulaştırılmasında değerli katkılarından dolayı tez danışmanım Sayın Dr. Öğr. Üyesi Tunç AŞUROĞLU'na sonsuz teşekkürlerimi ve saygılarımı sunarım.

Bu süreçte her zaman yanımda olan ve destekleyen aileme, öğreti ve katkılarından dolayı Behzad NADERALVOJOUD'a, süreci başlatan Ömürhan Avni SOYSAL'a, kitabına uygun olmasını sağlayan Mustafa Murat ARAT'a, beni geliştiren Ece KAPLAN ve Leyla SİVAS'a, motivasyonumu sağlayan Tansu YANIK TOKAT ve Beyza AKSONGUR'a çok teşekkür ederim.

ÖZET

Caner BOZKURT
MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KANSER HASTALIĞI TAKİBİ
Başkent Üniversitesi Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı
2022

Meme, akciğer, prostat ve mide kanserleri dünya genelinde en sık görülen kanser türleri olmuştur. Bu kanserlerin erken evrede tespiti ve teşhisi literatürde bir zorluk teşkil etmektedir. Hekimler kanser hastalarıyla uğraşırken, risk faktörü olan çeşitli tedavi yöntemleri arasından seçim yapmaktadır. Tedavinin riskleri faydalarından daha ağır basabileceğinden, klinik karar vermede tedavi programı kritik öneme sahiptir. Bu program hastanın önceki komorbiditilerine, aldığı ilaçlara ve geçirdiği tedavi prosedürlerine bakılarak hazırlanmaktadır. Hangi ilacın ve tedavinin kullanılacağına manuel olarak karar vermek çok zaman almakta ve zor olabilmektedir. Bu tez çalışmasında, meme, akciğer, prostat ve mide kanseri hastalarının hastane içi teşhis sonrası mortalite tahmini için tahmin oranını mümkün olduğunca yüksek tutan makine öğrenmesi yaklaşımları kullanılarak hesaplamalı bir çözüm sunulmuştur. Çözüm, elektronik sağlık sistemlerinden kolaylıkla elde edilebilen tanı, ilaç ve prosedür parametrelerinin analizine dayanmaktadır. Kanser hastalarının mortalite sonuçlarını tahmin etmek için sınıflandırmaya dayalı bir yaklaşım getirilmiş, model eğitimleri yapılmış ve bu sınıflandırıcıların performansları değerlendirilmiştir. Medical Information Mart in Intensive Care IV (MIMIC-IV) veri kümesi üzerinde Lojistik Regresyon, Karar Ağacı, Rastgele Orman, Destek Vektör Makinesi ve Çok Katmanlı Algılayıcı sınıflandırıcıları değerlendirilmiş ve bunlarla çeşitli deneyler yapılmıştır. Belirtilen kanser hastaları için tanı, ilaç ve tedavi özellikleri çıkarılmış ve Lojistik Regresyon ile ilişkili öznitelik seçimi yapılmıştır. Kolay erişilebilir elektronik sağlık verilerinin kullanılması ve yapılacak işlemlerin hafıza ve zaman kullanımı açısından hızlı ve etkin olabilmesi için az veri ile başarılı sonuç verecek şekilde sınıflandırıcı yapısı tasarlanmış ve doktorların yükünün azaltılması hedeflenmiştir. Makine öğrenimi modellerinin mortalite tahmin yetenekleri, F1 Makro Ortalaması ve AUC-ROC skor metrikleri ile değerlendirilmiştir. En iyi F1 skorları meme için 0.74, akciğer için 0.73, prostat için 0.82 ve

mide kanseri için 0.79 olarak bulunmuştur. En iyi AUC-ROC skorları meme için 0.94, akciğer için 0.91, prostat için 0.96 ve mide kanseri için 0.88 olarak bulunmuştur. Sonuç olarak, en ilişkili öznitelikler kullanılarak, çıkan sonuçların her kanser türü için temel sonucuna benzer olduğu görülmüş ve bu yaklaşımın, veri ve kaynağın sınırlı olduğu durumlarda sağlık tesislerinde kolaylıkla kullanılabileceği ortaya konulmuştur.

ANAHTAR KELİMELER: Kanser, MIMIC-IV, Makine öğrenimi, Mortalite tahmini, Elektronik sağlık kaydı

ABSTRACT

Caner BOZKURT
CANCER DISEASE TRACKING WITH MACHINE LEARNING METHODS
Baskent University Institute of Science and Engineering
Department of Computer Engineering
2022

Breast, lung, prostate and stomach cancers have been the most common types of cancer worldwide. Detection and diagnosis of these cancers at an early stage poses a challenge in the literature. When dealing with cancer patients, physicians choose among various treatment methods with risk factors. The treatment program is critical in clinical decision making, as the risks of treatment may outweigh the benefits. This program is prepared by looking at the previous comorbidities of the patient, the medications he took and the treatment procedures he had undergone. Manually deciding which drug and treatment to use can be time consuming and difficult. In this thesis, a computational solution is presented for breast, lung, prostate and stomach cancer patients' in-hospital post-diagnosis mortality prediction using machine learning approaches that keep the prediction rate as high as possible. The solution is based on the analysis of diagnostic, drug and procedural parameters that are easily available from electronic health systems. In order to predict the mortality outcomes of cancer patients, a classification-based approach has been introduced, model training has been carried out, and the performances of these classifiers have been evaluated. Logistic Regression, Decision Tree, Random Forest, Support Vector Machine and Multi Layer Perceptron classifiers were evaluated on the Medical Information Mart in Intensive Care IV (MIMIC-IV) dataset and various experiments were carried out with them. Diagnosis, drug and treatment features were extracted for the specified cancer patients, and feature selection related to Logistic Regression was made. In order to use easily accessible electronic health data and to make the procedures to be done quickly and effectively in terms of memory and time usage, the classifier structure was designed to provide successful results with less data and it was aimed to reduce the burden of doctors. The mortality prediction abilities of the machine learning models were evaluated with the F1 Macro Mean and AUC-ROC score metrics. The best F1 scores were 0.74 for breast, 0.73 for lung, 0.82 for prostate

and 0.79 for stomach cancer. The best AUROC scores were 0.94 for breast, 0.91 for lung, 0.96 for prostate and 0.88 for stomach cancer. As a result, using the most relevant features, it was seen that the results were similar to the main result for each cancer type, and it was revealed that this approach can be easily used in healthcare facilities where data and resources are limited.

KEYWORDS: Cancer, MIMIC-IV, Machine learning, Mortality prediction, Electronic health record

İÇİNDEKİLER

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

1. GİRİŞ

1.1. Motivasyon ve Problem Tanımı

Kanser, vücudun herhangi bir bölgesini etkileyebilen çok çeşitli hastalıkları kapsayan geniş bir terimdir. Kötü huylu tümörler ve neoplazmlar da kanser için kullanılabilen diğer terminolojilerdir. Kanser, normal hücrelerin genellikle kanser öncesi bir lezyondan çok aşamalı bir süreçte kötü huylu bir tümör hücresine dönüşmesi ile gelişir. Kanserin ayırt edici özelliklerinden biri, normal sınırlarının ötesinde büyüyerek vücudun diğer bölümlerini enfekte etmelerine ve diğer organlara yayılmalarına izin veren anormal hücrelerin hızla ortaya çıkmasıdır; bu metastaz olarak da bilinir. Kanser hastalarında en sık ölüm nedeni yaygın metastazdır [1].

2020'de dünya çapında tahminen 18,1 milyon kanser vakası olduğu belirtilmiştir [2]. Meme ve akciğer kanserleri 2020 yılında tüm yeni vakaların sırasıyla %12,5 ve %12,2'sini oluşturarak dünya genelinde en sık görülen kanser türü olmuşlardır. Prostat ve mide kanseri ise, 2020'de teşhis edilen tüm yeni vakaların sırasıyla %7,8 ve %6,0'ını oluşturan dördüncü ve beşinci en sık görülen kanser türü olmuşlardır [2]. 2020 yılında 2.261.419 yeni vaka ve 684.996 ölüm ile meme kanseri en sık görülen kanser olmuştur [3]. Bunu takiben akciğer kanseri, 2020 yılında 2.206.771 yeni vaka ve 1.796.144 ölüm sayısı ile sonuçlanmıştır [4]. Dördüncü sırada prostat kanseri 2020'de 1.414.259 yeni vakaya ve 375.304 ölüme sahiptir [5]. Bunu takiben mide kanseri, 2020'de 1.089.103 yeni vaka ve 768.793 ölüme sahiptir [6]. Vaka istatistiklerinden de anlaşılacağı üzere bu kanser türlerinin ölüm oranları oldukça yüksektir.

Hastaların ölüm oranını azaltmak için kanserin erken evrede tespiti çok önemlidir [7]. Hekimler ve araştırmacılar için kanserin tespiti ve teşhisi literatürde çeşitli zorluklar teşkil ettiği belirtilmiştir [26]. Kanserli hücrelerin tespiti esas olarak tıbbi görüntüleme ve laboratuvar testleri ile yapılmaktadır [8]. Bu prosedürler zaman almaktadır ve büyük miktarda işgücü gerektirmektedir [9].

Hekimler kanser hastalarıyla uğraşırken, her biri önemli risk taşıyan çeşitli tedavi yöntemleri arasında seçim yapmalıdır. Geç evre kanser hastalığı olan hastalar için mevcut tedaviler sadece küçük bir hayatta kalma şansı sağlayabilmektedir. Ek olarak, terapilerin tedavi edilen veya önlenen semptomlardan daha kötü olabilen yan etkileri de vardır [49].

Bazı tedavilerin sonuçlarının ortaya çıkması üç aya kadar sürebilir iken, olumsuz etkiler daha erken ortaya çıkabilmektedir ve altı haftaya kadar sürebilmektedir. Tedavinin riskleri faydalarından daha ağır basabileceğinden, tedavi kararlarının verilmesinde terapi takvimi kritik öneme sahiptir. Gereksiz toksisiteleri veya ikincil hasarı en aza indirmek için aşırı tedaviden kaçınılmalıdır [46]. Çok sayıda tedavi yaklaşımı yerine zamanında seçilen doğru tedaviler, tedavinin başarı oranını artırabilmektedir. Her hasta için ilaca karşı duyarlılık değişebileceğinden hassas tıp (Precision Medicine) da [43]-[45] kanser takibinde ve tedavisinde ön plana çıkmıştır. İlaç yanıtında olduğu gibi, bir hastanın komorbiditeleri (tanı) de hastane içi mortaliteyi etkileyebilmektedir, komorbiditeler de ilaç ve tedavi seçeneklerini değiştirebilmektedir.

Kanseri tedavi etmek için hangi ilaç ve tedavinin kullanılacağına manuel olarak karar vermek çok zaman almaktadır ve hekimler için sorunlu olabilir. Doktorlara kanser hastalarını değerlendirmede yardımcı olacak bir çözüm, makine öğrenimi yaklaşımlarıdır. Son çalışmalar [7]-[33], erken evre mortalite tahmininde önemli etkileri olduğunu gösterdikleri için ve kanser hastası verilerinin temel özelliklerini analiz etmek için makine öğrenimi yaklaşımlarına odaklanmışlardır. Makine öğrenimi yöntemleri, bir hasta verisinden temel bilgileri analiz edebilmekte ve çıkarabilmektedir. Ayrıca verilerden çeşitli örüntüleri öğrenebilir ve istenen sonuçları çok daha hızlı tahmin edebilmektedirler. Bu avantajlarından dolayı makine öğrenmesi yaklaşımları kanser araştırması alanında son yıllarda popülerlik kazanmıştır.

Makine öğrenimi yaklaşımları manuel karar vermeye göre daha hızlı olsa da, özniteliklerin sayısı arttıkça hesaplama süresini ve modelin gerektirdiği kaynaklar da artmaktadır. Dolayısıyla, bu sorunun üstesinden gelmek için modellere giren özniteliklerin sayısının az olması gerekmektedir [56]-[58]. Boyutsallığı makul bir şekilde azaltacak kadar küçük olmalı ve yine de tahmin görevlerinde kullanılmak üzere yeterli sonuçlar vermelidir.

Bir hastanın tanıları, ilaçları ve prosedürleri bir kanser hastasının hayatta kalıp kalmayacağını değerlendirmek için çok önemli öznitelikler olduğu birçok çalışmada belirtilmiştir [37]-[49]. Bu öznitelikler hastanenin elektronik sağlık veri tabanından kolaylıkla alınabilmektedir. Laboratuvar ölçümleri veya mikrobiyoloji sonuçları gibi diğer veriler çoğunlukla eksik veriler içerdiğinden, gerektirdiği ek ön işlemlerden ve toplanması zaman aldığından bu çalışmada kullanılmamıştır. Önceki çalışmalar bölümünde kullanım kapsamı ayrıntılı anlatılacak tanı, ilaç ve prosedürler bu çalışmanın temel verilerini oluşturmaktadır.

1.2. Önceki Çalışmalar

Literatürde kanserle ilgili tespit çözümleri için birçok çalışma bulunmaktadır. Şu anda, Medical Information Mart for Intensive Care IV (MIMIC-IV) veri kümesinde kanser vakaları ile mortalite tahmini üzerine önceden yapılmış herhangi bir çalışma bulunmamaktadır. Bu bölümde literatürdeki çalışmalar aşağıdaki gruplara ayrılarak açıklanmıştır:

- MIMIC'ten farklı veri kümeleri kullanılarak çeşitli kanser türleri üzerinden makine öğrenimi çalışmaları,
- MIMIC-III üzerindeki farklı kanser türlerinde yapılmış ilişkili diğer çalışmaları,
- MIMIC-IV üzerindeki farklı kanser türlerinde yapılmış ilişkili diğer çalışmaları.

Xie et al. [7] tarafından Çinli hastaların plazma metabolitlerini akciğer kanseri için tanısal biyobelirteçler (biomarkers) olarak keşfetmeye çalışılmış ve metabolitler ile makine öğrenimi yöntemlerini birleştirerek erken akciğer kanseri tanı biyobelirteçlerini saptamak için akciğer kanserine uygulanan disiplinler arası mekanizma kullanılmıştır. Erken teşhisin akciğer kanseri hastalarının hayatta kalma oranını iyileştirdiğini ve kan bazlı taramanın mevcudiyeti, erken akciğer kanseri hastalarının teşhisi artırabileceğini belirtmiştir. Akciğer kanseri için metabolik analizlerle daha spesifik ve duyarlı biyobelirteçlerin ortaya çıkarılması gerektiğini ve akciğer tümörü olan hastalarda 5 yıllık sağkalım oranının %18 ile düşük olduğunu ancak akciğer kanserinin erken teşhisi konulursa hayatta kalma oranı yaklaşık %55'e yükselebileceğini göstermiştir. Akciğer kanseri histolojik tip tahmini için, özellikle skuamöz karsinom ve adenokarsinom arasındaki ayrım, klinik pratikte önemli bir tanı gereksinimi olduğunu vurgulamıştır. 110 akciğer kanseri hastası ve 43 sağlıklı bireyden oluşan veri kümesi, Hubei Taihe Hastanesi'ne aittir. 61 plazma metabolitinin seviyeleri ile 6 metabolik biyobelirteç kombinasyonunun, birinci evre akciğer kanseri hastaları için uygun bulunduğu belirtilmiştir. Metabolik biyobelirteçlerin potansiyel kombinasyon şemaları lojistik regresyon analizine dayalı olarak yapılmış ve doğruluğunu artırmak için kontrol deneyi gerçekleştirmiştir. Tümörijenez ve tümör ilerlemesi ile ilgili işlevlere sahip en önemli 5 metabolik biyobelirteç olarak; taurin, Palmitoyll-karnitin, prolin, PE ve 2-DG bulunmuştur. Naive Bayes, erken akciğer tümörü tahmini için kullanılabilir bir araç olarak önerilmiştir.

Raoof et al. [8] tarafından Hindistan'da ve Dünya genelinde akciğer kanseri, nedenleri, semptomları, kansere bağlı ölüm oranı hakkında bir literatür araştırması yapılmış ve makine

öğrenimi teknikleri, sağlık hizmetlerindeki uygulamaları ve kanser prognozu ve tespiti incelenmiştir. Sağlık hizmetinin çeşitli alanlarında, Yapay Sinir Ağı (Artificial Neural Network) yaklaşımlarının meme kanseri, akciğer kanseri ve diğer ontoloji tahmini, tanı sistemleri, ilaç analizi alanlarında teşhis ve belirlemede yararlı olduğunu göstermiştir. Destek Vektör Makinesi (Support Vector Machine) yönteminin akciğer kanseri tespiti, teşhisi için kullanıldığını ve ayrıca akciğer kanseri olasılığını tahmin edebildiğini göstermiştir. Akciğer bölgesi ekstraksiyonundan önce akciğer bölgesi segmentasyonu yapıldığını, kanser nodül tespitinin Fuzzy Possibility CMean kümeleme algoritması ile gerçekleştiğini belirtmiştir. Ayrıca akciğer için görüntü işleme teknikleri ve intrapulmoner yapı segmentasyonu gri seviye eşiği, akciğer segmentasyonu için 3D etiketleme teknikleri ve matematiksel morfolojik tekniklerinin kullanıldığını göstermiştir. LUNA16 gibi büyük veri kümesinde Convolutional Neural Network (CNN) yönteminin görüntü analizinde, derin öğrenme yönteminin dil işleme görevinde ve Recurrent Neural Network (RNN) yönteminin metin analizi problemlerine uygulandığını göstermiştir. Çeşitli hastanelerden toplanan University of California Irvine (UCI) MLDb verileri ile akciğer kanserinden ölüme en çok sigara ve radon gazlarının neden olduğuna dikkat çekmişlerdir.

Cengil ve Çınar [9] tarafından derin öğrenme yöntemi ile SPIE-AAPM-LungX veritabanından Bilgisayarlı Tomografi görüntüleri kullanılarak akciğer nodülleri sınıflandırılmıştır. Kanser sınıflandırması için üç ayrı sinir ağı mimarisi kullanılmış ve veri işlemesi için evrişimli sinir ağları tercih edilmiştir. Kullanılan ağdaki katman sayısı arttıkça yöntemin başarısının da arttığını belirterek sistem eğitimi için 3B evrişim sinir ağı mimarisi kullanmıştır. Küçük boyutlu biyomedikal görüntülerden iyi huylu ve kötü huylu olarak akciğer nodülü sınıflandırması yapmasına rağmen yöntemin başarılı olduğunu gösterilmiştir.

Alam et al. [10] tarafından çok aşamalı Destek Vektör Makinesi sınıflandırıcısı ile akciğer kanseri saptama ve tahmin algoritması önerilmiş ve ayrıca akciğer kanseri olasılığı da tahmin edilmiştir. Akciğer hastalığının son aşamada ortaya çıkan evresinin, büyümenin akciğerde yayılma derecesine bağlı olduğunu, beklenmedik bir şekilde gelişen ve yayılan iki tür akciğer hastalığının, küçük hücreli akciğer maligniteleri ve küçük hücreli olmayan akciğer tümörleri olduğunu belirtmiştir. Yalnızca akciğer kanseri tespiti için yapılmış bu çalışmada eğitim için kullanılan veri kümesi, University of California Irvine (UCI) makine öğrenimi veritabanından alınmıştır. Çok aşamalı sınıflandırmada; giriş görüntüsünde kanserden etkilenen hücre yoksa algoritmanın akciğer kanseri olasılığını kontrol ettiğini, kanserden etkilenen hücre bulunursa da algoritmanın kanserin hangi aşamada olduğunu kontrol ettiğini açıklamış ve çıkan sonuçların yüksek olduğunu ortaya koymuştur.

Iyer et al. [11] tarafından önceden eğitilmiş VGG19 modeli kullanılarak, akciğer kanseri için Hint popülasyonlarına özgü PTEN, EGFR, ERBB2, BRAF ve CDKN2A mutasyonlarını içeren patolojik görüntülerden bilgi çıkarımı yapılmıştır. Genel olarak, tüm vakaların yüzde 80'i 'Küçük Hücreli Olmayan Akciğer Kanseri' ve kalan yüzde 20'sinin 'Küçük Hücreli Akciğer Kanseri' olarak sınıflandırmış ve tüm akciğer kanseri vakalarının yüzde 90'ının sigara içtiğini vurgulamıştır. Akciğer kanserinin epidemiyolojisi ve tedavilere verilen yanıtların farklı popülasyonlar arasında farklılık gösterdiğini, MiRNA'ların kanserle bağlantılı olduğunu ve kanser türlerinin sınıflandırılmasına yardımcı olduğunu belirtmiş ve bu sebeple seçilen genlere karşılık gelen görüntüleri, İnsan Protein Atlasından çıkarmıştır. 26 epoch ile çalıştırılmış derin evrişim ağı tabanlı görüntü sınıflandırma modeli kullanılmıştır. Genomik, epigenomik, çevresel ve metagenomiğe karşılık gelen single nucleotide polymorphism'lerin (SNP) akciğer kanserine yatkınlık tahmininde kullanılabileceğini ortaya koymuştur.

Patel and Nayak [12] tarafından görüntü kalitesi ve doğruluğu esas alınarak yerel enerji tabanlı şekil histogramı (Local Energy-based Shape Histogram - LESH) öznitelik çıkarma tekniği kullanılarak ve akciğer kanserini tespit etmek için LESH ve duyarlılık analizleri uygulanmıştır. Zaman faktörünün çok önemli olduğu erken teşhis ve tedavi aşamalarında görüntü iyileştirmenin önem kazandığını vurgulamıştır. The Japanese Society of Radiological Technology (JSRT) dijital görüntü veri tabanından göğüs radyografisi verileri seçilmiş ve 3 aşamada çalışma yapılmıştır. Kontrastı ayarlayan histogram eşitleme, görüntü ön işleme aşamasında kullanılmış, analizi kolay bir görüntü elde etmek için tüm görüntü birden çok bölüme ayrılmış ve LESH ile öznitelik çıkarma işlemi uygulanmıştır.

Wu and Zhao [13] tarafından akciğer kanserlerinin erken teşhisini kolaylaştırmak amacıyla bilgisayarlı tomografi (BT) görüntülerinden küçük hücreli akciğer kanserini (Small-Cell Lung Cancer - SCLC) tespit etmek için entropi bozunma yöntemi olarak adlandırılan bir sinir ağı tabanlı algoritma önerilmiştir. Makine öğrenimi tabanlı Bilgisayar Destekli Teşhis araştırmalarının, küçük hücreli olmayan akciğer kanserine odaklandığını ancak SCLC'li görüntünün olmayanla neredeyse aynı göründüğü için bunun tespiti konusunda az çalışma olduğunu vurgulamıştır. Eğitim ve test verileri, Ulusal Kanser Enstitüsü tarafından sağlanan yüksek çözünürlüklü akciğer BT taramalarından oluşmaktadır. Veritabanında 6'sı sağlıklı olan ve kalan 6'sı SCLC'li hastalardan gelen 12 akciğer BT taraması seçilmiştir. SCLC'li hastalarda tüm BT taramaları kanser hücrelerini göstermediğinden, akciğerin bulunduğu bölümler manuel olarak seçilmiştir. Model eğitiminde her gruptan 5 tarama alınarak vektörleştirilmiş histogram algoritmasına

beslenmiş ve kalan taramalar test amaçlı kullanılmıştır. Sonuç olarak %77,8 doğruluk elde edildiği belirtilmiştir.

Dekker et al. [14] tarafından 2 yıllık sağkalım tahmini için eksik verilerin olduğu veri setleri kullanılarak bir Bayes Ağı (BA) ve Destek Vektör Makinesi (DVM) modeli eğitilmiş ve performansları karşılaştırılmıştır. BA modeline, radyoterapi planına özel düğümler dahil edilmiş ve bu nedenle bu modeller, karar desteği olarak, yani tedavi planlama sürecinde uzak metastaz şansı ve komplikasyon arasında bir takas yapmak amacıyla kullanılabileceği belirtilmiştir. Eksik verilerin tıbbi alanda sıkça olduğu, bu nedenle makine öğrenimi modelleri, eksik veriler oluştuğunda bile tatmin edici performansa sahip olması gerektiğini vurgulamıştır. 322 akciğer kanseri hastasının ve eksik veriler açısından her birinin kendi özelliği olan üç ayrı harici veri kümesi de (35, 47, 33 hasta) kullanılmıştır. Modeller tümör boyutunu, klinik T ve N evresini, prognostik özellikler olarak lenf düğümlerini ve WHO performansını kullanmıştır. BA öğrenmesi sırasında tanımlanan özniteliklerle DVM modeli, Lagrange DVM sınıflandırması için sonlu bir Newton yöntemi aracılığıyla öğrenilmiştir. BA modelinin, Ghent ve Leuven kümesi özelinde DVM modelinden daha iyi performans gösterdiği ortaya konulmuştur. Ayrıca BA modelinde, lntumorload'ın en önemli çıkarım özniteliği olduğu belirtilmiş ve 0.77, 0.72. 0.70 Area under the ROC Curve (AUC) skorları ile tıbbi alanda kullanımının daha uygun olduğu sonucuna varılmıştır.

Shalini et al. [15] tarafından meme kanseri ve mamogram veri kümeleri ile, hastalığın tekrarlama sebebindeki kalıpları belirlemek için bazı derin öğrenme yöntemleri uygulanmıştır. Meme kanserinin nedeninin yaş faktörü, kadın hücresinde BRCA1, BRAC2 ve TP53 gen mutasyonu, meme kanserinin tekrarlaması, yoğun meme, erken evreye bağlı östrojen artışı, menopoz sonrası obezite, alkol alımı, veya dengesiz hormonal tedaviye bağlı olabildiği belirtilmiştir. University of California Irvine (UCI) makine öğrenimi veritabanından Meme Kanseri Wisconsin veri kümesi kullanılmış ve 699 örnek ve 10 öznitelik ile çalışılmıştır. Wisconsin'de veri kümesinde, kanser hücresinin kalınlığı, hücre boyutunun ve hücre şeklinin tutarlılığı, işaretleme özelliği, tek epitel hücre boyutu, sitoplazma kapsamı olmayan hücre, yumuşak kromatin oranı, nucleoli'nin görünürlüğünü gibi etkilenen hücre yapıları hakkında bilgiler içerdiği belirtilmiştir. Mamografik kütle veri kümesinde ise, mamogram değerlendirmesi, hastanın yaşı, kanser hücresinin şekli, kütle sınırı, kütle yoğunluğu ve 961 örneğinin iyi huylu veya kötü huylu olup olmadığı öznitelik bilgileri de bulunmaktadır. Yapay Sinir Ağı, Karar ağacı, Destek Vektör Makinesi ve Bayes ağı gibi çeşitli tahmin yöntemleri kullanılmıştır. Ayrıca hastalığın tekrarlama durumunu tahmin etmek için karar ağacı uygulanmış ve %72 doğruluk alındığı gösterilmiştir.

Thomas et al. [16] tarafından Wisconsin veri kümesinde birçok makine öğrenmesi yöntemi kullanılarak meme kanserini erken aşamada tahmin etmek için karşılaştırmalı bir çalışma sunulmuştur. 699 örnek ile çalışılmış ve 80:20 oranında eğitim-test kümesi belirlendiği belirtilmiştir. Örnek kod numarası gibi meme kanserini öngörmede ilişkisiz olan öznitelikler ön hazırlık sürecinde veri kümesinden çıkarılmıştır. En iyi doğruluğu %97 ile yapay sinir ağı tarafından verildiği ve bu nedenle kanseri tahmin etmek için kullanılabileceği ortaya konulmuştur.

Showrov et al. [17] tarafından Wisconsin veri kümesini kullanarak Yapay Sinir Ağı, Destek Vektör Makinesi ve Bayes ağı modelleri eğitilmiş ve bazı sendromlar üzerinden meme kanseri teşhisi yapılmıştır. Veri kümesinden 9 öznitelik sütunu ve 1 sınıf sütunu seçilmiş, giriş veri kümesinin boyutunu azaltmak için ReliefF algoritması kullanılmış ve 10 katlı çapraz doğrulama yöntemi uygulanmıştır. Sigmoid çekirdeğinin çok kötü bir sonuç verdiği ve Gaussian'ın meme kanseri veri kümesi için en iyi sonucu %95,86 doğrulukla verdiği ve Bernoulli kullanıldığında veri kaybı yaşandığı belirtilmiştir. Multinomial'in bu veri kümesi için %88 doğruluk ile ortalama sonucu verdiği, Radial basis function (RBF) değerinin İleri Beslemeli Sinir Ağından daha iyi olduğu ve RBF'nin doğruluğunun %95 olduğu açıklanmıştır. Veri kümesinin küçük olduğu ancak Doğrusal SVM'in bu veri kümesi için iyi çalıştığı belirtilmiştir. Bu dokuz algoritma karşılaştırıldığında, doğruluğu %96.72 ile en iyi model Doğrusal SVM, RBF Sinir Ağı ise en iyi ikinci doğruluğu verdiği ortaya konulmuştur.

Naveen et al. [18] tarafından Coimbra veri kümesi kullanılarak toplu makine öğrenmesi yöntemleri ile meme kanseri tespiti yapılmıştır. İzlenen ana adımların; öznitelik ölçekleme, çapraz doğrulama ve torbalama (bagging) tekniği ile çeşitli topluluk makine öğrenimi modellerinin eğitilmesi olduğu belirtilmiştir. Her torba farklı eğitilmiş model ile test edilip, tahmin sonucu için oylama yapılıp, sınıfın en yüksek tahmin sonucu modelinin tahmin sonuçlarını sonuca bağlandığı açıklanmıştır. Devamında doğruluk, karışıklık matrisi ve sınıflandırma raporları ile sonuçlar değerlendirilmiştir. Topluluk modelinin, önyargısız olarak sistem performansını iyileştirdiği, tüm modeller arasında, K-En Yakın Komşu (K Nearest Neighbor) ve karar ağacı algoritmasının, yüksek hassasiyet, geri çağırım ve F1 skoru verdiği gösterilmiştir.

Mishra et al. [19] tarafından kanserli görüntülerin sınıflandırılması ve kanserli görüntü olasılığını tespit etmek için öznitelik tabanlı makine öğrenmesi yöntemleri kullanılmıştır. Meme kanserinin erken teşhisinde, analitik kızılötesi termal görüntüleme uygulaması olan termografi yayılmacı ve iyonlaştırıcı olmayan çok güvenilir bir yöntem olduğu belirtilmiştir.

Visual Labs Database for Mastology Research (DMR) veri kümesinin kullanıldığı çalışmada Scale-Invariant Feature Transform (SIFT) ve Speeded Up Robust Features (SURF) öznitelik çıkarma teknikleri, çoğu vizyon girişimi ve nesne tespiti amacıyla kullanılmıştır. SIFT ile görüntülerden 128 öznitelik, SURF ile 64 öznitelik çıkarıldığı belirtilmiş ve daha sonra bunların eigen vektörleri alınarak, boyutsal olarak indirgenmiş veri kümesinin elde edildiği açıklanmıştır. Bu öznitelikler, parametrelerin daha iyi yorumlanması amacıyla Temel Bileşen Analizi kullanılarak daha da azaltılmıştır. Sonrasında farklı sınıflandırıcılara gönderilmiş ve sınıflandırıcıların etkinliğini belirlemek için doğruluk, özgüllük, duyarlılık, kesinlik ve F1 skoru gibi farklı parametreler ile değerlendirilmiş ve K-En Yakın Komşu (KNN) modelinin en iyi sonucu verdiği ortaya konulmuştur.

Bayrak et al. [20] tarafından Wisconsin veri kümesi ile meme kanserinin risk tahmini ve teşhisi için yapay sinir ağı ve destek vektör makinesi kullanılmıştır. Veri kümesinde 458 iyi huylu sınıf, 241 kötü huylu sınıf ile beraber; örnek kod numarası, küme kalınlığı, hücre boyutu tekdüzeliği, hücre şekli tekdüzeliği, marjinal yapışma, tek epitel hücre boyutu, yalın çekirdekler, yumuşak kromatin, normal çekirdekler ve mitoz öznitelikleri ile çalışıldığı belirtilmiştir. Waikato Environment for Knowledge Analysis (WEKA) aracı ile yapılan analizlerde doğruluk, kesinlik, hatırlama ve Receiver Operating Characteristic Curve (ROC) Alanı değerleri kullanılmış, destek vektör makinesinin en yüksek doğruluğu verdiği ortaya konulmuştur.

Bharat et al. [21] tarafından Wisconsin veri kümesi ile meme kanserinde iyi-kötü huylu olma durumu destek vektör makinesi kullanarak (DVM) sınıflandırma yapılmıştır. Kullanılan veri kümesi, bir meme kitlesinin ince iğne aspirasyon biyopsisinin sayısallaştırılmış görüntüsünden hesaplanan özellikleri içerdiği belirtilmiş ve belirli bir semptom setinin meme kanserine yol açıp açmadığını tahmin etmek için bir model analizi oluşturulmuştur. KNN, Naives Bayes ve Classification and Regression Tree (CART) ile eğitilip ve her bir algoritma için tahminin doğruluğu karşılaştırıldığında en iyi performansı verdiği görülmüştür. Girdi veri kümesinin standartlaştırılmasının, performansı artacağı vurgulanmış ve DVM'nin meme kanserinin tekrarlama tahmini için en uygun teknik olduğu gösterilmiştir. Test veri kümesinde %99,1 doğruluk elde edilen DVM'nin, tahmine dayalı analiz için güçlü bir teknik olduğu ve Gauss çekirdeği kullanıldığında meme kanserinin tekrarlama tahmini için en uygun teknik olduğu ortaya konulmuştur.

Khuriwal and Mishra [22] tarafından Wisconsin veri kümesini kullanarak teşhis edilen meme kanseri için uyarlanabilir topluluk oylama yöntemi önerilmiş ve değişkenler azaltılsa bile yapay sinir ağı ve lojistik regresyon (Logistic Regression) algoritmalarının nasıl daha

iyi çözüm sağladığını karşılaştırmıştır. Tüm süreç; veri ön işleme, öznitelik seçimi ve oylama modelleri olarak üç ana bölümden oluştuğu belirtilmiştir. Öznitelik seçimi için; Tek Değişkenli Öznitelik seçim yöntemi, çapraz doğrulama ve Özyinelemeli Öznitelik Seçim Algoritması kullanılmış ve en iyi 16 özniteliğin seçildiği açıklanmıştır. Tek değişkenli öznitelik seçim algoritmasında, negatif olmayan her özellik ve sınıf arasındaki istatistiklerini hesaplamak için Chi2 yöntemini kullanıldığı gösterilmiştir. Bu 16 öznitelik üzerinde lojistik regresyon ve sinir ağı algoritması ve sonuca göre son olarak oylama algoritması uygulanmış ve %98,50 doğruluk elde edildiği ortaya konulmuştur.

Kolay ve Erdoğmuş [23] tarafından University of California Irvine (UCI) veri kümesinden alınan meme kanseri verileri ile makine öğrenmesi yöntemleri kullanılarak kanser sınıflandırılması yapılmıştır. Yapay sinir ağı, kümeleme yöntemleri, karar ağaçları ile tahmin, öğrenme, sınıflandırma, kümeleme ve teşhis işlemleri yapıldığı belirtilmiştir. Genç hastalarda, erken teşhis aracı olarak, hastanın X-ışınına maruz kalmaması için ultrason ile muayene yapıldığı ancak orta yaşlılarda meme kanseri riski fazla olduğu için mamografi kullanıldığı açıklanmıştır. Breast Imaging Reporting and Data Systems (BIRADS) kullanılarak kanser riski 0-5 arası bir değerlendirme skoru aldığı belirtilmiştir. K-means ile Matlab'da bulunan dört uzaklık ölçüsüne için elde edilen sonuçlar RandIndex ile karşılaştırılmış ve en iyi kümeleme performansını Manhattan uzaklığının verdiği gösterilmiştir. Sonuç olarak, %79 doğruluk elde edildiği ortaya konulmuştur.

Gayathri and Sumathi [24] tarafından Wisconsin veri kümesi kullanılarak ve azaltılmış öznitelikler ile Lineer Diskriminant Analizinden (LDA) geçirilerek hastalık etkin bir şekilde teşhis edilmiştir. Boyutluluk sorunu, genellikle, verileri yüksek boyuttan düşük boyuta düzenlerken ortaya çıkması olarak açıklanmış ve öznitelik seçimi yapılmadan önce veri kümesinin normalize edilme gerektiği belirtilmiştir. Buradaki normalleştirmenin, verilerin anlaşılır olması adına tutarsız verilerin tutarlı verilere dönüştürülmesi anlamına geldiği ve veri temizleme işlemi gerçekleştirilirken, eksik veya yanlış verilerin keşfedilmesi olarak açıklanmıştır. Bu sebeple normalleştirme, yumuşatma, toplama ve verilerin Log10 değerleri hesaplanarak genelleştirme adımlarının uygulandığı gösterilmiştir. Bu öznitelikler LDA, KNN ve Çok Katmanlı Algılayıcı'ya (Multi Layer Perceptron) aktarılarak %92 doğruluk oranı elde edildiği ortaya konulmuştur.

Revett et al. [25] tarafından makine öğrenimini kullanarak 502 prostat kanseri hastası hakkında klinik bilgiler içeren tıbbi bir veri kümesi "rough set" tekniği ile araştırılmıştır. Veri kümesi, 27 eksik değerle karar özniteliği dahil olmak üzere 18 öznitelik içermektedir. Veri kümesinden fazla öznitelikler çıkarıldıktan sonra, sınıflandırma sürecinde rough set

kullanılmış, öznitelikler ve bunların değerleri karar sınıflarına eşlenmiştir. Oluşturulan kuralların, öznitelikleri ve değerlerini karar sınıflarına eşlediği belirtilmiştir. Eksik değerler, özniteliğin ait olduğu verilerin ortalaması ile doldurulmuştur. Evre, tedavi, yaş, Pf, Hx, Sbp, Dbp, Hg, tümör boyutu ve kemik metastazı öznitelikleri dikkate alınmıştır. Rough set'in temel amacı, benzer nesneler ve karşılık gelen bir karar sınıfı arasında bir eşleme olacak şekilde, her bir özniteliğin bilgi içeriğine bakarak karar tablosundaki gereksiz öznitelikleri azaltmak olduğu açıklanmıştır. Bu işlemin avantajı olarak, eşdeğerlik sınıfının herhangi bir üyesinin tüm sınıfı temsil etmek için kullanılabilmesi olduğu, böylece karar tablosundaki nesnelerin boyutsallığının azaldığı gösterilmiştir. Karar tablosundaki tüm nitelikler için Pearson Korelasyon katsayılarını hesaplanmış, Dtime özniteliğinin negatif korelasyon verdiği bulunmuş ancak belirli bir karar sınıfıyla güçlü ilişkisi olan bir öznitelik bulunamadığı açıklanmıştır. Sonuç olarak, yaklaşık %90 sınıflandırma doğruluğuna erişilmiştir.

Danilatou et al. [26] tarafından makine öğrenimi yöntemlerini kullanarak yoğun bakım hastaları üzerinde hastane içi ve taburculuk sonrası ölüm tahmini üzerine deneyler yapılmıştır. Venöz tromboembolizmin (VTE) yaygın kardiyovasküler durum olduğu belirtilerek, VTE tanısı konan bazı yüksek riskli hastaların, mortalite oranı yüksek olduğu için acil tedavi ve yoğun bakım ünitelerinde izlenmeye ihtiyaç duyulduğunu açıklamıştır. MIMIC-III veri tabanından yoğun bakımdaki 2468 VTE hastası kullanılarak, 1471 öznitelik çıkarılması yapılmıştır. 7 kategoriden oluşan veride; demografi, yoğun bakım ünitesinde kalış süresi, kabul sayısı, vücut ağırlığı, yaşamsal belirtiler, temel laboratuvar endeksleri, şiddet puanları, kan nakli gereksinimleri, prosedürler, ilaçlar ve mortalite bilgilerinin içerdiği belirtilmiştir. Synthetic Minority Over-sampling Technique (SMOTE) ve rastgele ormana dayalı ölüm tahmini kanser ve tromboz yaşı, analiz alt gruplarının çoğunda hem erken hem de geç ölüm için önemli tanımlayıcı olduğu vurgulanmıştır. Sınıf dengeleyici olarak kullanılmış Just Add Data BioMed (JADBIO) tarafından erken ölümleri tahmin etmek için seçilen en iyi makine öğrenimi modeli, Sapma ayırma kriteri ve minimum yaprak boyutu 2'ye eşit olan 500 ağaçlı Rastgele Orman olduğu gösterilmiştir. Ayrıca çıkarılan; ileri yaş, kanser, solunum, kardiyovasküler, böbrek hastalığı, vazopressör desteği ve mekanik ventilasyon özniteliklerinin yoğun bakım mortalitesinin iyi bilinen klinik belirleyicileri olduğu gösterilmiştir. Bireysel öznitelik analizi, varfarin, RDW ve kırmızı kan hücresi transfüzyonlarının erken ve uzun vadeli mortalitenin önemli belirleyicileri olduğu ortaya konulmuştur. Sonuç olarak; erken mortalitede 0.92 AUC skoru, geç mortalitede ise 0.82 AUC skoru alındığı belirtilmiştir.

Afrose et al. [27] tarafından dengesiz verilerin oluşturduğu önyargıları düzeltmek ve yeterince temsil edilmeyen alt popülasyonlar için tahmin doğruluğunu geliştirmek için çift öncelikli bir yöntem tasarlanıp, toplam dört tane ikili sınıflandırma görevi üzerinde çalışılmıştır. MIMIC-III ve Surveillance, Epidemiology, and End Results (SEER) kanser veri kümesi kullanılan bu çalışmada; hastane içi mortalite tahmini ve klinik tahmin kıyaslamasından dekompansasyon tahmini, 5 yıllık meme kanseri hayatta kalma tahmini ve 5 yıllık akciğer kanseri hayatta kalma tahmini yapıldığı gösterilmiştir. Dengesizlik kaynaklı tahmin eksikliklerinin çeşitli kategorilerini tanımlamanın yanı sıra, belirli azınlık demografik gruplarının tahmin doğruluğunu geliştiren bir yöntem olan çift öncelikli önyargı düzeltmesi öne sürülmüştür. Hastane içi mortalitede genel yaş spesifikliği, meme kanseri mortalite tahmininde daha zayıf olduğu ortaya konulmuştur. Diğer iki görev, 5 yıllık akciğer kanseri sağkalım tahmini ve 5 yıllık meme kanseri sağkalım tahmini için deneyler tekrarlanmış ve benzer kalıplar ile çift önceliklemenin mevcut örnekleme çözümlerinden 1,2 ila 58,8 kat daha fazla geri çağırma ve hassasiyet sağladığı gösterilmiştir.

Lee and Shin [28] tarafından üç klinik kıyaslama veri kümesi kullanılarak birleşik öğrenmenin (Federated Learning - FL) güvenilirliği ve performansı değerlendirilmiştir. Temel, dengesiz, çarpık ve bunların kombinasyonundan oluşan dört farklı deney, MIMIC-III ve EKG verileri kullanılarak gerçekleştirilmiştir. Dengesiz FL deneyi için, eğitim 1 ile 600 arasında değişen rastgele seçilmiş farklı boyutlarda veri kullanıldığı ve temel FL deneyi için, aynı boyuta sahip üç parçaya bölünmüş alt küme üzerinde tekrarlanmadan eğitildiği açıklanmıştır. Yerleşik yöntemlerin birçok cihazdan, kişiden veya kurumdan veri aktarımı gerektirmekle kalmadığı, aynı zamanda büyük veri kümeleri üzerinde eğitilmeleri gerektiğinden yüksek bir hesaplama maliyetine sahip olduğu vurgulanmıştır. İlk turlarda kullanılan az veriyle birlikte, FL'nin yerleşik yöntemlerden daha düşük performansa sahip olduğu, ancak sonraki turlarda, FL'nin performansının yerleşik yöntemlerin performansına benzer hale geldiği gösterilmiştir. MIMIC-III verileri kullanılarak hastane içi mortaliteye ilişkin temel FL, 0,85 AUC-ROC skoru ve 0,944 F1 skoru elde ettiği, dengesiz FL'nin ise 0,85 AUC-ROC skoru ve 0,943 F1 skoru elde ettiği ortaya konulmuştur.

Hammoud et al. [29] tarafından iki farklı veri kümesi üzerinde klinik durumun kötüye gideceğini işaret eden çeşitli olumsuz klinik olayların ve mortalitenin erken tahmini için bir erken uyarı modeli oluşturulmuştur. Güncel modellerin esas olarak belirli klinik olaylar için uyarlandığı ve diğer klinik olayları ne kadar iyi genelleyebileceğinin belirsiz olduğunu belirtmiştir. Kohortun Stony Brook Hastanesi veri kümesi ve MIMIC-III veri kümesi üzerinden seçildiği ve modelin ilk olarak her özniteliği birden fazla aralığa ayırdığı ve

ardından her bir özellikte alt küme seçmek için "lasso penalization" yaklaşımlı lojistik regresyon kullanıldığı açıklanmıştır. Modified Early Warning Scores (MEWS) ve Quick Sepsis-related Organ Failure Assessment (qSOFA) karşılaştırması sağlamak için, modelin erken uyarı skorlarından sadece belirli öznitelikler ile eğitildiği ve her hastane kalış verisinin bir dizi zaman vektörü olarak temsil edildiği vurgulanmıştır. COVID-19 pozitif ağırlıklı olarak seçilmiş veriler ile model; ventilasyon, yoğun bakım transferi, mortalite tahmini ve vazopresör ihtiyaç tahmini görevleri için erken uyarı skoru sağlayacak şekilde eğitildiği gösterilmiştir. Ayrıca, daha fazla öznitelik kullanımı ve düşük performans sonuçları göz önüne alınırsa daha karmaşık modellerin, eğitim kümesinde daha iyi performans gösterdiği ortaya konulmuştur. Sonuç olarak, qSOFA, MEWS ve tüm öznitelikler için sırasıyla 0.82, 0.84 ve 0.88 AUC skorlarının alındığı sunulmuştur.

Sauer et al. [30] tarafından yoğun bakıma kabul edilen kanser hastaları için kabul sayıları ve hasta özelliklerindeki eğilimleri açıklanmış, ortak değişkenler için düzeltme yapıldıktan sonra 28 günlük ve 1 yıllık mortalitenin 10 yıllık süre boyunca değişip değişmediği lojistik regresyon ile belirlenmiştir. 10 yıllık çalışma süresi boyunca, kanser hastalarının 28 günlük mortalitesinin %30 azaldığı ve bu eğilimin, ortak değişken ayarlaması yapıldıktan sonra da devam ettiği belirtilmiştir. Ayrıca MIMIC-III verileri üzerinden toplam 5102 kanserli hastadan ICD-9 koduna göre 3953'ünde onkolojik malignite, 1100'ünde hematolojik malignite ve 49'unda iki tanının da olduğu belirtilmiştir. Onkolojik hastalar ile kanser olmayan hastalar arasında benzer hasta karakteristiği olmasına rağmen; kanser hastalarının daha uzun yoğun bakım ve hastanede kalış süresine sahip oldukları ve daha yüksek 28 günlük mortalite oranına sahip oldukları gösterilmiştir. Acute Physiology and Chronic Health Evaluation (APACHE-III) skorunun dördüncü çeyreğindeki hastaların üçüncü çeyreğindeki hastalara göre 2 kat daha yüksek ölme olasılığının olduğu gösterilmiştir. Ayrıca Elixhauser skorunun dördüncü çeyreğindeki hastaların üçüncü çeyreğindeki hastalara göre 1,7 kat daha yüksek ölme olasılığının olduğu da gösterilmiştir. Sonuç olarak, kanser hastalarının, benzer klinik şiddet skorlarına sahip olmasına rağmen, 28 gün ve 1 yıl içinde ölme olasılıklarının daha yüksek olduğu ortaya konulmuştur.

Magna et al. [31] tarafından hastaların tıbbi geçmişleri kullanılarak meme kanseri teşhisi için doğal dil işleme yöntemi kullanan bir öneri sistemi geliştirilmiştir. Tıbbi dil işlemeye dayalı yöntemlerin; sağlık alanında klinik ontolojilerin kullanılması veya ICD-10'da verilen teşhisler ve isimler arasındaki anlamsal benzerliğin çıkarılması olduğu açıklanmıştır. Şili'deki bir hastaneden anonim hastaların tıbbi öyküleri ve MIMIC-III veri kümesindeki cinsiyet, klinik geçmiş, alışkanlık ve tıbbi teşhis kayıtları kullanılmış ve

bunların yalnızca %60'ının yazılı klinik bilgilerden oluştuğu belirtilmiştir. 126960 farklı kelimeden oluşan ifadeler tek bir sınıf olarak gruplandırılmış ve meme kanserine karşılık gelen hedef sınıfla karşılaştırılmıştır. Word2Vec, Bidirectional Encoder Representations from Transformers (BERT) ve Teerm Frequency-Inverse Document Frequency (TF-IDF) geleneksel kelime vektörü üretme yöntemleri ile rastgele orman ve KNN yöntemleri karşılaştırılmıştır. Sonuç olarak, 0.98 F1 skoru ile rastgele orman ve KNN algoritmalarının meme kanseri ve meme kanseri ameliyatları gibi çok benzer klinik geçmişleri olan vakalarda sınıflandırıcı olarak kullanılabileceği ortaya konulmuştur.

Wang et al. [32] tarafından meme kanseri hastalarının hastalığın ileride tekrarlama olasılığını tahmin etmek için elektronik sağlık verisi tabanlı doğal dil işleme modeli tasarlanmıştır. MIMIC-III'ten alınmış klinik notlar, doğal dil işleme araçları kullanılarak Concept birleşik tanımlayıcılara (Concept Unified Identifiers - CUI) eşleştirilmiştir. 6447 hastadan 446'sında ileride tekrarlama olduğu belirtilmiş, hastaların ilerleme notlarından ve patoloji raporlarından elde edilmiş kelime vektörleri ile CUI'lerin makine öğrenimi yöntemlerine beslendiği açıklanmıştır. Irk, etnik köken, medeni durum, sigara içme durumu, alkol kullanımı, ailede kanser öyküsü ve meme kanseri tanı yaşı öznitelikleri, tahmin performansını iyileştirmek için seçildiği belirtilmiştir. Sonuç olarak, Bilgi Güdümlü Evrişimli Sinir Ağı'nın (Knowledge-guided Convolutional Neural Network - K-CNN) 0,88 AUC skoru ve 0,5 F1 skoru aldığı ortaya konulmuştur.

Zeng et al. [33] tarafından Destek Vektör Makinesi yöntemi ile meme kanseri hastalarının hastalığın ileride tekrarlama olasılığını tahmin etmek için klinik notların ve elektronik sağlık verilerinin kullanıldığı bir model geliştirilmiştir. MIMIC-III'ten edinilmiş klinik metninleri ve 200 boyutlu word2vec kelime yerleşimi kullanılmış, yapılandırılmamış klinik notlardan 83 öznitelik ve yapılandırılmış klinik verilerden 18 özniteliğin birleştirildiği belirtilmiştir. Deneylerde, farklı öznitelik türleri üzerinde dört temel sınıflandırıcı eğitilmiş ve bunların; MetaMap tarafından etiketlenen tıbbi kavram kümesi, MetaMap tarafından etiketlenen filtrelenmiş tıbbi kavram kümesi, yapılandırılmış klinik veriler ve klinik notlardan standart sözcük kümesi olduğu anlatılmıştır. Sonuç olarak, eğitilmiş Destek Vektör Makinesi modeli ile 0,95 AUC skoru ve 0,78 F1 skoru alındığı gösterilmiştir.

Nowroozilarki et al. [34] tarafından MIMIC-IV veri kümesindeki sepsis hastaları üzerinden, hayatta kalma analiz yöntemi olan Boosted eXact Hazard Estimator with Dynamic Covariates (BoXHED 2.0) kullanılarak, gerçek zamanlı bir yoğun bakım içerisinde ölüm uyarı sistemi geliştirilmiştir. Makine öğrenimi modellerinin zamana bağlı veri kümeleri üzerindeki tahmin çözümlerini yeterli ölçüde iyileştiremediği belirtilmiş ve

performansın artırılması için elektronik sağlık verilerinden yararlanılması gerektiği vurgulanmıştır. Sepsis tahmin uygulamasındaki yaklaşım için veriler MIMIC-IV'ten kullanılmış ve bu risklerin güncellenmesi için kayan pencere yöntemi kullanıldığı belirtilmiştir. Elektronik sağlık verilerinden zamanla değişen öznitelikleri kullanarak klinik riski tahmin etmesi için yapılmış, gradyan destekli BoXHED yöntemi öne sürülmüştür. Bu yöntemin; üretilen risk ölçümleri son 8 saat boyunca belirli bir limitin üzerinde olan hastanın, yoğun bakımda öleceği sonucu ile işaretlenmesi şeklinde çalıştığı açıklanmıştır. Belirli bir zamanda ölüm oranı riski ölçümü üretmek üzere DeepHit modelinin buradaki yöntemin ölçütü olarak kullanılmasına uygun olmadığı belirtilmiştir. Burada bir hastanın belirlenmiş bir zamandan sonra yoğun bakımda ölüp ölmeyeceği araştırıldığı için kümülatif olasılığın uygun bir risk ölçütü olmadığı vurgulanmıştır. Sonuç olarak, hastanın risk ölçüm değerlerinden gerçek zamanlı ölüm tahmini oluşturan bu modelin 0.83 AUC-ROC skoru aldığı ortaya konulmuştur.

Meng et al. [35] tarafından MIMIC-IV veri kümesi üzerinden, hastane içi ölüm tahmini için derin öğrenme modellerinin yorumlanabilirliği ve tahmin doğruluğu analizleri ve veri kümesinin temsil edilmesindeki yanlılığı araştıran analizler gerçekleştirilmiştir. Öznitelik öneminin her bir veriye önem puanı atanmasıyla oluştuğu belirtilmiş ve sonuçların değerlendirilmesinin ikili sınıflandırma ile atanmış etiketin belirlenmiş problem için önemli olup olmadığına bakılarak yapıldığı açıklanmıştır. 122 elektronik sağlık verilerisi 5 demografik öznitelik, 4 kabul özniteliği, 33 komorbidite özniteliği ile MIMIC-IV veri kümesinden toplam 164 öznitelik kullanıldığı belirtilmiştir. Öznitelik önem tahminini değerlendirmek için, seçilen özniteliklerin kademeli olarak kaldırılmasıyla model performansındaki düşüş ölçülmüş ve veri dağıtımının tutarlılığını sağlamak için, değerlendirilen özniteliğin bilinen bir öznitelik ile değiştirildiği gösterilmiştir. Yorumlanabilirliğin; tüm model için küresel öznitelik önemi olduğu ve tahmin açısından tek bir örnek için yerel öznitelik önemi ile ilgili olduğu açıklanmıştır. Bu sebeple, ArchDetect modeli ile farklı tahmin modelleri tarafından verilen önemli öznitelikler araştırılmış ve karşılaştırılmıştır. Bununla; her bir veri örneği için sırasıyla yerel öznitelik önemi vermesi sebebiyle, küresel bir değerlendirme için yerel sonuçları toplandığı belirtilmiştir. Sonuç olarak, 164 öznitelik arasından etnik grup, cinsiyet ve yaş özniteliklerinde beyaz, erkek ve 78 yaşından büyük olmanın en yüksek öznitelik önemine sahip olduğu ortaya konulmuştur.

Önceki çalışmaların da gösterdiği gibi, kanseri tahmininde çeşitli makine öğrenme algoritmaları kullanılmış ve kanser başta olmak üzere birçok mortalite tahmin problemi

işlenmiştir. Ancak MIMIC-IV veri kümesi için bu kanser türleri ve bu öznitelik kümesi ile mortalite tahminini araştırmış hiçbir çalışma bulunmamaktadır.

1.3. Tez Çalışmasının Genel Katkıları

Bu çalışmanın yapıldığı sırada Medical Information Mart for Intensive Care IV (MIMIC-IV) veri kümesinde kanser vakalarında mortalite tahmini ile ilgili daha önce yapılmış bir çalışma bulunmamaktadır. Bu çalışmanın ilk katkısı; MIMIC-IV veri kümesinde mortalite tahmini için bu kanser türleri ile çalışılmasıdır. Bu nedenle bu çalışmadaki ana yaklaşım; meme, akciğer, prostat ve mide kanseri hastalarında hastane içi teşhis sonrası mortalite tahmini için tahmin oranını mümkün olduğunca yüksek tutan makine öğrenmesi yaklaşımlarının kullanılması ve kolayca elde edilebilir öznitelikleri bulmaktır. Bu amaçla, çeşitli öznitelik kümelerine sahip farklı makine öğrenmesi modelleri eğitilmiş ve bu sınıflandırıcıların performansları değerlendirilmiştir. Bu çalışma, belirtilen özniteliklerin sınırlı miktarda veri için makine öğrenimi yaklaşımlarıyla birlikte ne kadar iyi çalıştığını da araştırmaktadır.

Bu çalışmanın bir diğer katkısı; kolay erişilebilir elektronik sağlık verilerinin kullanılması ve yapılacak işlemlerin hafıza ve zaman kullanımı açısından hızlı ve etkin olabilmesi için az veri ile başarılı sonuç verecek şekilde tasarlanan sınıflandırıcı yapısı ile, doktorların yükünün azaltılmasıdır. Bu çalışmada, MIMIC-IV veri kümesinden hastanın tanıları, ilaçları ve prosedürleri kullanılarak, çeşitli kanser hastalarında hastane içi mortalite tahmini için en önemli öznitelikler Lojistik Regresyon ile seçilmiştir. Lojistik Regresyon, Karar Ağacı, Rastgele Orman, Destek Vektör Makinesi ve Çok Katmanlı Algılayıcı makine öğrenmesi modelleri ile çeşitli deneyler yapılmıştır. Makine öğrenimi modellerinin mortalite tahmin yetenekleri, F1 Makro Ortalaması ve AUC-ROC puan metrikleri ile değerlendirilmiştir.

1.4. Tez Planı

Tezin tamamı beş bölüm halinde düzenlenmiştir. Çalışmanın amaçları, kapsamı ve problem tanımı ile beraber önceki çalışmalar birinci bölümünde verilmiştir. Kullanılan veri kümesi, seçilen özniteliklerin neler olduğu, önemi, neden seçildiği ve son veriye

dönüşümündeki ön işleme adımları ikinci bölümde açıklanmıştır. Üçüncü bölümde, tasarlanan genel yöntem anlatılmış, öznitelik seçim aşamaları ve sebebi açıklanmış ve kullanılan makine öğrenmesi yöntemlerinin neler oldukları, nasıl çalıştıkları ve seçilme sebepleri sunulmuştur. Dördüncü bölümde, kullanılan değerlendirme ölçütleri, her bir kanser türüne uygulanmış yöntemlerin deneysel sonuçları ve özniteliklerin dağılım oranları verilmiştir. Altıncı bölüm, yapılan çalışmanın özetini, sonuçların yorumlanmasını ve gelecek çalışmalar için tavsiyeleri içerir.

2. TEZ ÇALIŞMASINDA KULLANILAN AÇIK ERİŞİM VERİ KÜMESİ

2.1. Medical Information Mart in Intensive Care (MIMIC) IV Veri Kümesi

Çalışma için, çok miktarda tıbbi tedavi verisi, halka açık olan Medical Information Mart for Intensive Care IV (MIMIC-IV v1.0) veritabanından erişilmiştir. MIMIC-IV, büyük bir üçüncü basamak hastanedeki yoğun bakım ünitelerine kabul edilmiş hastaların bilgilerini içeren geniş, tek merkezli bir veritabanıdır. MIMIC-IV, Boston, Massachusetts'teki Beth Israel Deaconess Tıp Merkezine (BIDMC) kabul edilen hastaların kimliği gizlenmiş olarak, gerçek hastane kalışlarını ve klinik verilerini içerir. Genellikle bir veri kullanım anlaşması kapsamında uluslararası araştırmacıların kullanımına açıktır. Veritabanı, yaşamsal belirtiler, ilaçlar, laboratuvar ölçümleri, grafikli gözlemler ve notlar, sıvı dengesi, prosedür ve tanı kodları, görüntüleme raporları, hastanede kalış süresi, hayatta kalma verileri vb. bilgileri içerir. Veritabanı, akademik ve endüstriyel araştırma, kalite geliştirme girişimleri ve yükseköğrenim kurslarından oluşan uygulamaları destekler [36].

MIMIC'e erişim elde etmek için, kişinin Collaborative Institutional Training Initiative (CITI) “Data or Specimens Only Research” kursunu bitirmesi ve sertifika alması gerekir. Kurs, Health Insurance Portability and Accountability Act (HIPAA) ve gizlilik korumaları dahil olmak üzere dokuz modülden oluşur. Kursu tamamladıktan sonra, PhysioNet'in barındırdığı klinik veritabanlarına sınırlı erişim talep etmek için veri kullanım sözleşme formu ve CITI “Sadece Veri veya Örnek Araştırma” eğitim programından tamamlama raporunun iletilmesi gereklidir. Talep başvurusunu gönderdikten sonra onay alındı ve MIMIC-IV v1.0 veritabanına erişim sağlandı. Veritabanı 2008 ve 2019 yılları arasında 500.000'den fazla acil servis ziyaretini kapsamaktadır. Veritabanı, kimliksizleştirilmiş olmasına rağmen, hastaların klinik bakımı hakkında BIDMC'nin MetaVision klinik bilgi sisteminden (iMDSoft) gelen kapsamlı bilgiler içerir [36].

Bu çalışmada ilgili veritabanına BigQuery ortamında eriştikten sonra, analiz için seçilen tablolar sorgulanmıştır. BigQuery, Google Cloud alt yapısında sunulan dağıtık, sunucusuz ve açık erişim bir veritabanı hizmetidir [60]. Tablo 2.1. de, bu iş için kullanılan tabloların listesini ve açıklamalarını içerir. Ayrıca, devamında BigQuery veritabanından veri çıkarmak için kullanılan örnek SQL sorguları ve genel veri çıkarma akışı sunulmaktadır.

Tablo 2.1. Kullanılan tablolar ve açıklamaları

Tablo Adı	**Açıklaması**	**Kayıt Sayısı**
diagnoses_icd	Hastaların tüm tanı kayıtlarını içerir	5.280.351
emar	Hastaya verilen ilacın kayıtlarını içerir	27.464.367
procedures_icd	Hastanın tüm prosedür kayıtlarını içerir	779.625
admissions	Hastanın hastane ziyaret kayıtlarını içerir	523.740

MIMIC-IV veri kümesinde 30 dan fazla tablo bulunmaktadır. Bu tablolar, her hasta için farklı türde verileri tutacak 5 farklı modüle bölünmüştür, bunlar; core, hosp, icu, ed ve cxr olarak adlandırılmıştır. Core modülü, hasta izleme verilerini içermektedir. Demografi, hastane kabulleri ve hastane içi bölüm transferleri burada açıklanmaktadır. Hosp modülü, hastane genelindeki elektronik sağlık kaydından elde edilen tüm verileri sağlamaktadır. Kapsanan bilgiler arasında laboratuvar ölçümleri, mikrobiyoloji, ilaç yönetimi ve faturalandırılmış teşhisler yer almaktadır. ICU modülü, yoğun bakım ünitesi içinde kullanılan klinik bilgi sisteminden toplanan bilgileri içermektedir. Belgelenmiş veriler intravenöz uygulamaları, ventilatör ayarlarını ve diğer grafik öğelerini içermektedir. ED modülü, acil serviste toplanan ilgili hastaların verilerini içerir. Buradaki bilgiler; hastanın kabul nedenini, triyaj değerlendirmesini, hayati belirtilerini ve ilaç mutabakatını içermektedir. CXR modülü, hasta tanımlayıcılarını bağlayan arama tabloları sağlar ve hasta göğüs röntgenlerinin analizinin diğer MIMIC-IV modüllerinden klinik verilerle ilişkilendirilmesine olanak tanımaktadır. Ayrıca değişken yapısı sebebiyle açık erişime sunulmamış, hastaların klinik not bilgilerinin serbest metin olarak tutulduğu bir note modülü de bulunmaktadır [36].

Bu çalışma için temel öznitelikler olarak core ve hosp modüllerinden aşağıdaki MIMIC-IV tabloları seçilmiştir: diagnoses_icd, emar, procedures_icd, ve admissions. Tablo diagnoses_icd, hastaların hastanede kaldıkları süre boyunca tıbbi eğitim almış kişiler tarafından belirlenmiş tüm tanılarını içermektedir. Verileri International Classification of Diseases (ICD) ICD-9 ve ICD-10 ontolojileri kullanılarak kaydedilmiştir. Tablo emar, yatak başı hemşire personeli tarafından doldurulmuş, belirli bir hastaya verilen ilacın yönetim kayıtlarını tutmaktadır. Electronic Medicine Administration Record (eMAR), ilaçların uygulama sırasındaki barkod taramasından gelen sonuç bilgileridir. Tablo procedures_icd, bir hastanın hastanede kaldığı süre boyunca faturalandırıldığı tüm prosedürleri tutmaktadır.

Verileri ICD-9 ve ICD-10 ontolojileri kullanılarak kaydedilmiştir. Tablo admissions, bir hastanın özgün hastane ziyaretlerini içermektedir. Bunlar; hastanın mevcut bilgileri, kabul ve taburcu olma zamanı bilgilerini, demografik bilgileri, kabulün kaynağı gibi bilgilerden oluşmaktadır [36].

Diagnoses_icd tablosu aşağıda belirtilen bilgileri tutmaktadır, tablo yapısı ve örnek veri Tablo 2.2. de gösterilmiştir. Subject_id, bireysel bir hastayı belirten benzersiz bir tanımlayıcıdır. Tek bir subject_id ile ilişkili tüm satırlar aynı kişiye aittir. Hadm_id, her hasta yatışı için benzersiz olan bir tamsayı tanımlayıcıdır. Seq_num, tanılara atanan öncelik bilgisidir. Öncelik, hangi tanıların "önemli" olduğunu bir sıra numarası ile gösterir. Icd_code, International Coding Definitions (ICD) kodudur. Bu kodlama sisteminin iki versiyonu vardır: versiyon 9 (ICD-9) ve versiyon 10 (ICD-10). Bunlar, icd_version sütunu kullanılarak ayırt edilebilmektedir. Genel olarak, ICD-10 kodları daha ayrıntılıdır, ancak ICD-9 kodlarını ICD-10 kodlarına dönüştüren kod eşlemeleri mevcuttur. Hem ICD-9 hem de ICD-10 kodları genellikle bir ondalık sayı ile sunulmuştur. Bu ondalık sayı, bir ICD kodunun yorumlanması için gerekli değildir; yani '0020' icd_code'u '002.0' ile eşdeğerdir [36].

Tablo 2.2. Diagnoses_icd tablosunun yapısı

Kolon Adı	**Tipi**	**Örnek Veri**
subject_id	INTEGER	19674536
hadm_id	INTEGER	22975928
seq_num	INTEGER	29
icd_code	CHAR(7)	5859
icd_version	INTEGER	9

Emar tablosu aşağıda belirtilen bilgileri tutar, tablo yapısı ve örnek veri Tablo 2.3. de gösterilmiştir. Emar_id, eMAR'da yapılan her emir için benzersiz bir tanımlayıcıdır. Emar_seq, eMAR siparişlerini kronolojik olarak numaralandıran ardışık bir tamsayıdır. emar_id, "subject_id-emar_seq" düzeniyle oluşur. Poe_id, emar'daki idareleri Provider Order Entry (POE) ve reçetelerdeki emirlere bağlayan bir tanımlayıcıdır. Pharmacy_id, emar'daki idareleri eczane tablosundaki eczane bilgilerine bağlayan bir tanımlayıcıdır. Charttime, ilacın verildiği saat bilgisidir. Medication, Uygulanan ilacın ad bilgisidir.

Event_txt, ilaç uygulaması hakkındaki bilgidir. Scheduletime, varsa, uygulamanın planlandığı saat bilgisidir. Storetime, ilaç uygulamasının eMAR tablosunda belgelendiği saat bilgisidir [36].

Tablo 2.3. Emar tablosunun yapısı

Kolon Adı	**Tipi**	**Örnek Veri**
subject_id	INTEGER NOT NULL	10146602
hadm_id	INTEGER NOT NULL	27382132
emar_id	VARCHAR(100) NOT NULL	10146602-464
emar_seq	INTEGER NOT NULL	464
poe_id	VARCHAR(25) NOT NULL	10146602-1081
pharmacy_id	INTEGER	90421304
charttime	TIMESTAMP NOT NULL	2185-01-05T22:00:00
medication	TEXT	Dexmedetomidine
event_txt	TEXT	Stopped - Unscheduled
scheduletime	TIMESTAMP	2185-01-05T22:00:00
storetime	TIMESTAMP NOT NULL	2185-01-06T04:02:00

Procedures_icd tablosu aşağıda belirtilen bilgileri tutar, tablo yapısı ve örnek veri Tablo 2.4. de gösterilmiştir. Seq_num, hastanede kalış sırasında işlemlerin gerçekleşme sırasıdır. Chartdate, ilgili prosedürlerin tarih bilgisidir [36].

Tablo 2.4. Procedures_icd tablosunun yapısı

Kolon Adı	**Tipi**	**Örnek Veri**
subject_id	INTEGER NOT NULL	11668016
hadm_id	INTEGER NOT NULL	26687342
seq_num	INTEGER NOT NULL	1
chartdate	DATE NOT NULL	2137-09-26
icd_code	CHAR(7)	016
icd_version	INTEGER	9

Admissions aşağıda belirtilen bilgileri tutar, tablo yapısı ve örnek veri Tablo 2.5. de gösterilmiştir. Admittime, hastanın hastaneye kabul edildiği tarih ve saat bilgisidir. Dischtime hastanın hastaneden taburcu olduğu tarih ve saat bilgisidir. Deathtime, eğer varsa hasta için hastanedeki ölüm zamanını tutar. Admission_type, 9 olasılık ile kabulün aciliyetini sınıflandırmak için kullanılan bilgidir. Admission_location, hastaneye gelmeden önce hastanın konumu hakkındaki bilgidir. Discharge_location, hastanın hastaneden taburcu olduktan sonraki konumu hakkındaki bilgidir. Insurance, language, marital_status ve ethnicity kolonları, belirli bir hastaneye yatış için hasta demografisi hakkındaki bilgilerdir. Bu veriler her hastaneye yatış için belgelendiğinden, kalıştan kalışa değişiklik gösterebilir. Edregtime ve edouttime, hastanın acil servise kaydedildiği ve taburcu edildiği tarih ve saat bilgilerdir. Hospital_expire_flag, hastanın belirtilen hastanede yatış süresi içinde ölüp ölmediğini gösteren ikili etikettir. 1 hastanede ölümü gösterir ve 0 hastaneden taburcu olana kadar hastanın hayatta kaldığını gösterir [36].

Tablo 2.5. Admissions tablosunun yapısı

Kolon Adı	Tipi	Örnek Veri
subject_id	INTEGER	10171525
hadm_id	INTEGER	21263495
admittime	TIMESTAMP(0)	2115-12-03T21:07:00
dischtime	TIMESTAMP(0)	2115-12-13T17:30:00
deathtime	TIMESTAMP(0)	null
admission_type	VARCHAR(40)	URGENT
admission_location	VARCHAR(60)	TRANSFER FROM HOSPITAL
discharge_location	VARCHAR(60)	PSYCH FACILITY
insurance	VARCHAR(255)	Medicaid
language	VARCHAR(10)	ENGLISH
marital_status	VARCHAR(80)	null
ethnicity	VARCHAR(80)	UNKNOWN
edregtime	TIMESTAMP(0)	2115-12-03T20:05:00
edouttime	TIMESTAMP(0)	2115-12-03T22:54:00
hospital_expire_flag	SMALLINT	0

2.2. Öznitelik Çıkarma ve Seçimi

2.2.1. Veri hazırlama ve kapsama kriterleri

Bu çalışmada ikili bayrak (binary flag) değeri olarak tanı, ilaç ve prosedür öznitelikleri kullanılmıştır. İkili bayrak; bir veri alanın 0 veya 1 olacak şekilde iki değerle tutulmasıdır [50]. Bunların her bir kanser türüne göre alınabilmesi için ilk olarak bu kanser hastalarının belirlendiği kümelerin oluşturulması gereklidir. Hastalara konulan tanılar ICD kodları ile tutulduğu için öncelikle kötü huylu meme, akciğer, prostat ve mide kanserlerine erişilmesi ve onların ICD kod bilgilerinin alınması gereklidir. Buna; tüm tanıların olduğu tabloya “malignant”, “breast”, “lung”, “prostate” ve “stomach” anahtar kelimeleri ile aşağıda gösterilen sorgulama yapılarak ulaşılmıştır (Şekil 2.1.).

```
select * from `physionet-data.mimic_hosp.d_icd_diagnoses`
        where  CONTAINS_SUBSTR (lower(long_title) , 'malignant') and
                ( CONTAINS_SUBSTR (lower(long_title) , 'breast') or
                CONTAINS_SUBSTR (lower(long_title) , 'lung') or
                CONTAINS_SUBSTR (lower(long_title) , 'prostate') or
                CONTAINS_SUBSTR (lower(long_title) , stomach))
```

Şekil 2.1. Kanser türlerini getiren sorgu

Sorgulama sonucu Tablo 2.6. de belirtilen ICD kodları ile kanser ilişki bilgileri bulunmuştur.

Tablo 2.6. Kanser türleri

Kanser Türü	**ICD-9 Kodu**	**ICD-10 Kodu**	**Literatür Adı**
Meme	174	C50	Malignant neoplasm of female breast
Akciğer	162	C33 / C34	Malignant neoplasm of trachea bronchus and lung
Prostat	185	C61	Malignant neoplasm of prostate
Mide	151	C16	Malignant neoplasm of stomach

Kanserlerin kod bilgilerine eriştikten sonra hangi hastalara bu teşhisin konulduğu bilgisi gereklidir. Bunun için tanı ve kabul tablolarına aşağıdaki sorgulama yapılarak çıkan sonuç "allDiagnosesAdmissions" adı ile ara bir tablo olarak kaydedilmiştir (Şekil 2.2.).

```
select diag.*, ad.admittime, ad.dischtime, ad.deathtime, ad.hospital_expire_flag
        from `physionet-data.mimic_hosp.diagnoses_icd`diag
        join `physionet-data.mimic_core.admissions` ad on ad.hadm_id = diag.hadm_id
```

Şekil 2.2. Hasta teşhislerini getiren sorgu

Hastanın hastaneye yatışı ile aynı gün ölmüş olması dışarıda tutularak, hastaların sadece ilk tanısı üzerinden kanser gruplarına ayrılmıştır. Burada, hastaneye kabul edildiği gün ölen hastalar üzerinde kanser hastalığını tedavi amaçlı herhangi bir ilaç ve prosedür uygulanmadığı için dışarıda bırakılmış ve modelin buradan muhtemel yanlış öğrenme hatası önlenmiştir. Benzer şekilde, bir hastaya hastalığı süresince aynı teşhis birden fazla ve farklı ICD versiyonlarında konulduğu görülmüştür. Bu sebeple kanser takibi için ilk konulan teşhis

dikkate alınmıştır. Aşağıda meme kanseri örneği üzerinden yapılan sorgulama ile belirtilen koşullardaki meme kanseri hastalarının tanı ve tekil hasta sayısı bilgileri elde edilmiş ve kendi kanser adı ile ayrı bir tablo olarak kaydedilmiştir (Şekil 2.3.).

```
select count(*) from (
        select *, row_number() over(partition by subject_id order by admittime asc) as r
                from `mimictez.allDiagnosesAdmissions`
                        where (icd_code like '174%'  and icd_version = 9) or
                        (lower(icd_code) like 'c50%'and icd_version = 10) ) A
where A.r = 1 and hospital_expire_flag = 0
```

Şekil 2.3. Kanser türlerine göre tanıları getiren sorgu

Benzer sorgulamalar diğer kanser türleri için de ilişkili ICD kodları kullanılarak yapılmış ve kendi kanser adı ile ayrı bir tablo olarak kaydedilmiştir. ICD-9 veya ICD 10 sürümleriyle beraber; göğüs kanseri için 3321 tanı ve 2065 farklı hasta, akciğer kanseri için 6677 tanı ve 3364 farklı hasta, prostat kanseri için 3112 tanı ve 1971 farklı hasta ve mide kanseri için 1146 tanı ve 583 farklı hasta bilgisi vardır.

Hastanın hastaneye yatışı ile aynı gün ölmüş olması hariç, hastaların sadece ilk tanısı dikkate alınmıştır. Bunun dışında başka herhangi bir daraltma kriteri uygulanmamıştır. Yine bu kriterlere göre aşağıdaki sorgulama yapılarak kohort bilgisi çıkarılmış ve hastaların ölüm sayısı bilgilerine erişilmiştir (Şekil 2.4.). Bu işlemler diğer kanser türleri için de tekrarlanarak meme, akciğer, prostat veya mide kanseri hastaları arasından kohort seçimi tamamlanmış ve verisi hazırlanmıştır.

```
select coh.admittime, ada.deathtime,
        from `mimictez.lung.lungInitialCohort` coh
        join `physionet-data.mimic_core.admissions` ada on coh.subject_id = ada.subject_id
                where ada.hospital_expire_flag = 1 and ada.deathtime > coh.admittime
```

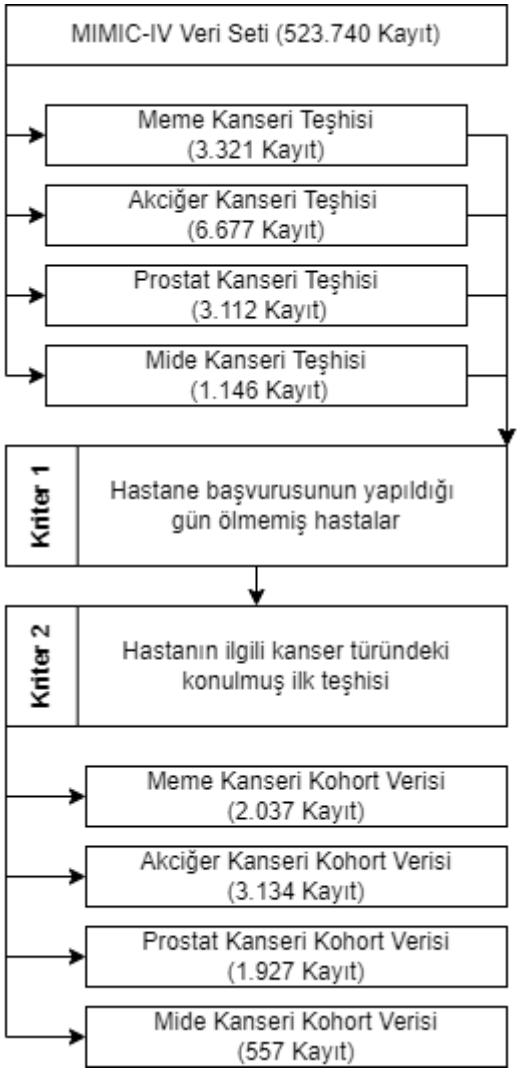
Şekil 2.4. Kohort bilgilerini getiren sorgu

Kohortta çalışma yapılabilecek; 2037 meme, 3134 akciğer, 1927 prostat, 557 mide kanseri hastasına ulaşılmıştır. Bu kohort için toplam ölen hasta sayısı: meme için 92, akciğer için 335, prostat için 121 ve mide kanseri için 63 dür. İşlem adımları boyunca elde edilen sayısal değerler Tablo 2.7. de gösterilmiştir.

Tablo 2.7. Kohortun sayısal değerleri

Kanser Türü	Tanı Sayısı	Tekil Hasta Sayısı	Başvuru Günü Ölmemiş Hasta Sayısı	Ölen Hasta Sayısı
Meme	3.321	2.065	2.037	92
Akciğer	6.677	3.364	3.134	335
Prostat	3.112	1.971	1.927	121
Mide	1.146	583	557	63

Verinin hazırlanmasındaki işlem akışı, seçim ve kapsama kriterleri Şekil 2.5. de özetlenmiştir.



Şekil 2.5. Veri hazırlığı işlem akışı

2.2.2. Kullanılan öznitelikler

Bu bölümde; kullanılan özniteliklerin önemi, neden seçildiği ve son veriye dönüşümü açıklanmıştır.

Aşağıdaki çalışmalarda [37]-[40]; hastane içi mortalite tahmininde bir hastanın tanılarının kullanılmasının, kanser sağkalımı, tedavi seçimi ve klinik uygulamalarda kullanımı ile yüksek oranda ilişkili olduğu belirtilmiştir. Bu çalışmalara göre kısaca; kemik metastazının, kişisel kanser deneyimlerinin, sigara içmek ve aile öyküsü arasında ilişkisi olduğu ve komorbidite yükünün tedavi seçimi ve mortalite oranında önemli olduğu gösterilmiştir.

Miao et al. [37] tarafından kemik ve kemik iliğinde sekonder malign neoplazmları olan hastalarda sağkalımı bağımsız olarak öngören faktörler belirlenmiştir. Tanı konulduktan sonra, kemik metastazlarının nadiren tedavi edilebildiğini, bunun da kanser hastalarında kısa vadeli bir prognoza işaret ettiği belirtilmiştir. Hasta kohortu MIMIC-III veri tabanından seçilmiştir ve bu hastaların genel sağkalımının tahminini iyileştirmek için prognostik bir nomogram geliştirilmiştir. Burada çoğu beyaz, evli ve sağlık sigortası öznitelikleri ile kemik ve kemik iliğinde sekonder malign neoplazmı olan 872 hasta vardır. Kullanılan eğitim kümesinde 610 ve doğrulama kümesinde 262 hasta ile çalışılmıştır. SOFA skorunun kanser hastalarının prognozunu değerlendirmede etkili olduğu gösterilmiştir. APACHE III ve SAPS II skorlarının aksine, daha yüksek SOFA skorları ve daha yüksek lojistik organ disfonksiyonu skorları, hem katı hem de hematolojik maligniteleri olan hastalarda artmış mortalite riski ile ilişkilendirilmiştir. OASIS puanlama sisteminin, kemik metastazı olan hastalarda 30 günlük mortalitenin önemli belirleyicileri olduğu gösterilmiştir. Ayrıca çok değişkenli Cox regresyon analizi, WBC sayısının ve koagülopati ile komorbiditelerin kemik metastazı olan hastalar için önemli bir bağımsız risk prognostik göstergesi olduğu bulunmuştur. Hayati belirtilerin değerlendirilmesi, kalp hızı, solunum hızı, vücut ısısı ve SpO2'nin kemik ve kemik iliğinde sekonder malign neoplazmları olan hastalar için bağımsız risk faktörleri olduğu ve düşük vücut ısısının artan riskle ilişkili olduğunu gösterilmiştir.

O'Rourke [38] tarafından prostat kanseri hastalarında diğer hastalıklarının klinik uygulamalarda kullanımı ile ilişkisini göstermiştir. Prostat kanseri tedavi karar sürecini etkileyen bir faktör olarak tedaviye bağlı potansiyel yan etkilerin etkisini tahmin etmenin zor olduğunu belirtmiştir. Amerika Birleşik Devletleri'nde hastalara radikal prostatektomi, radyoterapi ve bekleme yaklaşımı olarak üç ana tedavi seçeneği sunulurken; nihai tedavi

modalitesi seçiminin, tedavinin etkinliğine dayanması, yaşanabilecek olası komplikasyonlar ve tedaviye bağlı yan etkiler ve hastanın yaşam kalitesine ilişkin görüşleri dikkate alınması gerektiğini belirtmiştir. Brakiterapi tedavisi, kriyoterapi ve ışın tedavisi seçimleri ile ilgili varsayımsal görüşler sunulduğunda; erkeklerin doktorlarıyla birlikte kendi bağımsız kararları olarak değil kadınların tercihlerine göre çift kararları olarak görüldüğünü belirtmiştir. Sonuç olarak, bireysel hastalığa özgü veriler, yaş, kişisel değerler, dolaylı ve kişisel kanser deneyimleri ve uygun doktor-hasta ilişkisi gibi prostat kanseri tedavisinin seçiminde çeşitli faktörlerin olduğunu ortaya çıkarmıştır.

Osann [39] tarafından kadınlarda yapılan bir vaka kontrol çalışmasında, kadınlarda akciğer kanseri için sigara ve diğer faktörlerin önemi araştırılmıştır. Akciğer kanserlerinin histolojik dağılımdaki cinsiyet farklılıklarını açıklamak ve başka nedensel faktörlerin bulunabilmesi için vaka kontrol çalışmaları yapmıştır. Ailede akciğer öyküsü olan kişilerde kanser ve kronik obstrüktif akciğer hastalığı oluşumu gözlenmiştir. Ek olarak, kronik astımı veya alerjisi olan kişilerde beklenenden daha düşük bir akciğer kanseri oluşumu bildirilmiştir. Sigara ve aile öyküsü arasında önemli bir etkileşimin olduğunu ve ailesinde kanser öyküsü olan çok sigara içenlerin, aile öyküsü olmayan sigara içmeyenlere göre akciğer kanserine yakalanma olasılığının yaklaşık 50 kat daha fazla olduğunu bildirmiştir. Bronşit, pnömoni veya amfizem öyküsü olan kadınlar yüksek risk altındayken, astım veya saman nezlesi öyküsü, akciğer kanseri için önemli ölçüde daha düşük bir risk yaşamıştır. Kreyberg I vakaları arasında sigara içenlerin oranının daha yüksek olması nedeniyle, Kreyberg I tümörlü kadınların akrabalarının, Kreyberg II akciğer kanserli kadınların akrabalarına göre sigara içme olasılığının daha yüksek çıktığını ortaya koymuştur. Sonuç olarak, ailesel öykülerin hem akciğer kanseri hem de diğer akciğer hastalıklarının ortak bir öncüsü olduğu gösterilmiştir.

Piccirillo [40] tarafından baş ve boyun kanserli hastalarda aynı zamanda farklı seviyede kolorektum, akciğer, meme veya prostat kanserinin de olmasının komorbidite yükü gösterilmiştir. Ayrıca komorbiditenin genel sağkalım üzerindeki bağımsız etkisini ve ilk tedavi etkinliğinin değerlendirilmesinde komorbiditenin önemini göstermiştir. Kohort, 341 baş ve boyun, 307 kolorektum, 655 akciğer, 483 meme, 482 jinekolojik bölge ve 1.110 prostat kanseri dağılımı ile toplam 3.378 hastadan oluşmaktadır. Komorbidite yükünü karşılaştırmak için çok değişkenli analiz de dahil olmak üzere standart istatistiksel teknikler kullanılmıştır. Komorbiditenin bağımsız prognostik etkisini değerlendirmek için hayatta kalma teknikleri ve çok değişkenli lojistik regresyon analizi kullanılmıştır. Ayrıca baş ve boyun kanserli hastalar için farklı başlangıç tedavilerinin etkisini değerlendirmek için,

konjonktif konsolidasyon tekniği kullanılmıştır. Komorbiditenin baş boyun kanserli hastalarda önemli bir öznitelik olduğu gösterilmiş ve bu bilgilerinin eklenmesinin kanser istatistiklerinin değerini artıracağı ve kanser hastalarının bakımını iyileştireceğini belirtilmiştir. Sonuç olarak, şiddetli komorbiditesi olan hastalarda daha yüksek mortalite oranlarının sadece tedavi etkilerinden kaynaklanmadığı gösterilmiştir.

Komorbidite verilerinin; kanser sağkalımı, tedavi seçimi ve klinik uygulamalarda ilişkili olması ile bu çalışmanın motivasyonlarından biri olmuştur. Bununla birlikte, diagnoses_icd tablosundan icd_code verileri kullanılmıştır. Tanı olarak meme için 80855, akciğer için 158653, prostat için 93738 ve mide kanseri hastaları için 30141 tanı vardır. Aşağıda prostat kanseri üzerinden yapılan örnek sorgulama ile seçilen kohort içindeki hastalara konulmuş diğer teşhisler bulunmuştur (Şekil 2.6.). Aynı yöntem diğer kanser türleri için de tekrar edilerek ilgili teşhis sayıları elde edilmiştir.

```
select * from `mimictez.allDiagnosesAdmissions` diag
        join `mimictez.prostate.prostateInitialCohort` coh on coh.subject_id = diag.subject_id
                where diag.icd_code != coh.icd_code
```

Şekil 2.6. Hastaların diğer tanılarını getiren sorgu

Bu sayısal değerlere hastanın ilgili kanser türündeki teşhisleri dahil edilmemiştir. Buradaki amaç belirli kanser türü için başka hangi hastalığın etki edebileceğinin bulunmasıdır. Meme için 5625, akciğer için 6529, prostat için 5684 ve mide kanseri hastaları için 3193 tekil tanı bilgisi vardır (Tablo 2.8.).

Tablo 2.8. Kohort tanı sayısal değerleri

Kanser Türü	**Kohort Toplam Tanı Sayısı**	**Özgün Tanı Sayısı**
Meme	80.855	5.625
Akciğer	158.653	6.529
Prostat	93.738	5.684
Mide	30.141	3.193

Aşağıdaki araştırmalarda [41]-[45]; ilaca yanıt tahmininde, kanser sağkalımında ve hassas tıpta öznitelik olarak ilacın kullanılmasının kanser vakalarında hayati önem taşıdığına

işaret edilmiştir. Bu çalışmalara göre kısaca; genomik veya transkriptomik özniteliklerin, paklitakselin, diyabetin, düşmenin, paklitakselin ve immünoterapi ilaçlarının önemli olduğu gösterilmiştir.

Rafique et al. [41] tarafından makine öğrenimini kullanarak terapötik yanıt tahminindeki gelişmeleri ve klinik uygulama için terapi yanıtı tahminindeki zorlukları göstermiştir. İlaç duyarlılığını tahmin etmek için genomik veya transkriptomik özniteliklerin kullanılmasının yanı sıra, ilaçların kimyasal ve yapısal öznitelikleri öğrenme algoritmalarına dahil edilmiştir. Benzer bir gen ifade profiline sahip hücre hatları, belirli bir ilaca benzer tepkiler gösterirken, benzer kimyasal yapıya sahip ilaçlar, farklı hücre hatlarına karşı benzer inhibitör etkiler gösterdiğini açıklamıştır. İlaç hedefleri ve hücre dizileri arasındaki ilişkilerin dahil edildiği heterojen ağa dayalı bir yöntemin, bunlar arasındaki ilişkiyi daha iyi yakaladığını göstermiştir. NCI-ALMANAC, CCLE ve TCGA veritabanları üzerinde boyutu küçültülmüş ilaç kombinasyon verileriyle derin öğrenme, HOFM, autoencoder ve DeepGraph modelleri geliştirilmiştir. İlaç tanımlayıcıları ile birlikte gen ekspresyonu, mikroRNA ve proteom verilerinin kullanımının, en yüksek tahmin sonucu sağladığı gösterilmiştir. Ayrıca daha yüksek derece öznitelik kombinasyonlarının kullanılmasının tahmin performansını iyileştirebileceğini göstermiştir. Sonuç olarak, hassas tıp da ilaç yanıtını tahmin etmeye yönelik kullanılan modelde, ilaçlar için grafik verilerinin kullanılmasının tahmin performansını iyileştirdiğini belirtmiştir.

Brady et al. [42] tarafından kemoterapiye bağlı periferik nöropatinin (CIPN) metastatik meme kanseri (mBC) olan paklitaksel ve nab-paklitaksel kullanımına başlanan kadınlarda, klinik ve ekonomik sonuçlar üzerindeki etkisini değerlendirmiştir. CIPN'li hastalar ayrıca, CIPN'siz hastalarla karşılaştırıldığında, eşzamanlı ilaç karşılaştırmasında artan komorbidite yükünün ve daha fazla tedavi değişikliğinin dahil edildiğini belirtmiştir. Farklı kanserlerden hücre dizileri kullanılarak ilaç sinerjisini belirlemek için paklitaksel ve nab-paklitaksel kullanan mBC'si olan 5870 kadın ile çalışılmış ve bunların %42.7'sinde CIPN geliştiği gösterilmiştir. MarketScan veritabanı kullanılarak elde edilen bilgilerde CIPN'li ve CIPN'siz hastaların çoğunluğu paklitaksel ve nab-paklitaksel monoterapisi almıştır. Tedavi öncesi komorbiditeleri benzer olan kohortların takip sırasında CIPN olmayan hastalara kıyasla karaciğer fonksiyon bozukluğu, romatoid olmayan artrit, uykusuzluk, romatoid artrit, düşme ve kırık sorunları yaşadıklarını göstermiştir. CIPN'li hastaların, CIPN olmayan kohortla karşılaştırıldığında, depresyon, diyabet, uykusuzluk, karaciğer fonksiyon bozukluğu veya artrit geliştirme olasılığının daha yüksek olduğu ortaya konulmuştur. CIPN'li hastaların, CIPN'si olmayan kadınlara göre hastaneye yatış veya acil

servis ziyareti yapma olasılığı daha yüksek olduğu da gösterilmiştir. Bu bulgular kemoterapinin önemli uzun vadeli etkisini göstermiştir ve CIPN'nin hem hasta yükü hem de ekonomik etkisi hakkında bir tahmin sağlamıştır. Sağkalımın artması için CIPN gibi toksisitelerin daha uzun vadeli etkilerinin tedavi boyunca dikkate alınması gerektiği ortaya konulmuştur. Sonuç olarak, paklitaksel ve nab-paklitaksel tedavisi alan kadınlarda komorbidite gelişim ilişkisi gösterilmiştir.

Lin et al. [43] tarafından meme kanseri hastaları için oral anti-kanser ilaçlarının (OAM'ler) uyumunu artırmayı hedefleyen, değiştirilebilir psikososyal kolaylaştırıcılar ve bunların engelleri belirlenmiştir. Kadınlarda yaygın olan meme kanseri için OAM'lerin geliştirilmesi ve kullanımında artış olduğu ancak uyum oranlarının yetersiz olduğu, düşük hayatta kalma oranına yol açtığı, artan tekrarlama riskine ve daha yüksek sağlık hizmet maliyetlerine yol açtığı belirtilmiştir. Meme kanseri için OAM'lerin bulunması, ilaca uyum ve uyumun en az bir psikososyal yönü bulunabilmesi için sorgulama yapılmıştır. 1752 makale arasından; hasta-sağlayıcı ilişkileri ve ilaca ilişkin olumlu görüşler ve inançların olmasının ilaç uyumu açısından önemli olduğu belirtilmiştir. Ayrıca depresyon ve duyguların, hastalık algısının, yan etkiler endişesinin, ilaç yönetiminde ve karar vermede öz yeterliliğin, ilaç bilgisinin ve sosyal desteğin OAM üzerindeki etkisi olduğunu ortaya koymuştur. Sonuç olarak, OAM'lerin tedavi yönetim sorumluluğunu hastalara yüklediği, tedaviyi sağlayan ile sürekli iletişim ve hastalık hakkında eğitim gerektirdiğini ve buna uygun olarak tedavi planlaması yapılması gerektiğini belirtmiştir.

Deng and Nakamura [44] tarafından Kanser Hassas Tıp (CPM) sisteminde kullanılmış veya potansiyel olarak kullanılacak olan teknolojiler ve tedaviler gözden geçirilerek kanser hastalarında kullanılan hassas tıp iş akışı incelemesini öne sürmüştür. Temel ve klinik kanser araştırmalarındaki hızlanan ilerlemeyle birlikte, çok sayıda yeni keşif ve güçlü teknolojilerin, kanser hastaları için CPM uygulamalarını ortaya çıkardığını belirtmiştir. CPM sisteminin, kanser taramasını, tekrar ortaya çıkışının izlenmesini, etkili ilaçları, tedavi seçimini, tahmin edilmesini ve kişiselleştirilmiş immünoterapi dahil olmak üzere çok çeşitli kanser yönetimini kapsadığı belirtilmiştir. Sıvı biyopsi ile kanser tespiti, biyopsinin kolay olmadığı kanser türlerinde klinik olarak faydalı olduğu, ancak tedavi duyarlılığının yetersiz olduğu sadece hastalığın izlenmesinde faydalı olduğu belirtilmiştir. Kanser dokularında immün aktivite ile immün baskılama arasındaki dengenin immünoterapi başarısında belirleyici olduğunu ve iyi bir ilaç yanıtı için kanser hücrelerinde sitotoksik T lenfositlerin olması gerektiği vurgulanmıştır. Sonuç olarak, immünoterapinin bu hastaların klinik sonuçlarını ve yaşam kalitelerini iyileştirmeye katkıda bulunduğu gösterilmiştir.

Saarelainen et al. [45] tarafından medikal onkoloji polikliniğine başvuran hastalarda Potansiyel olarak uygunsuz ilaç (Potentially Inappropriate Medication - PIM) kullanım yaygınlığı ve ilişkili faktörleri araştırılmıştır. PIM kullanımının, ilacın kötü etkilerinde, hastaneye yatışta ve mortalitede artış ile ilişkilendirildiği belirtilmiş ve en yaygın beş PIM sınıfının, benzodiazepinler, trisiklik antidepresanlar, alfa-adrenoreseptör antagonistleri, iticiler ve steroidal olmayan antienflamatuar ilaçlar olduğu açıklanmıştır. 70 yaşından büyük 385 hastada yapılan çalışmada ilaç kullanımı, tanıları, önceki altı ayda kişinin bildirdiği düşme sayıları, ağrı ve acı bilgileri kullanılmıştır. 1 den fazla PIM kullanan 102 hastanın yöntem hesaplamalarında; lojistik regresyon, olasılık oranı ve PIM'lerin kullanımıyla ilişkili faktörler için %95 güven aralığı kullanılmıştır. PIM kullanmayan hastalarla karşılaştırıldığında, PIM kullanan hastalarda halsizlik, acı ve düşme yaşama olasılıklarının yükseldiği gözlemlenmiştir. Ayrıca PIM kullanımı ile polifarmasi arasında bir ilişki bulmuşlardır. Sonuç olarak, kanserli yaşlı hastaların PIM kullanımının, hastada dayanıksızlık oluşturduğu ortaya konulmuştur.

İlaç verilerinin; kanser sağkalımı ve hassas tıpla ilişkili olması ile bu çalışmanın motivasyonlarından biri olmuştur. Bunların ışığında emar tablosundan alınan medication verileri kullanılmıştır. İlaç sayısı meme için 392227, akciğer için 775811, prostat için 442399, mide kanseri hastaları için 205390'dır. Aşağıda mide kanseri üzerinden yapılan örnek sorgulama ile seçilen kohort içindeki hastalara reçete edilen ilaçlar bulunmuştur (Şekil 2.7.). Aynı yöntem diğer kanser türleri için de tekrar edilerek ilgili ilaç sayıları elde edilmiştir.

```
select * from `physionet-data.mimic_hosp.emar` emar
        join `mimictez.stomach.stomachInitialCohort` coh on coh.subject_id = emar.subject_id
```

Şekil 2.7. Hastaların ilaçlarını getiren sorgu

Meme için 992, akciğer için 1106, prostat için 982 ve mide kanseri hastaları için 659 tekil ilaç bilgisi vardır (Tablo 2.9.).

Tablo 2.9. Kohort ilaç sayısal değerleri

Kanser Türü	Kohort Toplam İlaç Sayısı	Özgün İlaç Sayısı
Meme	392.227	992
Akciğer	775.811	1.106
Prostat	442.399	982
Mide	205.390	659

Aşağıdaki çalışmalarda[46]-[49]; kanser sağkalımında, klinik kararda ve tedavi seçiminde yararsız tedavi uygulamasının önlenmesinin çok büyük bir rol oynadığı belirtilmiştir. Bu çalışmalara göre kısaca; kemoterapinin, radyasyon tedavisinin, immünoterapinin ve tedavi sıralamasının önemli olduğu gösterilmiştir.

Ali et al. [46] tarafından mesane kanserli hastalarda semptomları yönetmenin etkili bir yolu olan palyatif pelvik radyasyon tedavisinin (PRT) etkinliğini araştırılmış ve tedavi sonucu ile ilişkili faktörler belirlenerek hastanın yararsız tedavi görmesini engelleyen faktörler vurgulanmıştır. Hastalar yaş, evre, performans durumu, komorbiditeler, önceki kemoterapi, önceki radyasyon tedavisi ve radyasyon tedavisi protokolüne göre sınıflandırılmış ve radyasyon tedavisinden 6 hafta sonra takibe alınmıştır. 241 PRT alan hasta üzerinde çalışılarak haftalık hipofraksiyone radyasyon tedavisinin kullanımının, yaşlı hastalar da dahil olmak üzere, kabul edilebilir sonuçlarla iyi tolere edildiği gösterilmiş ve tedavisi sonrası 30 günlük mortalite oranının %18 olduğu belirtilmiştir. Sonuç olarak, iyi performans durumu ve erken evre hastalığı olan hastaların daha uzun süre hayatta kaldığı gösterilmiş ve hasta seçimi ve kapsamlı değerlendirmenin, yararsız tedavi uygulanmaması için çok önemli olduğu ortaya konulmuştur.

Choudhury and Nakamura [47] tarafından hasta seçimi ve izleme ile immün kontrol noktası inhibisyonuna özgü zorluklar ve bunları gidermeye yönelik yaklaşımlar gösterilmiş ve ayrıca, anti-immün kontrol noktası tedavisinde seçim ve izlemeyi yönlendirmek için ortaya çıkan immüno farmakogenomik alanının uygulamaları belirtilmiştir. Melanom'a ek olarak, bağışıklık kontrol noktası blokajı, metastatik mesane kanseri ve uyumsuz onarım eksiklikleri olan kolorektal kanserler dahil olmak üzere diğer kanser türlerinde önem kazandığı vurgulanmıştır. Pembrolizumab ile tedavi edilen hastalarda yapılan birinci faz çalışmada, tümör hücrelerinde %50 pozitif boyama oranına ulaşıldığı ve hücrelerin 8 aydan fazla hayatta kaldığı gösterilmiştir. Sonuç olarak, bağışıklık kontrol noktası blokajı tedavisinin, antikanser stratejisinde önemli olduğu ortaya konulmuştur.

Schonberg et al. [48] tarafından erken evre meme kanserli 80 yaş üstü kadınlarda meme kanseri tümör öznitelikleri, alınan ilk tedavileri ve sağkalımdaki farklılıkları daha genç kadınlara kıyaslayarak incelenmiştir. Evre içinde yapılan analizlerde, tümör ve sosyodemografik öznitelikler, alınan tedaviler ve komorbiditeler hastaların ölüm nedenleri incelemek için Cox orantılı tehlike modellerinde kullanılmıştır. 49616 kadın ile yapılan çalışmada, alınan tedavi türlerinin, yaş ve komorbidite ile önemli ölçüde ilişkili olduğu ortaya çıkarılmıştır. Agresif tedavi yöntemlerinden olan kemoterapinin sadece en kötü durumda ve en yaşlı olan hastalara uygulanması gerektiği belirtilmiştir. En yaygın tedavi yöntemleri olarak mastektomi ve lumpektomi (BCS) sadece doğru yaş grubundaki hastalara yapıldığında tedavinin etkili olduğu ve farklı yaş gruplarındaki hastalarda benzer tümör karakteristiği olduğu için yaşın sağkalımda tek başına etkili olmadığı gösterilmiştir. Sonuç olarak, yaşa ve hastanın genel durumuna bağlı olarak doğru tedavi seçiminin meme kanseri hastalarında sağkalım oranını artırdığı ortaya konulmuştur.

Simes [49] tarafından ileri yumurtalık kanserinde tedavi seçimini örnek olarak kullanarak optimal tedavi seçimi metodolojisinin güçlü ve zayıf yönleri değerlendirilmiştir. Optimal tedavi seçimi; kronik hastalığı, ilerlemiş kanser hastaları için, her tedavinin risklerini ve yararlarını tartarken dikkatli bir şekilde düşünmesi şeklinde tanımlanmıştır. İstatistiksel karar teorisinin bu tür problemlere uygulanması, riskler ve faydalar hakkındaki bilgileri, yaşam kalitesi konularında bireysel hasta tercihleriyle birleştirmenin açık ve sistematik bir yolunu sağladığı vurgulanmıştır. Genellikle yanıt veya sağkalım açısından daha umut verici olan terapilerin daha toksik olduğu, bu nedenle tedavi seçimindeki son kararın, yaşam kalitesi ile yaşam miktarı arasında veya bir hastalık durumu ile bir diğeri arasındaki dengenin korunması açısından dikkatlice verilmesi gerektiği belirtilmiştir. Karar teorisinin; klinik karar alınırken, tıbbi literatürü değerlendirmedeki zorlukların azaltılmasına yardımcı olduğu ve eleştirel değerlendirmenin önünü açtığı belirtilmiştir. Karar ağacı kullanılarak, ilerlemiş yumurtalık kanseri için önerilen tedavinin, büyük ölçüde sağkalım tahminlerine bağlı olduğu, ancak diğer olasılık tahminlerine veya yardımcı programları elde etme yöntemine daha az bağlı olduğu bulunmuştur. Sonuç olarak, tedavi seçimlerinde avantajların olabileceği ancak hayatta kalma sayısı ve kalitesi arasında takas olabileceği ortaya konulmuştur.

Prosedür verilerinin; kanser sağkalımıyla, klinik kararda ve tedavi seçimiyle ilişkili olması ile bu çalışmanın motivasyonlarından biri olmuştur. Bunlar göz önünde bulundurularak procedures_icd tablosundan icd_code verileri kullanılmıştır. Prosedür işlem sayısı meme için 10250, akciğer için 20160, prostat için 11084, mide kanseri hastaları için

5421'dir. Aşağıda meme kanseri üzerinden yapılan örnek sorgulama ile seçilen kohort içindeki hastalara uygulanan tedavi prosedürleri bulunmuştur (Şekil 2.8.). Aynı yöntem diğer kanser türleri için de tekrar edilerek ilgili prosedür sayıları elde edilmiştir.

```
select * from `physionet-data.mimic_hosp.procedures_icd` pro
        join `mimictez.breast.breastInitialCohort` coh on coh.subject_id = pro.subject_id
```

Şekil 2.8. Hastaların prosedürlerini getiren sorgu

Meme için 1686, akciğer için 2307, prostat için 1802 ve mide kanseri hastaları için 991 tekil prosedür bilgisi vardır (Tablo 2.10.).

Tablo 2.10. Kohort prosedür sayısal değerleri

Kanser Türü	**Kohort Toplam Prosedür Sayısı**	**Özgün Prosedür Sayısı**
Meme	10.250	1.686
Akciğer	20.160	2.307
Prostat	11.084	1.802
Mide	5.421	991

Tamamı birleştirildiğinde oluşan veri model eğitimlerinde kullanılacak olan tekil öznitelik listesini oluşturmaktadır. Meme için 8303, akciğer için 9942, prostat için 8468 ve mide kanseri hastaları için 4843 tekil öznitelik bilgisi oluşmuştur (Tablo 2.11.).

Tablo 2.11. Kohortun öznitelik sayısal değerleri

Kanser Türü	**Tanı Sayısı**	**İlaç Sayısı**	**Prosedür Sayısı**	**Toplam Öznitelik Sayısı**
Meme	5625	992	1686	8.303
Akciğer	6529	1106	2307	9.942
Prostat	5684	982	1802	8.468
Mide	3193	659	991	4.843

2.2.3. Ön işleme adımları

Bu bölümde; kullanılan özniteliklerin biçimlendirme prosedürleri, veri temizliği işlemleri ve öznitelik çıkarma yöntemi açıklanmıştır.

Aynı özniteliğin farklı yazım şekilleri ile eğitim modeline verilecek olması girdi boyutunu büyüterek amaçlanan performans kazanımını kötü yönde etkileyecektir. Bu sebeple aynı verinin farklı yazılışlarının tek bir ifadede gruplanması hem girdi boyutunu küçültecek hem de ilgili özniteliğe makine öğrenmesi sonrası atanacak ağırlık katsayılarındaki hatayı önleyecektir.

Veritabanı incelendiğinde özellikle ilaç verilerinde benzer değerlerin farklı formatta yazıldığı görülmüştür. Burada aynı değer için; tamamı büyük, tamamı küçük, bazı harflerin büyük veya sonunda fazla boşluk karakteri bulunarak, verinin karışık formatta tutulduğu görüldü. Tanı ve tedavi verileri için de, ICD-9 ve ICD-10 versiyonlarındaki yazımlar, benzer şekilde farklıydı, bu da özniteliklerin yanlış gruplanmasına sebep oluyordu. Boş (NULL) değerler de herhangi bir gerçekleşme bilgisi içermediği için onların da gruplamaya dahil edilmesi ilişkisiz veri yaratıyordu. Teşhis ve prosedürler benzer ICD kodlarını kullandığı için; karışıklığı önlemek ve hatalı işlemin engellenmesi adına bunların da birbirinden ayırt edilebilmesi gerekmektedir.

Özniteliklerin doğru bir şekilde gruplanabilmesi için aşağıdaki veri temizleme ve biçimlendirme prosedürleri uygulandı:

- ilk olarak boş değerler çıkarılmış,
- ikinci olarak veri sonunda veya başında bulunan fazla boşluk karakteri kaldırılarak veriler kısaltılmış,
- üçüncü olarak veriler küçük harf olarak formatlanmış,
- son olarak tanı, ilaç ve prosedür öznitelikleri değerlerine sırasıyla diag-, med- ve pro- ön ekleri eklenmiştir.

Boşluk karakteri temizliği ve küçük harfe formatlama işlemleri aşağıda akciğer kanseri örneği üzerinden belirtilen sorgular ile yapılmıştır (Şekil 2.9.). Daha sonra diğer kanser türleri için de aynı işlemler gerçekleştirilmiştir ve tanı, ilaç ve prosedür öznitelikleri değerlerine sırasıyla diag-, med- ve pro- ön ekleri eklenmiştir.

```
update `mimictez.lung.l_emar`
        set medication = trim(medication), medication = LOWER(medication) where TRUE
update `mimictez.lung.l_diag`
        set icd_code = trim(icd_code),  icd_code = LOWER(icd_code) where TRUE
update `mimictez.lung.l_procedures`
        set icd_code = trim(icd_code), icd_code = LOWER(icd_code) where TRUE
```

Şekil 2.9. Veri formatlaması yapan sorgu

Tüm hastalar için, karşılaşılan her tanı, ilaç ve prosedür özniteliği için, o kanser türüne özgün bir özellik sözlüğü oluşturacak bir haritaya yerleştirildi. Burada özniteliklerin gruplu olması ve sırasının bozulmaması için indeks ataması yapıldı. Böylece belirli bir sayı indeksi ile sorgulama yapıldığında her zaman aynı özniteliğe ulaşılabilir olmutur. Daha sonra satır sütun ilişkisi yaratılarak her hasta için bu öznitelikler sütun olarak atandı. Atanan sütunlar için genel hasta-öznitelik listesi gezilerek, hasta için ilgili öznitelikle karşılaşılırsa değeri 1 artırılacak şekilde bir algoritma çalıştırıldı. Böylece ilişkili özniteliğin seçili hastadaki frekansı çıkarılmış oldu. Algoritma sonunda karşılaşılmamış öznitelikler için 0 ataması yapıldı.

Aşağıdaki çalışmalarda [54],[55]; hesaplama yükünü ve zamanını azaltmak için öznitelik verilerinin modele gönderilmeden önce ikili bayrak değerine dönüştürülebileceği belirtilmiştir. Ayrıca, verilerin tek bitlik gösteriminin hızlı sonuçlar ve düşük kaynak maliyetleri ile sonuçlandığı gösterilmiştir.

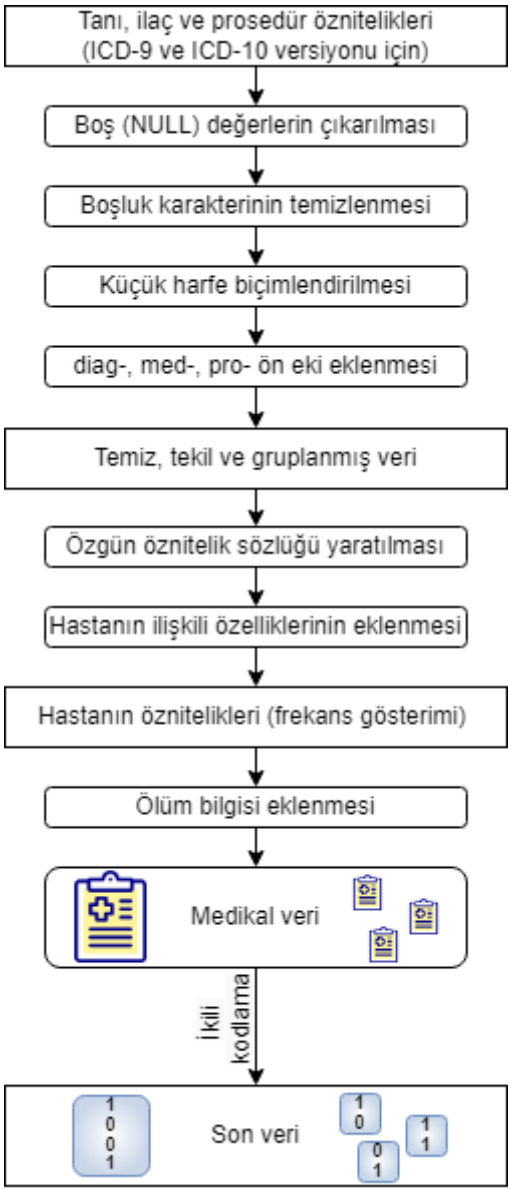
Needell et al. [54] tarafından algoritma tasarımında sıkça karşılaşılan ikili bayrak değeri ile veri sınıflandırma probleminin faydaları incelenmiş, düşük hesaplama ve kaynak maliyetleri olan bir yapı önerilmiş ve sayısal deneylerle sonuçlar gösterilip, temel bir örnek ile teorik analizler sağlanmıştır. İkili bayrak değeri kullanarak tasarlanan yöntemin ilk aşaması, bilinen sınıf üyeliğine sahip veriler üzerinde eğitim gerçekleştirdiği ve ikinci aşaması, önceden bilinmeyen sınıf üyeliğine sahip yeni veri noktalarını sınıflandırmak için kullanıldığı belirtilmiştir. Kullanılan DVM yaklaşımında algoritmanın hiperdüzlemler ile eğitim verileri arasındaki ilişkiyi çıkardığı ve sınıflandırdığı açıklanmıştır. Sonuç olarak, ikili bayrak değeri ile yüksek düzeyde nicelleştirilmiş veri temsillerinden sınıflandırma gibi öğrenme çıkarımları gerçekleştirecek bir yapı sunarak daha iyi performans alındığı gösterilmiştir.

Fitkov-Norris et al. [55] tarafından seyircilerin tekrar izleme eğilimini tahmin etmek amacıyla, kategorik girdi kodlama ve ölçekleme yaklaşımlarının sinir ağı duyarlılığı ve genel sınıflandırma performansı üzerindeki etkisini araştırılmıştır. Sinir ağı sınıflandırıcısının

performansını iyileştirebilmek için; öznitelik seçimi, yeniden örnekleme ve sürekli özniteliklerin ayrıklaştırılması adımlarının olduğu bir dizi veri ön işleme aşamasının yapılması gerektiği açıklanmıştır. Yalnızca kodlama türünün ve gizli katman sayısının sinir ağının geneli ve duyarlılık sınıflandırma doğruluğu üzerinde bir etkisi olduğu gösterilmiştir. Kodlamanın minimum duyarlılık ve genel sınıflandırma doğruluğu üzerindeki etkisi küçük olsa bile, termometre kodlamasının, bu ölçütlerde tutarlı olarak, diğer kodlamalardan önemli ölçüde daha iyi performans gösterdiği vurgulanmıştır. Sonuç olarak, girdilerin kodlanmasının, sınıflandırma doğruluğu üzerinde önemli bir etkiye sahip olduğu ve sıralı veya termometre kodlama yaklaşımlarının kullanılmasının, sinir ağı sınıflandırıcısının performansını önemli ölçüde artırdığı ortaya konulmuştur.

Bu nedenle, model eğitimine harcanan hesaplama maliyetini ve zamanı azaltmak için aşağıdaki öznitelik çıkarma yöntemi uygulanmıştır: o hasta için ilgili öznitelik değeri 0'dan büyükse değeri 1 olarak güncellenmiştir. Son olarak, eğer hasta ölmüşse, bağımlı değişken 1, aksi durumda 0 olarak atanmıştır.

Bu işlem ile beraber, son veri kümesi yaratılmıştır. Belirtilen işlemler dışında başka bir ön işleme adımı uygulanmamıştır. Öznitelik biçimlendirme, temizlik ve çıkarma yöntemi ile işlem akışı Şekil 2.10. de özetlenmiştir.



Şekil 2.10. Veri kümesi yaratma işlem akışı

Tablo 2.12. de ön işleme adımları sonrası model eğitimlerinde kullanılacak verinin oluşan son sayısal durumu özetlenmiştir.

Tablo 2.12. Eğitim verisi sayısal değerleri

Kanser Türü	**Hasta Sayısı**	**Ölen Hasta Sayısı**	**Öznitelik Sayısı**
Meme	2.037	92	8.303
Akciğer	3.134	335	9.942
Prostat	1.927	121	8.468
Mide	557	63	4.843

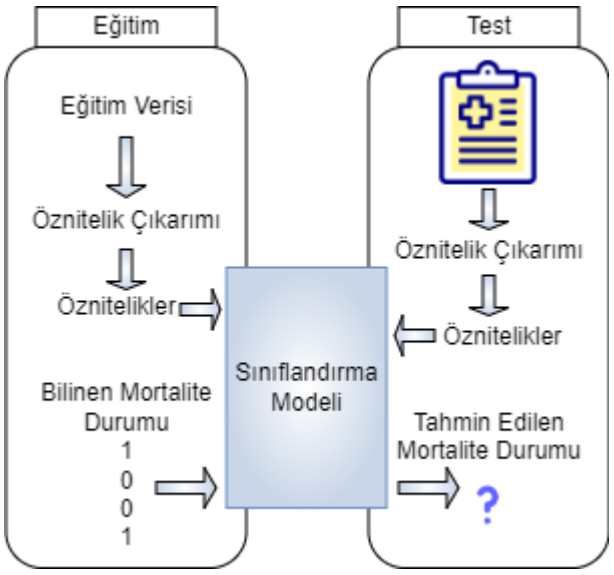
3. YÖNTEMLER

Bu bölümde çalışmada tasarlanan genel yöntem anlatılmış, öznitelik seçim aşamaları ve sebebi açıklanmış ve kullanılan makine öğrenmesi yöntemlerinin neler oldukları, nasıl çalıştıkları ve seçilme sebepleri ile birlikte çalışma içerisindeki kullanılışları sunulmuştur.

3.1. Genel Yapı

Bu çalışmada açık erişim olan MIMIC-IV adlı sağlık veri kümesi kullanılmıştır. Bu veri kümesinde meme, akciğer, prostat ve mide kanseri hastaları için tanı, ilaç ve prosedür öznitelikleri çıkarılmıştır. Bu yapı birden fazla öznitelik alarak seçilen kanser türleri için mortalite oranını tahmin eden bir sınıflandırma modelidir. Bu yapıda, tipik ikili sınıflandırma problemini çözmek için eğitim ve test aşamalarından oluşan parametrik ve parametrik olmayan gözetimli öğrenme modellerleri kullanılmıştır (Şekil 3.1.).

Parametrik modeller, eğitim verilerini kullanarak bir dizi parametreyi tahmin etmek için kullanılan fonksiyon kümelerinden oluşur. Parametrik olmayan modeller; alt kümesi (parametreleri), tahmin sırasında kullanımda olan eğitim kümesi tarafından belirlenen modellerdir. Gözetimli öğrenme, etiketlenmiş eğitim verilerinden ve bir dizi eğitim örneğinden bir işlev çıkarsaması yapar. İkili sınıflandırma, belirli bir durum için sınıf etiketini doğru veya yanlış olarak tahmin eden bir görevdir [50].



Şekil 3.1. Genel yapı

Öncelikle, hastaların tanı, ilaç ve prosedür bilgileri veri kümesinden alınmış ve hastanın mortalite durumu ile etiketlenmiştir. Daha sonra, ön işleme aşamasında boş değerler çıkarılmış, diğer verilerin doğru şekilde gruplanabilmesi için fazla boşluk karakterlerinden temizlenmiş ve hepsi küçük harf olarak güncellenmiştir. Bu öznitelik vektörleri Lojistik Regresyon, Karar Ağacı, Rastgele Orman, Destek Vektör Makinesi ve Çok Katmanlı Algılayıcı sınıflandırıcıları ile eğitilmiştir. Sonuç olarak, eğitilen verilerden bir makine öğrenimi modeli devreye alınmış ve mortalite değerini tahmin etmek için bir test örneği de modellere verilerek sonuçlar elde edilmiştir. Son olarak, sınıflandırıcı değerlendirmesi için çoklu metrikler kullanılarak karşılaştırma sonuçları rapor edilmiştir.

3.1.1. Öznitelik seçim yöntemi

Bu bölümde öznitelik seçimi için incelenen çalışmalar ve yöntemlerinin detayları gösterilmiştir. Ayrıca neden öznitelik seçimi yapıldığı ve bu çalışma için seçilen yöntemin açıklaması da yapılmıştır.

Aşağıdaki çalışmalarda [56]-[58]; farklı Lojistik Regresyon metotlarının öznitelik seçim yöntemi olarak nasıl kullanıldığı açıklanmıştır. Ayrıca sağlık alanında öznitelik seçiminin önemi de bu araştırmalarda vurgulanmış ve olabilecek en az sayıda özniteliğin neden seçilmesi gerektiği de gösterilmiştir.

Khandezamin et al. [56] tarafından meme kanserinin iyi ve kötü huylu örneklerinin teşhisinde kullanılacak sinir ağı modeline Lojistik Regresyon ile öznitelik ağırlıklandırması ve bu ağırlıklar kullanılarak ilgisiz olanların elenmesi yapılmıştır. Lojistik Regresyona dayalı öznitelik seçim yönteminin performansı; bu seçim yönteminin uygulandığı ve uygulanmadığı iki grup küme ile aynı modele verilerek yapılmıştır. Burada sıralama sonrası katsayısına göre her değer tek tek eklenerek AUC değeri hesaplanmış, ekleme işlemi yeterli görülen bir değerde kesilmiş ve yakın AUC değerleri arasından en küçük öznitelik kümesi seçilmiştir. Bu seçim yönteminin raporlanan doğruluk değerlerini %1 lik artış ile ortalama 98.5 e çıkardığı gösterilmiştir. Daha az ilgili özniteliklerin kaldırılması ve her veri kümesi için sınırlı sayıda öznitelik seçilmesinin daha iyi doğruluk sağladığı ve hesaplama maliyetini azalttığı ortaya konulmuştur.

Liu et al. [57] tarafından akciğer kanseri sınıflandırmaları için kullanılabilecek gen biyobelirteç seçimi yapan LogSum+L2 düzenleyicili Lojistik Regresyon yöntemi geliştirilmiştir. Gen biyobelirteçlerinin (özniteliklerin) doğru seçilmesi, teşhisin doğruluğunu önemli ölçüde artırabileceği belirtilmiştir. Biyolojik açıdan, her genin hastalıkla ilgili olmadığı bu sebeple en az sayıda gen seçiminin hedef kanseri güçlü bir şekilde gösterdiği vurgulanmıştır. İlişkisiz genlere sahip verilerin gürültü oluşturduğu ve makine öğrenimi yaklaşımlarında hastalığa neden olan patojenik genlerin bulmasının zorlaştığı belirtilmiştir. Bu yöntem ile %97.86 doğruluk elde edilmiştir.

Huang et al. [58] tarafından prostat, lenf ve akciğer kanserleri için Lojistik Regresyon ile gen (öznitelik) seçimi yapan hibrit L1/2 +2 düzenleme yöntemi geliştirilmiştir. Öznitelik seçiminin, genomik verilerdeki bilgi keşfinde ve kanser sınıflandırmasında önemli bir rol oynadığı vurgulanmıştır. Bununla birlikte hesaplama süresinin öznitelik boyutuyla doğrusal olarak arttığı ve en az sayıda öznitelik seçilmesi gerektiği belirtilmiştir. Bu yöntem ile %96.55 doğruluk elde edilmiştir.

Yukarıda belirtilen yöntemlerin ışığında, bu çalışmadaki temel öznitelik seçim algoritması Khandezamin et al. [56] tarafından yapılan çalışmadan esinlenilmiş ancak farklı bir şekilde yorumlanmıştır. Liu et al. [57], Huang et al. [58] ve önceki çalışmalar bölümünde bahsedilen Hammoud et al. [29] tarafından yapılmış çalışmalarla; öznitelik seçiminin medikal alandaki önemi, sınıflandırma performansı için az sayıda seçilmesi gerektiği ve buradaki probleme uygunluğu görülmüş ve bu çalışmada ilke olarak benimsenmiştir.

Her bir kanser türü verileri için Lojistik Regresyon yöntemi çalıştırılmış ve özniteliklere katsayı ağırlık ataması yaptırılarak mortalite sonucu ile uygunlukları ortaya çıkarılmıştır. Sınıflandırma sonucu, 0 ile 1 arasındaki bir olasılık değeri, ile aşağıdaki formülde yer alan özniteliklerin ağırlıklı katsayılarının toplam değeri ile belirlenmektedir. Burada katsayı (β), bağımlı değişkenin olasılığının logaritmasındaki değişimi ilişkinin yönü ve büyüklüğü ile göstermektedir [50].

$$P(y^{(i)} = 1) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_p x_p^{(i)}))} \quad (1)$$

Katsayı hesaplaması yapılırken lojistik fonksiyon tarafından ağırlıklı toplam bir olasılığa dönüştürülür. ln() işlevindeki terim "odds" olarak adlandırılmaktadır ve buna log oranları denilmektedir. Odds; bir olayın olasılığının, olmaması olasılığına bölünmesidir.

Yorumlanabilmesi için aşağıdaki formül yalnızca doğrusal terim olacak şekilde formatlanır [52].

$$ln\left(\frac{P(y=1)}{1-P(y=1)}\right) = log\left(\frac{P(y=1)}{P(y=0)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \quad (2)$$

Bu formül, lojistik regresyon modelinin log oranları için doğrusal bir model olduğunu göstermektedir. X özniteliklerinden biri 1 birim değiştirildiğinde tahminin nasıl değiştiğinin anlaşılabilmesi için her tarafa exp() fonksiyonu uygulanmaktadır. Olasılıkların doğal log dönüşümü "logit dönüşümü" olarak da adlandırılmaktadır. Bir öznitelik 1 birim arttığında sonucu nasıl etkilediği iki tahminin arasındaki fark yerine, oranlarına bakılarak karşılaştırılmaktadır [52].

$$\frac{P(y=1)}{1-P(y=1)} = odds = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p) \quad (3)$$

Sonunda, bir özniteliğin ağırlığının exp() değeri kadar basit bir sonuç kalmaktadır. Bir öznitelikteki 1 birimlik değişim, olasılık oranını exp(β_j) faktörü kadar değiştirmekte yani log olasılık oranına karşılık gelen ağırlığın değeri kadar artırmaktadır. Log olasılığını en üst düzeye çıkarabilmek için, regresyon katsayıları sınıflandırma tamamlanana kadar yinelemeli olarak yeniden ağırlıklandırılmaktadır [52].

$$\frac{odds_{x_j+1}}{odds_{x_j}} = \exp\big(\beta_j(x_j+1) - \beta_j x_j\big) = \exp(\beta_j) \quad (4)$$

Lojistik regresyonda, pozitif değerlerin üsü 1'den büyük bir katsayı üretirken sıfır değerine sahip bir katsayı, 1'e eşit bir exp(β) üretmekte, bu da bağımsız değişkenin bağımlı değişkenin olma olasılığını etkilemediğini göstermektedir [56]. Her bir özniteliğe atanmış katsayı ile mortalite arasındaki ilişkinin yönü ve ilgililiği pozitif ve negatif olarak görülmüştür. Ana çalışmada olduğu gibi ilgisiz öznitelik ağırlıkları 0 civarında kalması sebebi ile, bunların modele en son girebilmesi için bir sıralama yapılmıştır. Ancak sıralama sonrası, negatif katsayılı olan öznitelikler sıranın sonunda olacağı için ters ilişkideki verilerden faydalanamayacağı görülmüştür.

Pozitif ilişkili öznitelikler sonucu iyileştirici yönde etkilerken negatif olarak ilişkili olanlar ise sonucu azalan bir şekilde etkilemektedir, ancak yine de ilişkilidir [50]. Ana çalışmanın aksine her yöndeki ilişkilerden yararlanılmak istenmiştir. Böylece, ters ilişkili özniteliklerin de modele erken girebilmesi için özniteliklerin ağırlıkları mutlak değerlerine göre sıralanıp öncelikli listesi oluşturulduktan sonra modele asıl değerleri üzerinden gönderilmiştir.

3.1.2. Kullanılan makine öğrenmesi yöntemleri

Bu bölümde, kullanılan makine öğrenmesi yöntemlerinin neler oldukları, nasıl kullanıldıkları, olumlu ve olumsuz yönleri ve çalışmada nasıl kullanıldığı açıklanmıştır. Bu çalışmada aşağıdaki sınıflandırıcılar kullanılmıştır: Lojistik Regresyon, Karar Ağacı, Rastgele Orman, Destek Vektör Makinesi ve Çok Katmanlı Algılayıcı.

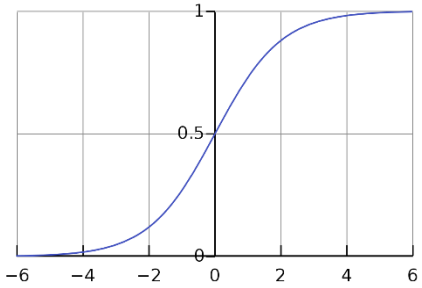
3.1.2.1. Lojistik Regresyon

Lojistik regresyon (Logistic Regression), lojistik eğri üzerindeki verileri eşleştirerek bir sonucu belirleyen bir veya daha fazla bağımsız değişken içeren veri kümesini analiz etmek için kullanılan denetimli bir sınıflandırma algoritmasıdır. Çıkan sonuç ikili değer ile gösterilir. Lojistik regresyonun (LR) amacı, bir model oluşturmak için en az değişkeni kullanarak bir veya daha fazla bağımsız değişken arasındaki ilişkiyi tanımlamaktır. Olasılık için lojistik regresyon modeli kullanılabileceği gibi, sınıflandırma için de kullanılabilir [51].

Sigmoid fonksiyonu, lojistik regresyon yönteminin temelidir; pozitif sonsuz ile negatif sonsuz arasında girdi alabilir, ancak çıktılar her zaman sıfır ile bir arasındadır. Fonksiyondaki x değişkeni modelde kullanılan tüm bağımsız değişkenleri temsil ederken f(x) bu değişkenlerin varlığının olasılığını gösterir [50].

$$f(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

Lojistik regresyon, ilgili özniteliklerin varlığının olasılığını tahmin etmek için formül katsayılarını ve bununla beraber ilgililik seviyelerini üretir [50].



Şekil 3.2. Lojistik regresyon

Grafikten görülebileceği üzere bu fonksiyonun olabilecek minimum çıktı değeri 0 ve maksimum çıktı değeri ise 1'dir. x değeri 0 iken sigmoid fonksiyonunun çıktısı 0,5'tir, dolayısıyla sınıflandırma problemleri için, çıktı 0,5'ten büyükse sonuç 1, küçükse 0 olarak değerlendirilir (Şekil 3.2.). Sonuç olarak, bir olayın gerçekleşme olasılığını hesaplamak için kullanılır [51].

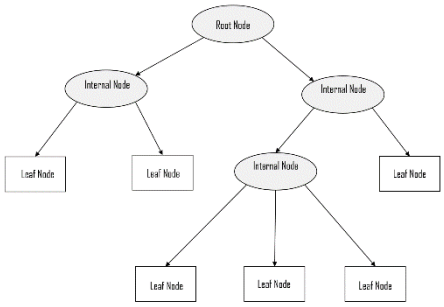
Bu çalışmada Lojistik Regresyonun seçilmesinin sebebi, değişkenler arasındaki ilişkiyi tanımlaması ve özniteliğin varlığının olasılığını tahmin edebilmesi sebebiyle kullanılmıştır.

3.1.2.2. Karar Ağacı

Karar Ağaçları (Decision Tree), verilerin belirli bir parametreye göre sürekli olarak bölündüğü, parametrik olmayan bir Denetimli Makine Öğrenimi türüdür. Veri özniteliklerinden çıkarılan basit karar kurallarını öğrenerek bir hedef değişkenin değerini tahmin eden bir model oluşturulur. Karar ağacının (KA) ana yöntemi ID3 olmakla birlikte, C4.5 ve CART teknikleri de vardır. Bu teknik, lineer sınırlı problem türleri ile daha iyi çalışır; yüksek doğruluk ve zamanın önemli olduğu durumlarda uygulanabilir. Ayrıca C4.5 Karar Ağacı algoritması yüksek hassasiyet sağlama yeteneğine sahipken, CART algoritması hızlı karar gerektiren durumlarda kullanılmaktadır [52].

Kararları sağlamak için bir grafiksel bir ağaç kullanılır. Bir akış şemasına benzer şekilde çizilen ağaca sınıflandırma ağacı denir. Karar ağaçlarının temel amacı, böl-yönet

yöntemini kullanarak veri kümesini belirli kurallar uygulayarak küçük alt gruplara bölmektir. Ağacı oluşturmak için eğitim verilerinin hangi sırayla kullanılması gerektiğini belirlemek için entropi ölçümü kullanılır. Entropinin ölçüsü ne kadar fazlaysa, o kadar belirsizdir. Bu nedenle karar ağacının kökünde Entropi ölçüsü en düşük olan alanlar kullanılır (Şekil 3.3.). Bu algoritmanın her düğümü, sınıf üzerinde bir olasılık dağılımı ile işaretlenir. Sınıflandırma işlemi veya kuralı, kökten yaprak düğümüne giden yol ile temsil edilir. Karar düğümleri, veri kümesinde karar vermek, sınıflandırmak veya tahmin yapmak için kullanılır. Ağacın dallanmamış son düğümü yaprak düğümüdür. Bir karara varmak için ağacın kökünden yaprak düğümlerine kadar belli bir yol izlenir [51].



Şekil 3.3. Karar ağacı

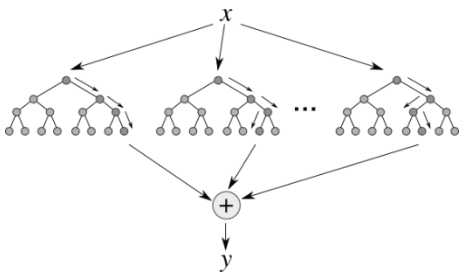
Karar ağaçları, düşük maliyeti, anlaşılması, yorumlanması, veri tabanları ile entegrasyonu ve iyi güvenilirliği nedeniyle en yaygın kullanılan sınıflandırma tekniklerinden biridir. Sınıflandırma doğruluğu diğer öğrenme yöntemlerine göre oldukça etkilidir. Verilerde gürültü oldukça, her yaprağın saf olmasını sağlanmaya çalışıldığı için büyük bir karar ağacı ortaya çıkar ve verinin ezberlenmesi durumuna neden olur [52].

Bu çalışmada Karar Ağacının seçilmesinin sebebi, düşük maliyetli oluşu ve sınıflandırma doğruluğu sebebiyle kullanılmıştır.

3.1.2.3. Rastgele Orman

Rastgele Orman (Random Forest), regresyon için de kullanılabilen, birbiriyle ilişkisiz birkaç karar ağacından oluşan denetimli bir sınıflandırma yöntemidir. Tüm karar ağaçları, eğitim sürecinde rastgele olarak büyür. Bir sınıflandırma işlemi için, buradaki her ağaç bir

sınıfa karar verir ve yüksek oyu alan sınıf seçilir. Rastgele Orman (RO) ayrıca özniteliklerin önem derecesini de gösterir. Alt veri kümeleri, orijinal veri kümesinden çıkartılarak oluşturulur (Şekil 3.4.). Rastgele Orman algoritması, her düğümde rastgele seçilen değişkenler arasından en iyisini kullanarak dallandırma yapar ve bu işlem n ağaç oluşturmak için n kez tekrar edilerek bir orman oluşturulur [53].



Şekil 3.4. Rastgele orman

Bölme kriteri olarak kullanılan Gini indeksindeki bir azalma ile saflık artar ve bu değer sıfır olduğunda maksimum saflığa ulaşılır. Belirli bir düğüm Gini indeksi arttıkça sınıf heterojenliği artarken, Gini indeksi azaldıkça sınıf homojenliği artmaktadır. Bir alt düğümün Gini indeksi, bir üst düğümün Gini indeksinden küçük olduğunda dal tamamlanır. Gini indeksi sıfır olunca yani her yaprak düğümünde bir sınıf kaldığında ağaç tamamlanır [52].

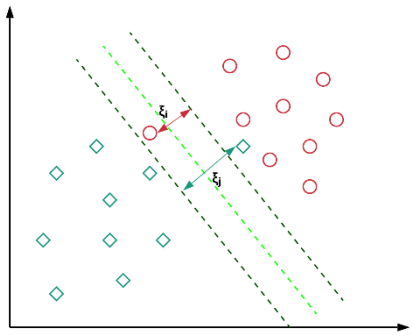
$$GINI(t) = \sum_j [p(j/t)]^2 \tag{6}$$

Sınıflandırıcı, düşük hesaplama karmaşıklığına sahip olduğu için eğitim aşamasında çok hızlıdır. Bu avantaj, tek bir karar ağacının kısa eğitim süresinden ve orman için eğitim süresinin ağaç sayısı ile doğrusal olarak artmasından kaynaklanmaktadır. Bir test örneğinin değerlendirilmesi her ağaçta ayrı ayrı gerçekleşir, bu nedenle çok hızlı değerlendirir. Aynı zamanda büyük ölçekli veriler için çok verimli ve doğruluğu yüksektir. Karar ağaçlarının en büyük sorunlarından biri olan verinin ezberlenmesi durumu burada farklı veri kümeleri üzerinde eğitilerek çözülmüştür [53].

Bu çalışmada Rastgele Ormanın seçilmesinin sebebi, eğitim aşamasının hızlı oluşu ve veriyi ezberlemeyen yapısı sebebiyle kullanılmıştır.

3.1.2.4. Destek Vektör Makinesi

Destek Vektör Makineleri (Support Vector Machine), sınıflandırma ve regresyon analizi için kullanılan, veri analizi ilişkili öğrenme algoritmalarına sahip denetimli bir sınıflandırma yöntemidir. Destek Vektör Makinesi (DVM), karar sınırlarını yine karar seviyeleri ile tanımlayarak derecelendirir. Karar seviyesi, farklı sınıf üyelerine sahip bir grup veriyi ayırır. Algoritma, veri kümesini sınıflara ayıran bir hiper düzlem veya bunlardan oluşan bir küme oluşturur. İki boyutta iki sınıfı ayıran bir hiper düzlem çizgi olarak oluşur. Karar sınırına en yakın nokta arasındaki mesafeye marj denir. Karar sınırına en yakın olan her bir sınıftaki veri noktalarına destek vektörleri denir (Şekil 3.5.). Destek Vektör Makinesi için en iyi hiper düzlem, destek vektöründeki veri noktaları için maksimum marja sahip olan olarak tanımlanır [50].



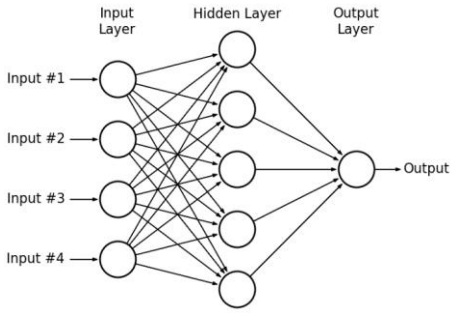
Şekil 3.5. Destek vektör makinesi

Destek Vektör Makineleri, karar işlevi için farklı çekirdek metodlarını destekler. Çekirdek, girdi verilerini, sorunun verimli çözüldüğü daha yüksek boyutlu bir alana dönüştürür. Doğrusal veya değişken dereceli polinomlar gibi çekirdek metodlarıyla, farklı karar sınırları bulunabilir. Destek Vektör Makinesi, çok güçlü ve çok yönlü bir makine öğrenimi modelidir. Makul boyuttaki veri kümelerinde önemli ölçüde daha iyi sınıflandırma performansı verir [52].

Bu çalışmada Destek Vektör Makinesinin seçilmesinin sebebi, farklı çekirdek metodları ile esnek oluşu ve sağlık verileri gibi fazla özniteliğin olduğu durumlarda efektif olması sebebiyle kullanılmıştır. 2 nci dereceden polinom çekirdeği ise verilerin doğrusal olarak ayrılamadığı görüldüğü için seçilmiştir.

3.1.2.5. Çok Katmanlı Algılayıcı

Çok Katmanlı Algılayıcı (Multi Layer Perceptron), girdi verilerini bir dizi uygun çıktıya eşleyen ileri beslemeli bir yapay sinir ağı modelidir. Basit bir ifadeyle, insan beynindeki bir sinir hücresinin (nöron) öğrenme sürecini matematiksel olarak taklit eder. Bir Çok Katmanlı Algılayıcı (ÇKA), yönlendirilmiş bir grafikteki birden çok düğüm katmanından oluşur ve her katman bir sonrakine tam bağlıdır. Çok Katmanlı Algılayıcı mimarisi, giriş katmanı ve çıkış katmanına ek olarak en az bir gizli (orta) katman içerir (Şekil 3.6.). Giriş katmanında probleme konu olan veri kabul edilir. Gelen değerler çeşitli ağırlıklarla çarpılarak gizli katmana iletilir. Gizli katmanların ve gizli katmanlardaki nöronların sayısının arttırılması, genellikle Çok Katmanlı Algılayıcının başarısı ile birlikte hesaplama süresini de arttırır. Çıktı katmanı, ara katmanda hesaplanan değerlerin ağırlıklı varyasyonlarını işleyerek son çıktıyı üretir [50].



Şekil 3.6. Çok katmanlı algılayıcı

Giriş düğümleri dışında, her düğüm, doğrusal olmayan bir aktivasyon fonksiyonuna sahip bir nörondur. Aktivasyon fonksiyonları, nöronlara girdi olarak verilen değerlerin ağırlıkları toplanarak nöronun çıkış değerini üretmek için kullanılır. Biyolojik nöronlarda bu görev için basit bir eşik kontrolü yapıldığı kabul edilirken, Çok Katmanlı Algılayıcı modellerinde aktivasyon fonksiyonu olarak birçok doğrusal olmayan fonksiyon seçilebilir. Bunlar genel olarak: ReLU, tanh ve Sigmoid fonksiyonlarıdır. Rectified Linear Unit (ReLU), yalnızca pozitif öğeleri koruyup negatif öğeleri atarak doğrusal olmayan dönüşüm sağlar ve iyi performanslıdır. Çok Katmanlı Algılayıcı, ağı eğitmek için geri besleme (hata geri bildirimi) adı verilen denetimli bir öğrenme tekniği kullanır. Burada, kontrollü bir eğitim aşamasından sonra ağın sınıflandırabilmesi için bağlantıların ağırlıkları değiştirilir. Çok

Katmanlı Algılayıcı, eğitimden sonra hızlı tahminler verebilecek şekilde karmaşık ve doğrusal olmayan problemlere uygulanabilir. Ağ derinliği arttıkça hesaplamalar zorlaşır ve zaman alıcı hale gelir. Ayrıca model sonucu eğitimin kalitesine de büyük ölçüde bağlıdır. Aynı doğruluk oranı hem daha küçük hem de büyük girdi verileriyle elde edilebilir [53].

Bu çalışmada Çok Katmanlı Algılayıcının seçilmesinin sebebi, büyük veriler ile iyi çalışabilen esnek bir yapıda oluşu ve hızlı tahmin verebilmesi sebebiyle kullanılmıştır. Katmanların yapısı ve maksimum iterasyon sayısı [50] de belirtilen "stretch pants" yöntemi ile seçilmiştir. Bu; gereğinden fazla nöronlu katmanla başlayarak her katmanda azaltılması, maliyetli olmaması için az iterasyonlu olması ve verinin ezberlenmemesi için erken durdurma yapılması olarak tanımlanmıştır. ReLU aktivasyon fonksiyonu, pozitif değerleri (bu durumda ölenleri) ön plana çıkardığı için seçilmiştir. Adam optimizasyon yöntemi ise büyük verilerde daha iyi çalıştığı için seçilmiştir [50].

4. SONUÇLAR

Bu bölümde kullanılan değerlendirme ölçütleri ile deneysel sonuçlar paylaşılmıştır. Ayrıca özniteliklerin dağılım oranları da bu bölümde verilmiştir. Deneysel sonuçlar, her bir kanser türüne uygulanmış yöntemlerin detaylı sonuçlar ve genel sonuçlar olmak üzere iki bölümde incelenmiştir. Genel sonuçlar ise F1 Makro Ortalama ve AUC-ROC karşılaştırması olarak ikiye ayrılmıştır.

4.1. Değerlendirme Ölçütleri

Tüm kanser türleri için, belirtilen makine öğrenme algoritmaları kullanılmıştır. Test kümesinin boyutu ilgili kanser kohortundaki tüm verinin %30'u, eğitim kümesinin boyutu ise ilgili kanser kohortundaki tüm verinin %70'i olarak seçilmiştir. Oluşturulan veri kümesi, sonuçların tekrar üretilebilmesi için sabit bir rastgele durum değeri ile modele verilerek deneysel sonuçlar oluşturulmuştur.

Performans karşılaştırmada temel olarak F1 skoru özelindeki Makro Ort. değeri esas alınmıştır. Makro Ort. değerinin ön plana alınmasının sebebi, veri kümesindeki sınıfların dengesiz dağılmış olmasındandır. Ölen sınıfı için alınan başarı, örnek sayısı yetersizliğinden dolayı, yaşayan sınıfından daha az olduğu görülmüştür. Ağırlıklı Ort. ve Doğruluk değerleri bu dengesizlikten kaynaklı sonuçları yukarı çektiği ve ölen sınıfı için gerçeği yansıtmadığından değerlendirme kriteri olarak kullanılmamış ancak genel durumu yansıtması amaçlı paylaşılmıştır. AUC-ROC değeri, her sınıfın F1 skoru ve Makro Ort. özelinde Kesinlik ve Duyarlılık değerleri de ayrıca paylaşılmıştır. Bu değerlerin detaylı açıklamaları ve formülasyonları aşağıda belirtilmiştir.

Performans ölçümü sırasında odaklanılan ölen sınıfı Pozitif (P), yaşayan sınıfı ise Negatif (N) olarak adlandırıldı. Verilen bir sınıflandırıcı ve bir örnek için dört olası sonuç (Şekil 4.1.) vardır:

- Doğru Pozitif (TP), modelin pozitif vakaları doğru bir şekilde pozitif olarak tahmin etmesi olarak tanımlanır,
- Yanlış Pozitif (FP), modelin negatif vakaları hatalı bir şekilde pozitif olarak tahmin etmesi olarak tanımlanır,

- Doğru Negatif (TN), modelin negatif vakaları doğru bir şekilde negatif olarak tahmin etmesi olarak tanımlanır,
- Yanlış Negatif (FN), modelin pozitif vakaları hatalı bir şekilde negatif olarak tahmin etmesi olarak tanımlanır [50].

	Negatif Tahmin (Yaşayan)	Pozitif Tahmin (Ölen)	Toplam
Negatif Gerçek (Yaşayan)	Doğru Negatif (TN)	Yanlış Pozitif (FP)	Negatif Örnek (N)
Pozitif Gerçek (Ölen)	Yanlış Negatif (FN)	Doğru Pozitif (TP)	Pozitif Örnek (P)

Şekil 4.1. Karışıklık matrisi

Kesinlik, tanımlanan sınıflardan kaçının pozitif olduğunu gösteren ölçüdür. Gerçek pozitif etikete sahip tüm örnekler üzerinde doğru tahmin edilen pozitif sınıf örneklerinin oranını ölçer. Bir modelin pozitif örnekleri tahmin etmede ne kadar tutarlı olduğunu gösterir [50].

$$Kesinlik\ (Precision) = \frac{TP}{TP + FP} \tag{7}$$

Duyarlılık, pozitif sınıflardan kaçının tanımlandığını gösteren ölçüdür. Pozitif olarak tahmin edilen tüm örnekler üzerinde doğru tahmin edilen pozitif sınıf örneklerinin oranını ölçer. Bir modelin tüm veri kümesinden pozitif örnekleri bulmada ne kadar başarılı olduğunu gösterir [50].

$$Duyarlılık\ (Sensitivity) = \frac{TP}{TP + FN} \tag{8}$$

Doğruluk, doğru tahmin edilen örneklerin tüm örneklere oranını gösteren ölçüdür. Doğru sınıflandırılmış örneklerin toplamının toplam örnek sayısına bölünmesiyle hesaplanır. Doğru sınıflandırılmış örneklerin oranını gösterir [50].

$$Doğruluk\ (Accuracy) = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

Receiver Operating Characteristics (ROC) eğrisi, dengesiz veri kümelerindeki performanslarına göre sınıflandırıcıları görselleştirmek ve seçmek için kullanılabilen bir araçtır. ROC eğrisi, genel hata oranlarına bağlı olarak her model için 1-kesinliğe karşı duyarlılığın grafiksel temsilidir. ROC eğrisi altındaki alan (AUC) ise bunun skaler değeridir. Bir sınıflandırıcının AUC değeri, onun rastgele seçilen pozitif bir örneği, rastgele seçilen negatif bir örnekten daha yüksek sıraya koyma olasılığıdır. AUC, sınıf dağılımlarından ve yanlış sınıflandırma hatalarından etkilenmez. Değeri, tek bir eğitim ve test çifti kullanılarak hesaplandığı için kullanılan test testine bağımlıdır. Daha yüksek AUC değerine sahip sınıflandırıcı genel olarak daha iyi performans gösterir. 1.0 değeri kusursuz bir sınıflandırıcıyı belirtirken, 0,5 değeri sınıflandırıcının rastgele çalıştığını gösterir [50].

$$AUC = \int_0^1 TPR\left(FPR^{-1}(x)\right)dx = \int_{-\infty}^{\infty} TPR(t) * FPR'(t)dt \tag{10}$$

F1 skoru, dengesiz sınıf dağılımına sahip veri kümeleri için ortalama performans ölçüsü veren bir fonksiyondur. F1 skoru, kesinlik ve duyarlılığın harmonik (dengeli) ortalamasıdır. Hem yanlış pozitifleri hem de yanlış negatifleri hesaba kattığı için dengesiz sınıf dağılımında doğruluktan daha faydalıdır. F1 skorunun değeri 0 ile 1 arasındadır. 1 modelin en iyi çıktısını, 0 ise modelin en kötü çıktısını temsil eder [50].

$$F_1 = \frac{TP}{TP + \left(\frac{FN + FP}{2}\right)} \tag{11}$$

Makro ortalama, bir grup değerin aritmetik ortalamasını gösteren ölçüdür. Makro ortalamalı F1 skoru, sınıf başına tüm F1 skorlarının ağırlıksız ortalaması kullanılarak hesaplanır. Destek değerlerinden bağımsız olarak tüm sınıflar eşit olarak ele alınır [50].

Bu bölümde belirtilen yöntemler üç şekilde incelenmiştir. Öncelikle tüm öznitelikler kullanılarak TEMEL sonuçlar toplanmıştır. Genel olarak TEMEL sonuç, karşılaştırma amaçlı kullanılacak ana sonuçlara karşılık gelir. Tüm metodlarda, tüm öznitelikler kullanılarak bu sonuçlar elde edilmiş ve raporlanmıştır. Böylece, modellerin temel performansları ortaya konulmuştur. TEMEL sonucunu geçen birden fazla öznitelik grubu olursa en az sayıdaki grup ele alınmıştır. Eğer TEMEL sonucu geçilememişse öncelikli

olarak en büyük F1 skoru, sonra AUC-ROC değeri en son da öznitelik sayısı önceliğindeki sıralamaya göre seçimler yapılmıştır.

İkinci olarak, lojistik regresyon ile ilgililiklerine göre sıralanmış en iyi 100 öznitelik seçilmiştir. Daha sonra sadece bu öznitelikler kullanılarak aynı işlem prosedürleri uygulanmış ve İLK 100 sonucu not edilmiştir. Bu işlem, seçilen yöntemlerin yalnızca minimum öznitelik kümesiyle ne kadar iyi performans gösterdiğini göstermek için yapılmıştır.

Son olarak aynı işlem prosedürleri, ilgililiklerine göre sıralı ilk 100 öznitelikten başlayarak o kanser türü için olası tüm öznitelik sayısına kadar her seferinde bir sonraki 100 grup daha eklenerek tekrarlanmıştır. Tüm sonuçlar toplandıktan sonra, bulunan bu değerler TEMEL sonucunu aşarsa veya mümkün olduğunca yakınsa aday olarak işaretlenmiştir. Adaylar arasında en az sayıdaki öznitelik de, EN İYİ X sonucu oluşturmuştur. Burada X değeri, o kanser türü için o modelle belirtilen en iyi sonucu oluşturan özniteliklerin sayısını gösterir.

4.2. Deneysel sonuçlar

Bu bölümde her bir kanser türü üzerinde uygulanmış farklı makine öğrenmesi modellerinin deneysel sonuçları paylaşılmıştır. İlgili tablolarda 100'erli gruplar halinde verilen sonuçlar görsel olarak renklendirilmiştir. Burada TEMEL ve karşılaştırma için kullanılan sonuç mor renk ile ifade edilmiş ve ilk satırda verilmiştir. İlgili öznitelik grubunun F1 Makro skoru TEMEL değerine yakınsa ya da geçiyorsa ilgili satır sarı ile işaretlenmiştir. Bu değerlerden en iyisi ve eşitlik durumunda en küçük öznitelik grubuna sahip olan satır yeşil ile gösterilmiştir.

4.2.1. Meme Kanseri

Meme kanseri kohort verilerinin test kümesi üzerinde çalıştırılmış 5 modelin; sınıflandırma raporu, karışıklık matrisleri ve 100'erli gruplarla verilmiş öznitelik küme sonuçları bu bölümde gösterilmiş ve değerlendirilmiştir.

4.2.1.1. Lojistik Regresyon sonuçları

Meme kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Lojistik Regresyon modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

612 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.2. de gösterilmiştir. Burada 586 yaşayan hasta için F1 skoru 0.98 ve 26 ölen hasta için F1 skoru 0.43 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.70 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %70 Duyarlılığında %71 Kesinlik elde edilmiştir.

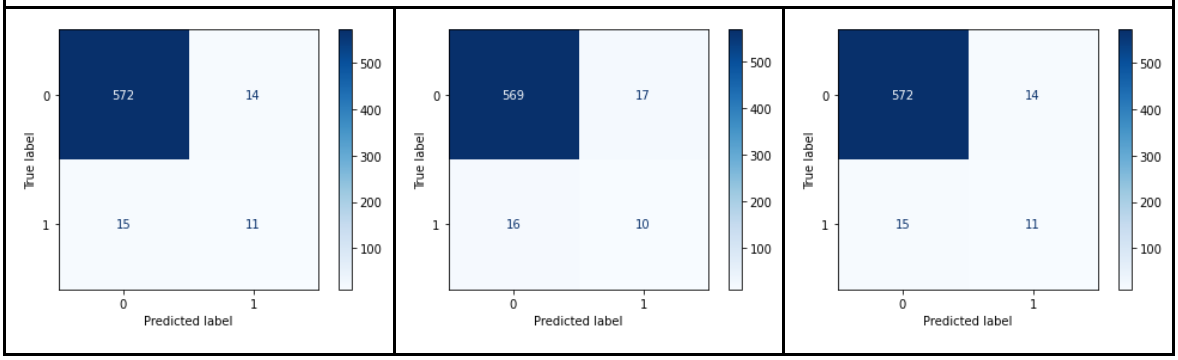
TEMEL öznitelikleri ile 586 yaşayan hastanın 572'si doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 11'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 14'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 15'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 586 yaşayan hastanın 569'u doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 10'u doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 17'si hatalı bir şekilde ölen olarak, ölen hastaların ise 16'sı hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna çok yaklaşılmış ancak geçilememiştir.

EN İYİ öznitelikleri (1500) ile 586 yaşayan hastanın 572'si doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 11'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 14'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 15'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada meme kanseri için belirlenmiş toplam özniteliklerin %18'i kullanılarak TEMEL sonucunun aynısını elde edebilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.2. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.97	0.98	0.98
1	0.44	0.42	0.43
accuracy			0.95
macro avg	0.71	0.70	0.70
weighted avg	0.95	0.95	0.95



Şekil 4.2. Meme kanseri Lojistik Regresyon için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.38 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.67 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %68 Duyarlılığında %67 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (1500) sınıflandırmaların F1 skoru yaşayan için 0.98, ölen için 0.43 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.70 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %70 Duyarlılığında %71 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.759, İLK 100 için: 0.855 ve EN İYİ için: 0.759 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.1. üzerinde paylaşılmıştır.

Tablo 4.1. Meme kanseri Lojistik Regresyon için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.98**	**0.43**	**0.95**	**0.70**	**0.95**	**0.759**	**0.71**	**0.70**
100	0.97	0.38	0.95	0.67	0.95	0.855	0.67	0.68
200	0.97	0.36	0.94	0.67	0.94	0.855	0.66	0.68
300	0.97	0.37	0.94	0.67	0.94	0.795	0.65	0.69
400	0.97	0.38	0.95	0.67	0.95	0.759	0.67	0.68
500	0.97	0.38	0.95	0.67	0.95	0.749	0.69	0.66
600	0.97	0.37	0.95	0.67	0.95	0.731	0.68	0.66
700	0.98	0.41	0.95	0.69	0.95	0.752	0.70	0.68
800	0.98	0.38	0.95	0.68	0.95	0.736	0.70	0.66
900	0.98	0.38	0.95	0.68	0.95	0.761	0.70	0.66
1000	0.98	0.38	0.95	0.68	0.95	0.758	0.70	0.66
1100	0.98	0.38	0.95	0.68	0.95	0.759	0.70	0.66
1200	0.97	0.37	0.95	0.67	0.95	0.762	0.68	0.66
1300	0.98	0.41	0.95	0.69	0.95	0.763	0.70	0.68
1400	0.98	0.41	0.95	0.69	0.95	0.765	0.70	0.68
1500	0.98	0.43	0.95	0.70	0.95	0.759	0.71	0.70

EN İYİ öznitelik grubunun (1500) dağılımı şu şekilde olduğu bulunmuştur; tanı: 952, ilaç: 351, prosedür: 197. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.1.2. Karar Ağacı sonuçları

Meme kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Karar Ağacı modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 2000 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

612 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.3. de gösterilmiştir. Burada 586 yaşayan hasta için F1 skoru 0.98 ve 26 ölen hasta için F1 skoru

0.54 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.76 ve Doğruluk sonucu 0.96 olarak bulunmuştur. Makro Ort. özelinde %78 Duyarlılığında %74 Kesinlik elde edilmiştir.

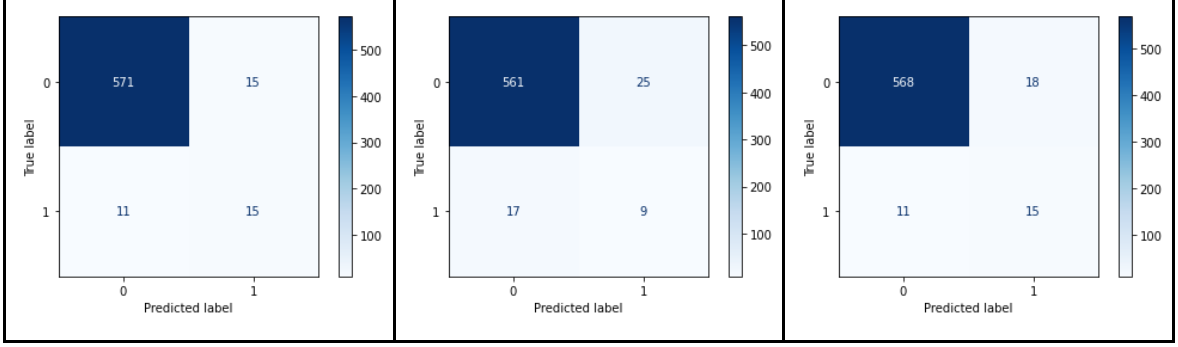
TEMEL öznitelikleri ile 586 yaşayan hastanın 571'si doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 15'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 15'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 11'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 586 yaşayan hastanın 561'i doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 9'u doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 25'i hatalı bir şekilde ölen olarak, ölen hastaların ise 17'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna yaklaşılamamıştır.

EN İYİ öznitelikleri (1900) ile 586 yaşayan hastanın 568'i doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 15'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 18'i hatalı bir şekilde ölen olarak, ölen hastaların ise 11'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada meme kanseri için belirlenmiş toplam özniteliklerin %23'ü kullanılarak TEMEL sonucuna çok yaklaşılmış ancak geçilememiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.3. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.98	0.97	0.98
1	0.50	0.58	0.54
accuracy			0.96
macro avg	0.74	0.78	0.76
weighted avg	0.96	0.96	0.96



Şekil 4.3. Meme kanseri Karar Ağacı için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.96, ölen için 0.30 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.63 ve Doğruluk sonucu 0.93 olarak bulunmuştur. Makro Ort. özelinde %65 Duyarlılığında %62 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (1900) sınıflandırmaların F1 skoru yaşayan için 0.98, ölen için 0.51 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.74 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %77 Duyarlılığında %72 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.775, İLK 100 için: 0.650 ve EN İYİ için: 0.773 AUC-ROC sonuçları ile gösterilmiştir. 100 den 2000 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.2. üzerinde paylaşılmıştır.

Tablo 4.2. Meme kanseri Karar Ağacı için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.98**	**0.54**	**0.96**	**0.76**	**0.96**	**0.775**	**0.74**	**0.78**
100	0.96	0.30	0.93	0.63	0.94	0.650	0.62	0.65
200	0.96	0.30	0.93	0.63	0.94	0.651	0.62	0.65
300	0.97	0.38	0.94	0.67	0.94	0.693	0.66	0.69
400	0.97	0.39	0.94	0.68	0.94	0.694	0.66	0.69
500	0.97	0.35	0.94	0.66	0.94	0.658	0.66	0.66
600	0.97	0.39	0.94	0.68	0.95	0.694	0.66	0.69
700	0.97	0.44	0.95	0.71	0.95	0.732	0.69	0.73
800	0.97	0.41	0.94	0.69	0.95	0.712	0.67	0.71
900	0.97	0.44	0.95	0.70	0.95	0.716	0.69	0.72
1000	0.97	0.43	0.95	0.70	0.95	0.715	0.69	0.72
1100	0.97	0.43	0.95	0.70	0.95	0.715	0.69	0.72
1200	0.97	0.44	0.95	0.70	0.95	0.716	0.69	0.72
1300	0.97	0.34	0.94	0.65	0.94	0.672	0.64	0.67
1400	0.97	0.47	0.95	0.72	0.95	0.753	0.70	0.75
1500	0.97	0.47	0.95	0.72	0.95	0.752	0.70	0.75
1600	0.97	0.46	0.95	0.71	0.95	0.734	0.70	0.73
1700	0.97	0.46	0.95	0.71	0.95	0.734	0.70	0.73
1800	0.97	0.38	0.94	0.67	0.94	0.693	0.66	0.69
1900	0.98	0.51	0.95	0.74	0.96	0.773	0.72	0.77
2000	0.97	0.47	0.95	0.72	0.95	0.753	0.70	0.75

EN İYİ öznitelik grubunun (1900) dağılımı şu şekilde olduğu bulunmuştur; tanı: 1238, ilaç: 411, prosedür: 251. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.1.3. Rastgele Orman sonuçları

Meme kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Rastgele Orman modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL,

İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

612 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.4. de gösterilmiştir. Burada 586 yaşayan hasta için F1 skoru 0.98 ve 26 ölen hasta için F1 skoru 0.07 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.53 ve Doğruluk sonucu 0.96 olarak bulunmuştur. Makro Ort. özelinde %52 Duyarlılığında %98 Kesinlik elde edilmiştir.

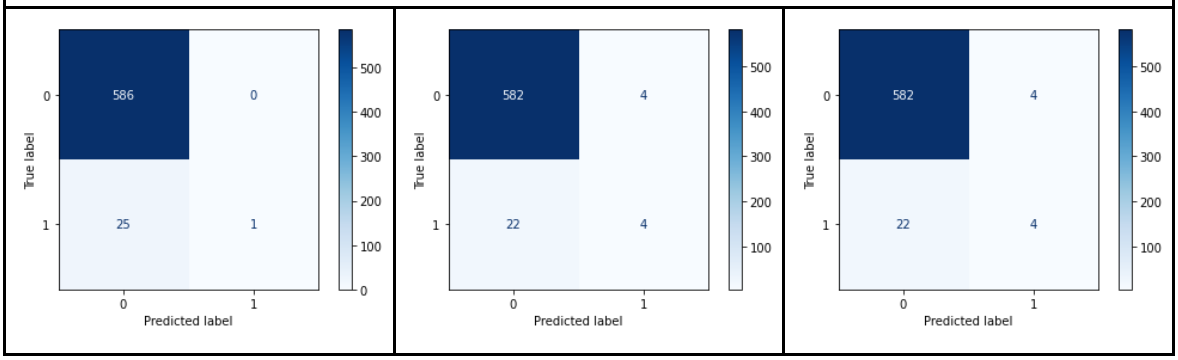
TEMEL öznitelikleri ile 586 yaşayan hastanın 586'sı doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 1'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 0'ı hatalı bir şekilde ölen olarak, ölen hastaların ise 25'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 586 yaşayan hastanın 582'si doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 4'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 4'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 22'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucu geçilmiştir.

EN İYİ öznitelikleri (100) ile 586 yaşayan hastanın 582'si doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 4'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 4'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 22'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada meme kanseri için belirlenmiş toplam özniteliklerin %1'i kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.4. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.96	1.00	0.98
1	1.00	0.04	0.07
accuracy			0.96
macro avg	0.98	0.52	0.53
weighted avg	0.96	0.96	0.94



Şekil 4.4. Meme kanseri Rastgele Orman için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.98, ölen için 0.24 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.61 ve Doğruluk sonucu 0.96 olarak bulunmuştur. Makro Ort. özelinde %57 Duyarlılığında %73 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (100) sınıflandırmaların F1 skoru yaşayan için 0.98, ölen için 0.24 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.61 ve Doğruluk sonucu 0.96 olarak bulunmuştur. Makro Ort. özelinde %57 Duyarlılığında %73 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.894, İLK 100 için: 0.906 ve EN İYİ için: 0.906 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.3. üzerinde paylaşılmıştır.

Tablo 4.3. Meme kanseri Rastgele Orman için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.98**	**0.07**	**0.96**	**0.53**	**0.94**	**0.894**	**0.98**	**0.52**
100	0.98	0.24	0.96	0.61	0.95	0.906	0.73	0.57
200	0.98	0.07	0.96	0.52	0.94	0.903	0.73	0.52
300	0.98	0.14	0.96	0.56	0.94	0.890	0.98	0.54
400	0.98	0.14	0.96	0.56	0.94	0.869	0.81	0.54
500	0.98	0.00	0.96	0.49	0.94	0.925	0.48	0.50
600	0.98	0.14	0.96	0.56	0.94	0.931	0.98	0.54
700	0.98	0.00	0.96	0.49	0.94	0.944	0.48	0.50
800	0.98	0.07	0.96	0.53	0.94	0.884	0.98	0.52
900	0.98	0.07	0.96	0.52	0.94	0.928	0.73	0.52
1000	0.98	0.00	0.95	0.49	0.94	0.917	0.48	0.50
1100	0.98	0.00	0.95	0.49	0.94	0.901	0.48	0.50
1200	0.98	0.07	0.96	0.52	0.94	0.935	0.73	0.52
1300	0.98	0.07	0.96	0.52	0.94	0.905	0.73	0.52
1400	0.98	0.07	0.96	0.53	0.94	0.914	0.98	0.52
1500	0.98	0.14	0.96	0.56	0.94	0.943	0.98	0.54

EN İYİ öznitelik grubunun (100) dağılımı şu şekilde olduğu bulunmuştur; tanı: 73, ilaç: 15, prosedür: 12. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.1.4. Destek Vektör Makinesi sonuçları

Meme kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Destek Vektör Makinesi modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

612 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.5. de gösterilmiştir. Burada 586 yaşayan hasta için F1 skoru 0.98 ve 26 ölen hasta için F1 skoru

0.29 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.63 ve Doğruluk sonucu 0.96 olarak bulunmuştur. Makro Ort. özelinde %59 Duyarlılığında %76 Kesinlik elde edilmiştir.

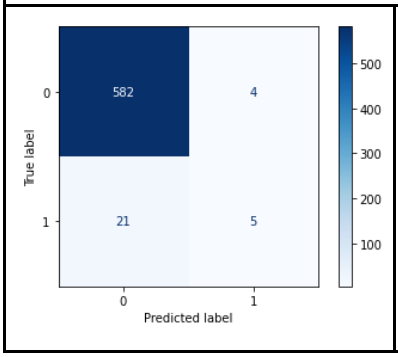
TEMEL öznitelikleri ile 586 yaşayan hastanın 582'si doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 5'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 4'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 21'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

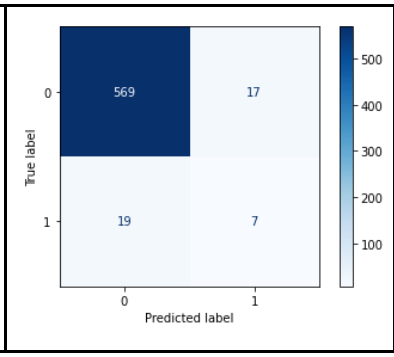
İLK 100 öznitelikleri ile 586 yaşayan hastanın 569'u doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 7'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 17'si hatalı bir şekilde ölen olarak, ölen hastaların ise 19'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna çok yaklaşılmış ancak geçilememiştir.

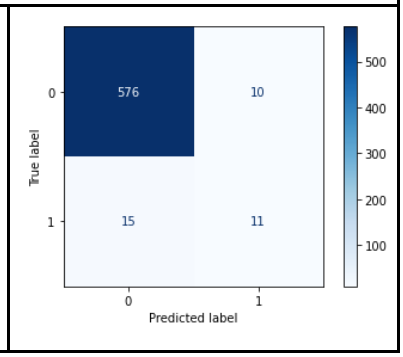
EN İYİ öznitelikleri (400) ile 586 yaşayan hastanın 576'sı doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 11'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 10'u hatalı bir şekilde ölen olarak, ölen hastaların ise 15'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada meme kanseri için belirlenmiş toplam özniteliklerin %5'i kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.5. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.97	0.99	0.98
1	0.56	0.19	0.29
accuracy			0.96
macro avg	0.76	0.59	0.63
weighted avg	0.95	0.96	0.95







Şekil 4.5. Meme kanseri Destek Vektör Makinesi için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.28 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.62 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %62 Duyarlılığında %63 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (400) sınıflandırmaların F1 skoru yaşayan için 0.98, ölen için 0.47 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.72 ve Doğruluk sonucu 0.96 olarak bulunmuştur. Makro Ort. özelinde %70 Duyarlılığında %75 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.921, İLK 100 için: 0.794 ve EN İYİ için: 0.850 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.4. üzerinde paylaşılmıştır.

Tablo 4.4. Meme kanseri Destek Vektör Makinesi için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.98**	**0.29**	**0.96**	**0.63**	**0.95**	**0.921**	**0.76**	**0.59**
100	0.97	0.28	0.94	0.62	0.94	0.794	0.63	0.62
200	0.98	0.46	0.96	0.72	0.96	0.828	0.74	0.70
300	0.97	0.42	0.95	0.70	0.95	0.861	0.70	0.70
400	0.98	0.47	0.96	0.72	0.96	0.850	0.75	0.70
500	0.97	0.30	0.95	0.64	0.94	0.922	0.66	0.62
600	0.97	0.19	0.94	0.58	0.94	0.877	0.61	0.57
700	0.97	0.20	0.95	0.58	0.94	0.896	0.61	0.57
800	0.98	0.22	0.96	0.60	0.94	0.929	0.68	0.57
900	0.98	0.23	0.96	0.60	0.95	0.923	0.70	0.57
1000	0.98	0.19	0.96	0.58	0.94	0.936	0.73	0.56
1100	0.98	0.19	0.96	0.58	0.94	0.938	0.73	0.56
1200	0.98	0.18	0.96	0.58	0.94	0.939	0.70	0.55
1300	0.98	0.18	0.96	0.58	0.94	0.939	0.70	0.55
1400	0.98	0.18	0.96	0.58	0.94	0.946	0.70	0.55
1500	0.98	0.28	0.96	0.63	0.95	0.946	0.73	0.59

EN İYİ öznitelik grubunun (400) dağılımı şu şekilde olduğu bulunmuştur; tanı: 269, ilaç: 90, prosedür: 41. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.1.5. Çok Katmanlı Algılayıcı sonuçları

Meme kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Çok Katmanlı Algılayıcı modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

612 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.6. de gösterilmiştir. Burada 586 yaşayan hasta için F1 skoru 0.98 ve 26 ölen hasta için F1 skoru

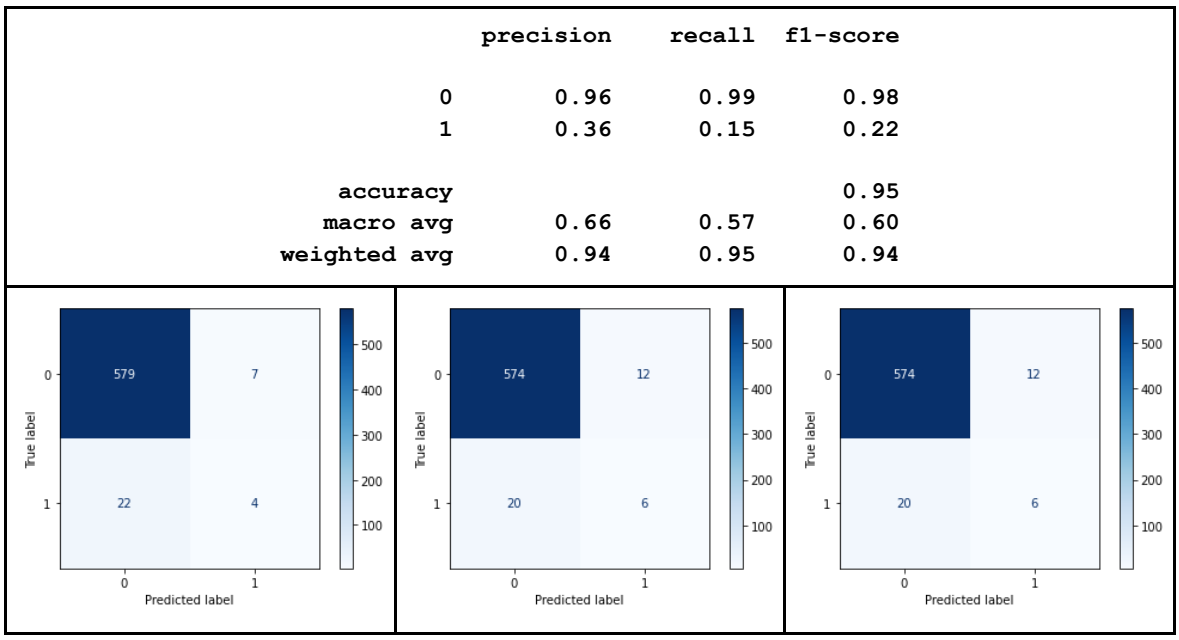
0.22 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.60 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %57 Duyarlılığında %66 Kesinlik elde edilmiştir.

TEMEL öznitelikleri ile 586 yaşayan hastanın 579'u doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 4'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 7'si hatalı bir şekilde ölen olarak, ölen hastaların ise 22'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 586 yaşayan hastanın 574'ü doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 6'sı doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 12'si hatalı bir şekilde ölen olarak, ölen hastaların ise 20'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucu geçilmiştir.

EN İYİ öznitelikleri (100) ile 586 yaşayan hastanın 574'ü doğru bir şekilde yaşayan olarak, 26 ölen hastanın ise 6'sı doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 12'si hatalı bir şekilde ölen olarak, ölen hastaların ise 20'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada meme kanseri için belirlenmiş toplam özniteliklerin %1'i kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.6. üzerinde görselleştirilmiştir.



Şekil 4.6. Meme kanseri Çok Katmanlı Algılayıcı için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.27 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.62 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %61 Duyarlılığında %65 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (100) sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.27 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.62 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %61 Duyarlılığında %65 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.784, İLK 100 için: 0.733 ve EN İYİ için: 0.733 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.5. üzerinde paylaşılmıştır.

Tablo 4.5. Meme kanseri Çok Katmanlı Algılayıcı için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.98**	**0.22**	**0.95**	**0.60**	**0.94**	**0.784**	0.66	0.57
100	0.97	0.27	0.95	0.62	0.94	0.733	0.65	0.61
200	0.97	0.00	0.95	0.49	0.93	0.706	0.48	0.49
300	0.98	0.00	0.96	0.49	0.94	0.677	0.48	0.50
400	0.98	0.00	0.95	0.49	0.94	0.767	0.48	0.50
500	0.96	0.11	0.92	0.53	0.92	0.559	0.53	0.54
600	0.98	0.14	0.96	0.56	0.94	0.512	0.81	0.54
700	0.98	0.00	0.95	0.49	0.94	0.654	0.48	0.50
800	0.98	0.00	0.96	0.49	0.94	0.764	0.48	0.50
900	0.98	0.00	0.96	0.49	0.94	0.576	0.48	0.50
1000	0.98	0.00	0.96	0.49	0.94	0.554	0.48	0.50
1100	0.98	0.00	0.96	0.49	0.94	0.690	0.48	0.50
1200	0.97	0.06	0.95	0.52	0.94	0.696	0.56	0.51
1300	0.98	0.07	0.95	0.52	0.94	0.622	0.60	0.52
1400	0.98	0.00	0.96	0.49	0.94	0.504	0.48	0.50
1500	0.98	0.00	0.96	0.49	0.94	0.617	0.48	0.50

EN İYİ öznitelik grubunun (100) dağılımı şu şekilde olduğu bulunmuştur; tanı: 73, ilaç: 15, prosedür: 12. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.2. Akciğer kanseri

Akciğer kanseri kohort verilerinin test kümesi üzerinde çalıştırılmış 5 modelin; sınıflandırma raporu, karışıklık matrisleri ve 100'erli gruplarla verilmiş öznitelik küme sonuçları bu bölümde gösterilmiş ve değerlendirilmiştir.

4.2.2.1. Lojistik Regresyon sonuçları

Akciğer kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Lojistik Regresyon modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

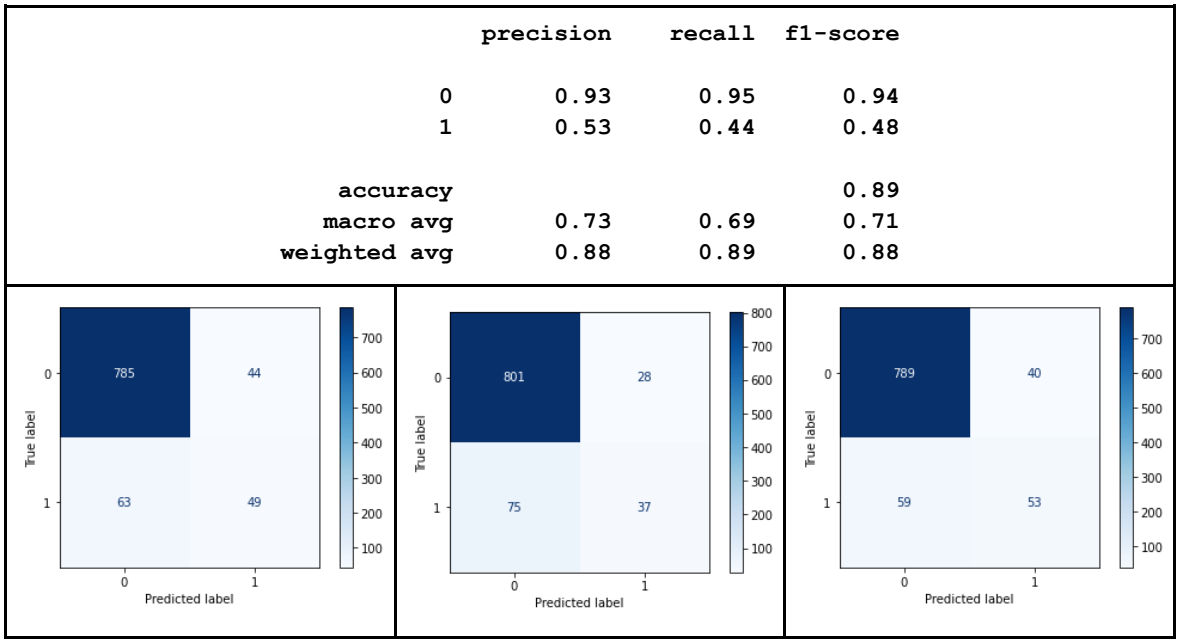
941 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.7. de gösterilmiştir. Burada 829 yaşayan hasta için F1 skoru 0.94 ve 112 ölen hasta için F1 skoru 0.48 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.71 ve Doğruluk sonucu 0.89 olarak bulunmuştur. Makro Ort. özelinde %69 Duyarlılığında %73 Kesinlik elde edilmiştir.

TEMEL öznitelikleri ile 829 yaşayan hastanın 785'i doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 49'u doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 44'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 63'ü hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 829 yaşayan hastanın 801'i doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 37'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 28'i hatalı bir şekilde ölen olarak, ölen hastaların ise 75'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna yaklaşılmış ancak geçilememiştir.

EN İYİ öznitelikleri (1000) ile 829 yaşayan hastanın 789'u doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 53'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 40'ı hatalı bir şekilde ölen olarak, ölen hastaların ise 59'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada akciğer kanseri için belirlenmiş toplam özniteliklerin %10'u kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.7. üzerinde görselleştirilmiştir.



Şekil 4.7. Akciğer kanseri Lojistik Regresyon için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.94, ölen için 0.42 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.68 ve Doğruluk sonucu 0.89 olarak bulunmuştur. Makro Ort. özelinde %65 Duyarlılığında %74 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (1000) sınıflandırmaların F1 skoru yaşayan için 0.94, ölen için 0.52 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.73 ve Doğruluk sonucu 0.89 olarak bulunmuştur. Makro Ort. özelinde %71 Duyarlılığında %75 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.778, İLK 100 için: 0.863 ve EN İYİ için: 0.803 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.6. üzerinde paylaşılmıştır.

Tablo 4.6. Akciğer kanseri Lojistik Regresyon için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.94**	**0.48**	**0.89**	**0.71**	**0.88**	**0.778**	**0.73**	**0.69**
100	0.94	0.42	0.89	0.68	0.88	0.863	0.74	0.65
200	0.94	0.45	0.89	0.69	0.88	0.855	0.74	0.67
300	0.93	0.47	0.88	0.70	0.88	0.834	0.72	0.69
400	0.93	0.47	0.88	0.70	0.88	0.810	0.71	0.69
500	0.93	0.47	0.88	0.70	0.88	0.798	0.72	0.69
600	0.93	0.49	0.88	0.71	0.88	0.813	0.72	0.71
700	0.93	0.51	0.89	0.72	0.88	0.805	0.73	0.72
800	0.93	0.50	0.88	0.72	0.88	0.798	0.72	0.71
900	0.93	0.47	0.88	0.70	0.88	0.801	0.72	0.69
1000	0.94	0.52	0.89	0.73	0.89	0.803	0.75	0.71
1100	0.94	0.51	0.89	0.72	0.89	0.796	0.74	0.71
1200	0.94	0.48	0.89	0.71	0.88	0.792	0.73	0.70
1300	0.94	0.50	0.89	0.72	0.89	0.797	0.74	0.71
1400	0.94	0.51	0.89	0.73	0.89	0.789	0.75	0.71
1500	0.94	0.52	0.89	0.73	0.89	0.787	0.75	0.72

EN İYİ öznitelik grubunun (1000) dağılımı şu şekilde olduğu bulunmuştur; tanı: 715, ilaç: 153, prosedür: 132. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.2.2. Karar Ağacı sonuçları

Akciğer kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Karar Ağacı modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

941 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.8. de gösterilmiştir. Burada 829 yaşayan hasta için F1 skoru 0.93 ve 112 ölen hasta için F1 skoru 0.49 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.71 ve Doğruluk sonucu 0.88 olarak bulunmuştur. Makro Ort. özelinde %71 Duyarlılığında %72 Kesinlik elde edilmiştir.

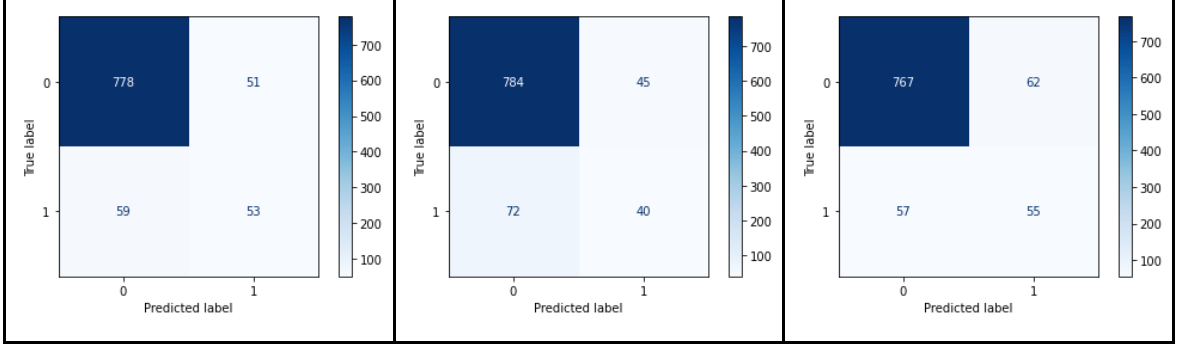
TEMEL öznitelikleri ile 829 yaşayan hastanın 778'i doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 53'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 51'i hatalı bir şekilde ölen olarak, ölen hastaların ise 59'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 829 yaşayan hastanın 784'ü doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 40'ı doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 45'i hatalı bir şekilde ölen olarak, ölen hastaların ise 72'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna yaklaşılamamıştır.

EN İYİ öznitelikleri (1400) ile 829 yaşayan hastanın 767'si doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 55'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 62'si hatalı bir şekilde ölen olarak, ölen hastaların ise 57'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada akciğer kanseri için belirlenmiş toplam özniteliklerin %14'ü kullanılarak TEMEL sonucuna çok yaklaşılmış ancak geçilememiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.8. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.93	0.94	0.93
1	0.51	0.47	0.49
accuracy			0.88
macro avg	0.72	0.71	0.71
weighted avg	0.88	0.88	0.88



Şekil 4.8. Akciğer kanseri Karar Ağacı için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.93, ölen için 0.41 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.67 ve Doğruluk sonucu 0.88 olarak bulunmuştur. Makro Ort. özelinde %65 Duyarlılığında %69 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (1400) sınıflandırmaların F1 skoru yaşayan için 0.93, ölen için 0.48 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.70 ve Doğruluk sonucu 0.87 olarak bulunmuştur. Makro Ort. özelinde %71 Duyarlılığında %70 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.705, İLK 100 için: 0.679 ve EN İYİ için: 0.708 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.7. üzerinde paylaşılmıştır.

Tablo 4.7. Akciğer kanseri Karar Ağacı için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.93**	**0.49**	**0.88**	**0.71**	**0.88**	**0.705**	**0.72**	**0.71**
100	0.93	0.41	0.88	0.67	0.87	0.679	0.69	0.65
200	0.92	0.43	0.87	0.68	0.87	0.694	0.68	0.67
300	0.92	0.37	0.86	0.64	0.85	0.637	0.65	0.64
400	0.92	0.38	0.86	0.65	0.85	0.644	0.66	0.64
500	0.93	0.44	0.87	0.68	0.87	0.678	0.69	0.68
600	0.93	0.44	0.87	0.68	0.87	0.678	0.69	0.68
700	0.93	0.44	0.88	0.69	0.87	0.674	0.71	0.67
800	0.93	0.42	0.88	0.68	0.87	0.658	0.71	0.66
900	0.93	0.42	0.88	0.67	0.87	0.656	0.70	0.66
1000	0.92	0.39	0.86	0.65	0.86	0.648	0.66	0.65
1100	0.93	0.42	0.87	0.67	0.87	0.663	0.68	0.66
1200	0.93	0.44	0.87	0.68	0.87	0.682	0.69	0.68
1300	0.92	0.35	0.85	0.63	0.85	0.626	0.64	0.63
1400	0.93	0.48	0.87	0.70	0.87	0.708	0.70	0.71
1500	0.93	0.41	0.87	0.67	0.87	0.653	0.69	0.65

EN İYİ öznitelik grubunun (1400) dağılımı şu şekilde olduğu bulunmuştur; tanı: 972, ilaç: 232, prosedür: 196. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.2.3. Rastgele Orman sonuçları

Akciğer kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Rastgele Orman modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

941 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.9. de gösterilmiştir. Burada 829 yaşayan hasta için F1 skoru 0.94 ve 112 ölen hasta için F1 skoru

0.05 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.49 ve Doğruluk sonucu 0.88 olarak bulunmuştur. Makro Ort. özelinde %51 Duyarlılığında %82 Kesinlik elde edilmiştir.

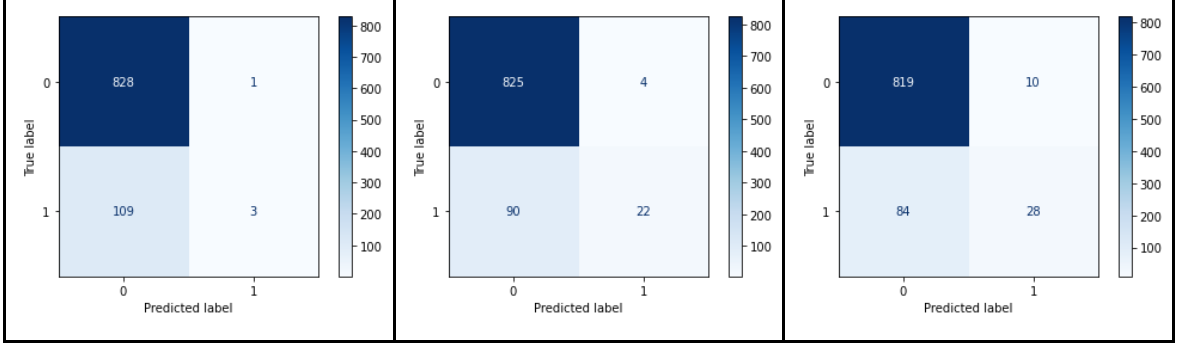
TEMEL öznitelikleri ile 829 yaşayan hastanın 828'i doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 3'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 1'i hatalı bir şekilde ölen olarak, ölen hastaların ise 109'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 829 yaşayan hastanın 825'i doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 22'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 4'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 90'ı hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucu geçilmiştir.

EN İYİ öznitelikleri (200) ile 829 yaşayan hastanın 819'u doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 28'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 10'u hatalı bir şekilde ölen olarak, ölen hastaların ise 84'ü hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada akciğer kanseri için belirlenmiş toplam özniteliklerin %2'si kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.9. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.88	1.00	0.94
1	0.75	0.03	0.05
accuracy			0.88
macro avg	0.82	0.51	0.49
weighted avg	0.87	0.88	0.83



Şekil 4.9. Akciğer kanseri Rastgele Orman için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.95, ölen için 0.32 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.63 ve Doğruluk sonucu 0.90 olarak bulunmuştur. Makro Ort. özelinde %60 Duyarlılığında %87 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (200) sınıflandırmaların F1 skoru yaşayan için 0.95, ölen için 0.37 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.66 ve Doğruluk sonucu 0.90 olarak bulunmuştur. Makro Ort. özelinde %62 Duyarlılığında %82 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.881, İLK 100 için: 0.881 ve EN İYİ için: 0.901 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.8. üzerinde paylaşılmıştır.

Tablo 4.8. Akciğer kanseri Rastgele Orman için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.94**	**0.05**	**0.88**	**0.49**	**0.83**	**0.881**	**0.82**	**0.51**
100	0.95	0.32	0.90	0.63	0.87	0.881	0.87	0.60
200	0.95	0.37	0.90	0.66	0.88	0.901	0.82	0.62
300	0.94	0.26	0.89	0.60	0.86	0.906	0.81	0.58
400	0.95	0.31	0.90	0.63	0.87	0.909	0.87	0.59
500	0.94	0.13	0.89	0.54	0.84	0.913	0.84	0.53
600	0.94	0.18	0.89	0.56	0.85	0.909	0.87	0.55
700	0.94	0.13	0.89	0.53	0.84	0.904	0.78	0.53
800	0.94	0.15	0.89	0.54	0.85	0.905	0.85	0.54
900	0.94	0.12	0.89	0.53	0.84	0.900	0.88	0.53
1000	0.94	0.18	0.89	0.56	0.85	0.901	0.87	0.55
1100	0.94	0.10	0.89	0.52	0.84	0.905	0.82	0.53
1200	0.94	0.05	0.88	0.49	0.83	0.907	0.74	0.51
1300	0.94	0.07	0.88	0.50	0.83	0.909	0.78	0.52
1400	0.94	0.10	0.89	0.52	0.84	0.906	0.82	0.53
1500	0.94	0.05	0.88	0.49	0.83	0.892	0.82	0.51

EN İYİ öznitelik grubunun (200) dağılımı şu şekilde olduğu bulunmuştur; tanı: 165, ilaç: 12, prosedür: 23. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu prosedür ve ilaç öznitelikleri takip etmiştir.

4.2.2.4. Destek Vektör Makinesi sonuçları

Akciğer kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Destek Vektör Makinesi modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

941 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.10. de gösterilmiştir. Burada 829 yaşayan hasta için F1 skoru 0.93 ve 112 ölen hasta için F1 skoru 0.42 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.67 ve Doğruluk sonucu 0.87 olarak bulunmuştur. Makro Ort. özelinde %66 Duyarlılığında %69 Kesinlik elde edilmiştir.

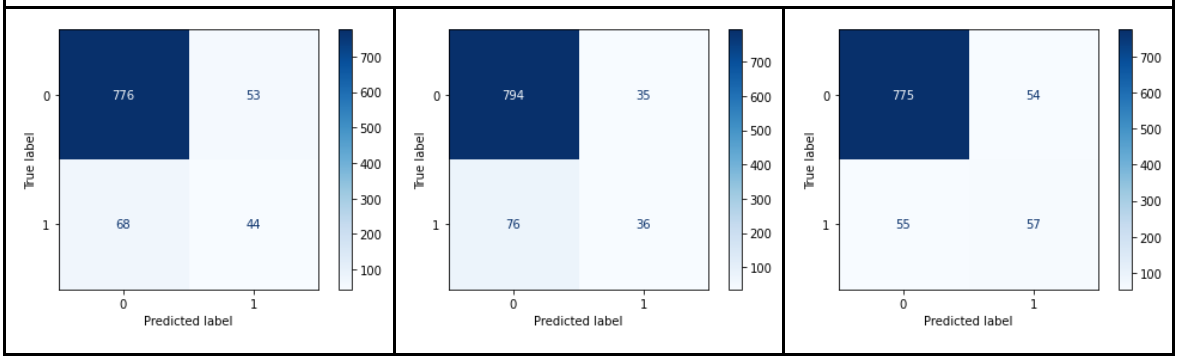
TEMEL öznitelikleri ile 829 yaşayan hastanın 776'sı doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 44'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 53'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 68'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 829 yaşayan hastanın 794'ü doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 36'sı doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 35'i hatalı bir şekilde ölen olarak, ölen hastaların ise 76'sı hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna çok yaklaşılmış ancak geçilememiştir.

EN İYİ öznitelikleri (600) ile 829 yaşayan hastanın 775'i doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 57'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 54'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 55'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada akciğer kanseri için belirlenmiş toplam özniteliklerin %6'sı kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.10. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.92	0.94	0.93
1	0.45	0.39	0.42
accuracy			0.87
macro avg	0.69	0.66	0.67
weighted avg	0.86	0.87	0.87



Şekil 4.10. Akciğer kanseri Destek Vektör Makinesi için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.93, ölen için 0.39 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.66 ve Doğruluk sonucu 0.88 olarak bulunmuştur. Makro Ort. özelinde %64 Duyarlılığında %71 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (600) sınıflandırmaların F1 skoru yaşayan için 0.92, ölen için 0.51 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.72 ve Doğruluk sonucu 0.88 olarak bulunmuştur. Makro Ort. özelinde %72 Duyarlılığında %72 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.826, İLK 100 için: 0.784 ve EN İYİ için: 0.843 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.9. üzerinde paylaşılmıştır.

Tablo 4.9. Akciğer kanseri Destek Vektör Makinesi için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.93**	**0.42**	**0.87**	**0.67**	**0.87**	**0.826**	**0.69**	**0.66**
100	0.93	0.39	0.88	0.66	0.87	0.784	0.71	0.64
200	0.93	0.41	0.87	0.67	0.87	0.824	0.69	0.65
300	0.93	0.43	0.88	0.68	0.87	0.804	0.70	0.67
400	0.93	0.46	0.87	0.69	0.87	0.799	0.69	0.69
500	0.92	0.47	0.87	0.70	0.87	0.801	0.69	0.71
600	0.92	0.51	0.88	0.72	0.88	0.843	0.72	0.72
700	0.93	0.49	0.88	0.71	0.88	0.827	0.71	0.71
800	0.92	0.45	0.87	0.69	0.87	0.799	0.68	0.69
900	0.93	0.46	0.88	0.69	0.87	0.800	0.70	0.69
1000	0.93	0.47	0.87	0.70	0.87	0.812	0.70	0.70
1100	0.93	0.47	0.88	0.70	0.88	0.804	0.71	0.69
1200	0.94	0.49	0.89	0.71	0.88	0.806	0.73	0.70
1300	0.93	0.48	0.88	0.71	0.88	0.806	0.71	0.70
1400	0.94	0.49	0.89	0.71	0.88	0.810	0.72	0.70
1500	0.94	0.49	0.89	0.71	0.88	0.809	0.73	0.70

EN İYİ öznitelik grubunun (600) dağılımı şu şekilde olduğu bulunmuştur; tanı: 439, ilaç: 91, prosedür: 70. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.2.5. Çok Katmanlı Algılayıcı sonuçları

Akciğer kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Çok Katmanlı Algılayıcı modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

941 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.11. de gösterilmiştir. Burada 829 yaşayan hasta için F1 skoru 0.94 ve 112 ölen hasta için F1

skoru 0.31 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.62 ve Doğruluk sonucu 0.88 olarak bulunmuştur. Makro Ort. özelinde %59 Duyarlılığında %72 Kesinlik elde edilmiştir.

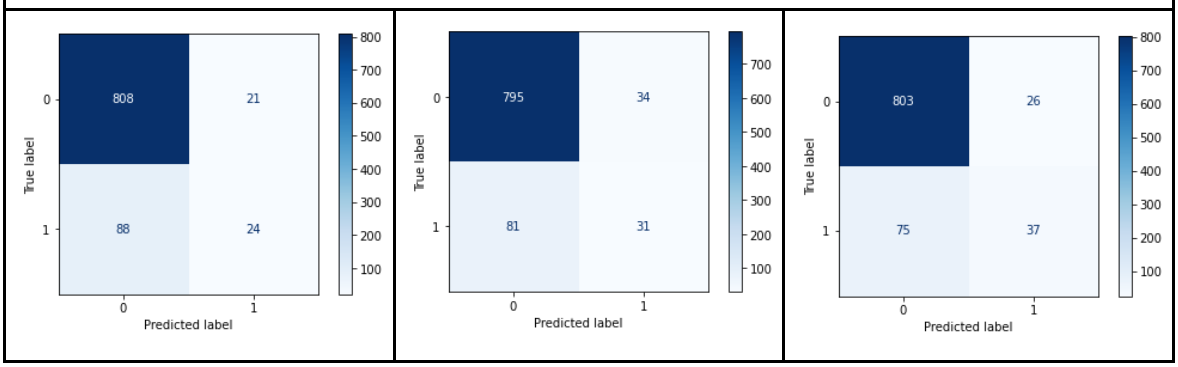
TEMEL öznitelikleri ile 829 yaşayan hastanın 808'i doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 24'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 21'i hatalı bir şekilde ölen olarak, ölen hastaların ise 88'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 829 yaşayan hastanın 795'i doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 31'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 34'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 81'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucu geçilmiştir.

EN İYİ öznitelikleri (1100) ile 829 yaşayan hastanın 803'ü doğru bir şekilde yaşayan olarak, 112 ölen hastanın ise 37'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 26'sı hatalı bir şekilde ölen olarak, ölen hastaların ise 75'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada akciğer kanseri için belirlenmiş toplam özniteliklerin %11'i kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.11. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.90	0.97	0.94
1	0.53	0.21	0.31
accuracy			0.88
macro avg	0.72	0.59	0.62
weighted avg	0.86	0.88	0.86



Şekil 4.11. Akciğer kanseri Çok Katmanlı Algılayıcı için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.93, ölen için 0.35 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.64 ve Doğruluk sonucu 0.88 olarak bulunmuştur. Makro Ort. özelinde %62 Duyarlılığında %69 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (1100) sınıflandırmaların F1 skoru yaşayan için 0.94, ölen için 0.42 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.68 ve Doğruluk sonucu 0.89 olarak bulunmuştur. Makro Ort. özelinde %65 Duyarlılığında %75 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.680, İLK 100 için: 0.726 ve EN İYİ için: 0.702 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.10. üzerinde paylaşılmıştır.

Tablo 4.10. Akciğer kanseri Çok Katmanlı Algılayıcı için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.94**	**0.31**	**0.88**	**0.62**	**0.86**	**0.680**	0.72	0.59
100	0.93	0.35	0.88	0.64	0.86	0.726	0.69	0.62
200	0.94	0.38	0.90	0.66	0.88	0.728	0.78	0.62
300	0.93	0.02	0.88	0.48	0.83	0.724	0.52	0.50
400	0.94	0.27	0.88	0.60	0.86	0.634	0.71	0.58
500	0.93	0.33	0.87	0.63	0.86	0.694	0.67	0.61
600	0.94	0.27	0.89	0.60	0.86	0.622	0.73	0.58
700	0.94	0.18	0.88	0.56	0.85	0.619	0.71	0.55
800	0.94	0.05	0.88	0.49	0.83	0.576	0.63	0.51
900	0.94	0.02	0.88	0.48	0.83	0.624	0.57	0.50
1000	0.93	0.23	0.87	0.58	0.85	0.611	0.65	0.56
1100	0.94	0.42	0.89	0.68	0.88	0.702	0.75	0.65
1200	0.94	0.00	0.88	0.47	0.82	0.513	0.44	0.50
1300	0.93	0.00	0.87	0.47	0.82	0.613	0.44	0.50
1400	0.93	0.38	0.87	0.65	0.86	0.668	0.69	0.64
1500	0.94	0.00	0.88	0.47	0.82	0.566	0.44	0.50

EN İYİ öznitelik grubunun (1100) dağılımı şu şekilde olduğu bulunmuştur; tanı: 781, ilaç: 173, prosedür: 146. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.3. Prostat Kanseri

Prostat kanseri kohort verilerinin test kümesi üzerinde çalıştırılmış 5 modelin; sınıflandırma raporu, karışıklık matrisleri ve 100'erli gruplarla verilmiş öznitelik küme sonuçları bu bölümde gösterilmiş ve değerlendirilmiştir.

4.2.3.1. Lojistik Regresyon sonuçları

Prostat kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Lojistik Regresyon modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

579 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.12. de gösterilmiştir. Burada 541 yaşayan hasta için F1 skoru 0.97 ve 38 ölen hasta için F1 skoru 0.55 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.76 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %72 Duyarlılığında %84 Kesinlik elde edilmiştir.

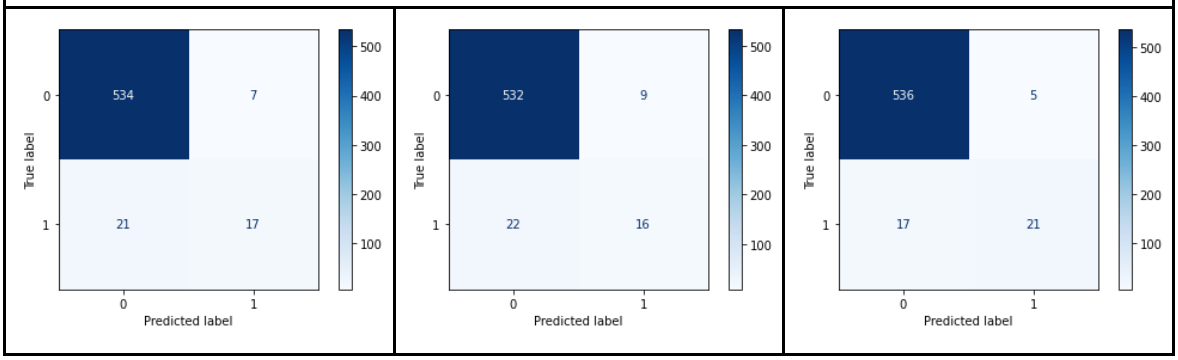
TEMEL öznitelikleri ile 541 yaşayan hastanın 534'ü doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 17'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 7'si hatalı bir şekilde ölen olarak, ölen hastaların ise 21'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 541 yaşayan hastanın 532'si doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 16'sı doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 9'u hatalı bir şekilde ölen olarak, ölen hastaların ise 22'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna yaklaşılmış ancak geçilememiştir.

EN İYİ öznitelikleri (700) ile 541 yaşayan hastanın 536'sı doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 21'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 5'i hatalı bir şekilde ölen olarak, ölen hastaların ise 17'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada prostat kanseri için belirlenmiş toplam özniteliklerin %8'i kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.12. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.96	0.99	0.97
1	0.71	0.45	0.55
accuracy			0.95
macro avg	0.84	0.72	0.76
weighted avg	0.95	0.95	0.95



Şekil 4.12. Prostat kanseri Lojistik Regresyon için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.51 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.74 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %70 Duyarlılığında %80 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (700) sınıflandırmaların F1 skoru yaşayan için 0.98, ölen için 0.66 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.82 ve Doğruluk sonucu 0.96 olarak bulunmuştur. Makro Ort. özelinde %77 Duyarlılığında %89 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.746, İLK 100 için: 0.931 ve EN İYİ için: 0.803 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.11. üzerinde paylaşılmıştır.

Tablo 4.11. Prostat kanseri Lojistik Regresyon için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.97**	**0.55**	**0.95**	**0.76**	**0.95**	**0.746**	**0.84**	**0.72**
100	0.97	0.51	0.95	0.74	0.94	0.931	0.80	0.70
200	0.97	0.41	0.94	0.69	0.93	0.890	0.76	0.65
300	0.96	0.37	0.93	0.67	0.92	0.799	0.71	0.64
400	0.97	0.44	0.94	0.71	0.93	0.823	0.76	0.67
500	0.97	0.48	0.94	0.72	0.94	0.795	0.76	0.70
600	0.97	0.52	0.95	0.74	0.94	0.783	0.81	0.70
700	0.98	0.66	0.96	0.82	0.96	0.803	0.89	0.77
800	0.98	0.58	0.96	0.78	0.95	0.788	0.86	0.73
900	0.98	0.58	0.96	0.78	0.95	0.778	0.86	0.73
1000	0.98	0.57	0.95	0.77	0.95	0.772	0.84	0.73
1100	0.98	0.58	0.96	0.78	0.95	0.774	0.86	0.73
1200	0.98	0.57	0.95	0.77	0.95	0.776	0.84	0.73
1300	0.97	0.55	0.95	0.76	0.95	0.799	0.84	0.72
1400	0.97	0.55	0.95	0.76	0.95	0.799	0.84	0.72
1500	0.98	0.58	0.96	0.78	0.95	0.789	0.86	0.73

EN İYİ öznitelik grubunun (700) dağılımı şu şekilde olduğu bulunmuştur; tanı: 487, ilaç: 119, prosedür: 94. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.3.2. Karar Ağacı sonuçları

Prostat kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Karar Ağacı modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

579 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.13. de gösterilmiştir. Burada 541 yaşayan hasta için F1 skoru 0.98 ve 38 ölen hasta için F1 skoru

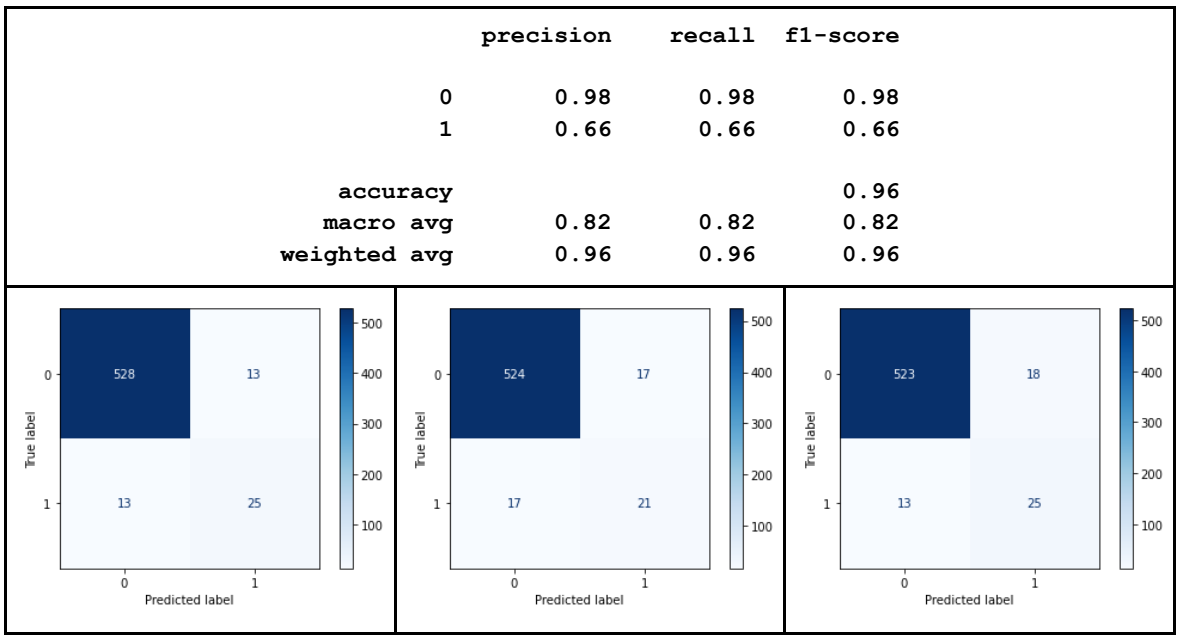
0.66 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.82 ve Doğruluk sonucu 0.96 olarak bulunmuştur. Makro Ort. özelinde %82 Duyarlılığında %82 Kesinlik elde edilmiştir.

TEMEL öznitelikleri ile 541 yaşayan hastanın 528'i doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 25'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 7'si hatalı bir şekilde ölen olarak, ölen hastaların ise 21'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 541 yaşayan hastanın 524'ü doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 21'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 17'si hatalı bir şekilde ölen olarak, ölen hastaların ise 17'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna yaklaşılamamıştır.

EN İYİ öznitelikleri (1000) ile 541 yaşayan hastanın 523'ü doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 25'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 18'i hatalı bir şekilde ölen olarak, ölen hastaların ise 13'ü hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada prostat kanseri için belirlenmiş toplam özniteliklerin %12'si kullanılarak TEMEL sonucuna yaklaşılmış ancak geçilememiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.13. üzerinde görselleştirilmiştir.



Şekil 4.13. Prostat kanseri Karar Ağacı için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.55 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.76 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %76 Duyarlılığında %76 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (700) sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.62 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.79 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %81 Duyarlılığında %78 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.799, İLK 100 için: 0.770 ve EN İYİ için: 0.812 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.12. üzerinde paylaşılmıştır.

Tablo 4.12. Prostat kanseri Karar Ağacı için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.98**	**0.66**	**0.96**	**0.82**	**0.96**	**0.799**	**0.82**	**0.82**
100	0.97	0.55	0.94	0.76	0.94	0.770	0.76	0.76
200	0.97	0.58	0.95	0.78	0.95	0.764	0.79	0.76
300	0.97	0.50	0.94	0.73	0.94	0.722	0.75	0.72
400	0.97	0.53	0.94	0.75	0.94	0.757	0.74	0.76
500	0.97	0.49	0.94	0.73	0.94	0.721	0.74	0.72
600	0.97	0.55	0.94	0.76	0.94	0.749	0.77	0.75
700	0.97	0.58	0.94	0.77	0.94	0.774	0.77	0.77
800	0.96	0.45	0.93	0.71	0.93	0.694	0.72	0.79
900	0.97	0.57	0.94	0.77	0.94	0.785	0.76	0.79
1000	0.97	0.62	0.95	0.79	0.95	0.812	0.78	0.81
1100	0.96	0.50	0.93	0.73	0.93	0.742	0.72	0.74
1200	0.97	0.54	0.94	0.75	0.94	0.758	0.75	0.76
1300	0.97	0.47	0.93	0.72	0.93	0.707	0.73	0.71
1400	0.97	0.52	0.94	0.74	0.94	0.745	0.74	0.75
1500	0.97	0.47	0.93	0.72	0.93	0.707	0.73	0.71

EN İYİ öznitelik grubunun (1000) dağılımı şu şekilde olduğu bulunmuştur; tanı: 679, ilaç: 177, prosedür: 144. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.3.3. Rastgele Orman sonuçları

Prostat kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Rastgele Orman modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

579 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.14. de gösterilmiştir. Burada 541 yaşayan hasta için F1 skoru 0.97 ve 38 ölen hasta için F1 skoru 0.10 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.53 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %53 Duyarlılığında %97 Kesinlik elde edilmiştir.

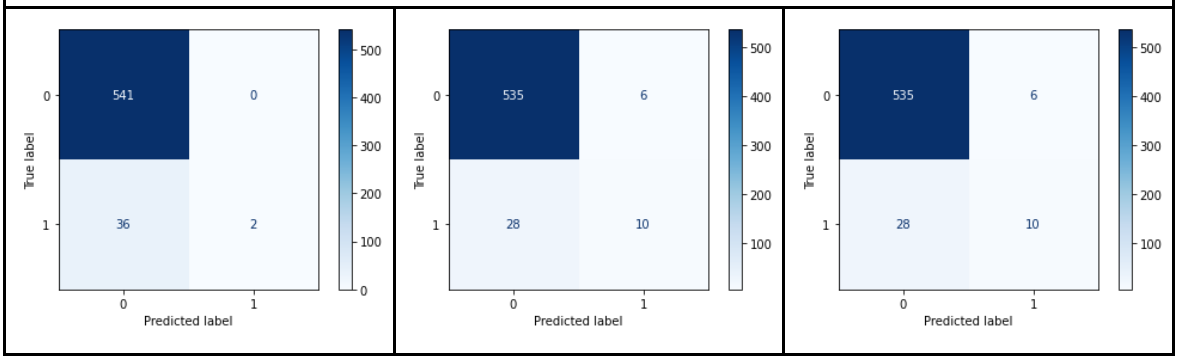
TEMEL öznitelikleri ile 541 yaşayan hastanın 541'i doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 2'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 0'ı hatalı bir şekilde ölen olarak, ölen hastaların ise 36'sı hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 541 yaşayan hastanın 535'i doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 10'u doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 6'sı hatalı bir şekilde ölen olarak, ölen hastaların ise 28'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucu geçilmiştir.

EN İYİ öznitelikleri (100) ile 541 yaşayan hastanın 535'i doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 10'u doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 6'sı hatalı bir şekilde ölen olarak, ölen hastaların ise 28'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada prostat kanseri için belirlenmiş toplam özniteliklerin %1'i kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.14. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.94	1.00	0.97
1	1.00	0.05	0.10
accuracy			0.94
macro avg	0.97	0.53	0.53
weighted avg	0.94	0.94	0.91



Şekil 4.14. Prostat kanseri Rastgele Orman için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.37 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.67 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %63 Duyarlılığında %79 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (100) sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.37 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.67 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %63 Duyarlılığında %79 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.941, İLK 100 için: 0.938 ve EN İYİ için: 0.938 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.13. üzerinde paylaşılmıştır.

Tablo 4.13. Prostat kanseri Rastgele Orman için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.97**	**0.10**	**0.94**	**0.53**	**0.91**	**0.941**	**0.97**	**0.53**
100	0.97	0.37	0.94	0.67	0.93	0.938	0.79	0.63
200	0.97	0.33	0.94	0.65	0.93	0.951	0.87	0.60
300	0.97	0.22	0.94	0.60	0.92	0.961	0.83	0.56
400	0.97	0.19	0.94	0.58	0.92	0.954	0.87	0.55
500	0.97	0.19	0.94	0.58	0.92	0.939	0.87	0.55
600	0.97	0.23	0.94	0.60	0.92	0.941	0.89	0.56
700	0.97	0.18	0.94	0.57	0.91	0.947	0.76	0.55
800	0.97	0.15	0.94	0.56	0.91	0.954	0.97	0.54
900	0.97	0.22	0.94	0.60	0.92	0.954	0.83	0.56
1000	0.97	0.10	0.94	0.53	0.91	0.953	0.97	0.53
1100	0.97	0.10	0.94	0.53	0.91	0.950	0.97	0.53
1200	0.97	0.14	0.94	0.56	0.91	0.950	0.84	0.54
1300	0.97	0.10	0.94	0.53	0.91	0.952	0.80	0.53
1400	0.97	0.10	0.94	0.53	0.91	0.957	0.97	0.53
1500	0.97	0.05	0.94	0.51	0.91	0.942	0.72	0.51

EN İYİ öznitelik grubunun (100) dağılımı şu şekilde olduğu bulunmuştur; tanı: 82, ilaç: 4, prosedür: 14. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu prosedür ve ilaç öznitelikleri takip etmiştir.

4.2.3.4. Destek Vektör Makinesi sonuçları

Prostat kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Destek Vektör Makinesi modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

579 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.15. de gösterilmiştir. Burada 541 yaşayan hasta için F1 skoru 0.97 ve 38 ölen hasta için F1 skoru

0.43 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.70 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %66 Duyarlılığında %76 Kesinlik elde edilmiştir.

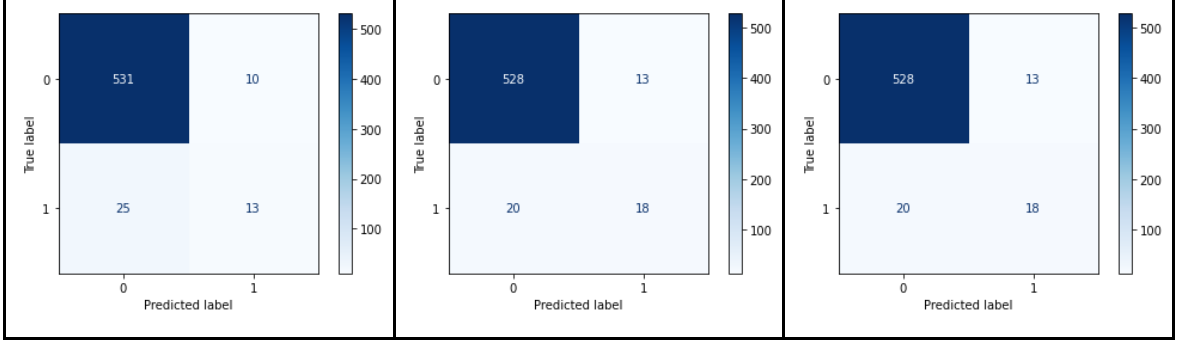
TEMEL öznitelikleri ile 541 yaşayan hastanın 531'i doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 13'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 10'u hatalı bir şekilde ölen olarak, ölen hastaların ise 25'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 541 yaşayan hastanın 528'i doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 18'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 13'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 20'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucu geçilmiştir.

EN İYİ öznitelikleri (100) ile 541 yaşayan hastanın 528'i doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 18'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 13'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 20'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada prostat kanseri için belirlenmiş toplam özniteliklerin %1'i kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.15. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.96	0.98	0.97
1	0.57	0.34	0.43
accuracy			0.94
macro avg	0.76	0.66	0.70
weighted avg	0.93	0.94	0.93



Şekil 4.15. Prostat kanseri Destek Vektör Makinesi için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.52 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.75 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %72 Duyarlılığında %77 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (100) sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.52 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.75 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %72 Duyarlılığında %77 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.849, İLK 100 için: 0.785 ve EN İYİ için: 0.785 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.14. üzerinde paylaşılmıştır.

Tablo 4.14. Prostat kanseri Destek Vektör Makinesi için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.97**	**0.43**	**0.94**	**0.70**	**0.93**	**0.849**	**0.76**	**0.66**
100	0.97	0.52	0.94	0.75	0.94	0.785	0.77	0.72
200	0.97	0.46	0.94	0.71	0.93	0.913	0.76	0.69
300	0.96	0.42	0.93	0.69	0.93	0.895	0.72	0.67
400	0.96	0.38	0.93	0.67	0.92	0.935	0.69	0.65
500	0.96	0.35	0.92	0.65	0.92	0.924	0.67	0.64
600	0.96	0.43	0.93	0.70	0.93	0.929	0.71	0.68
700	0.97	0.41	0.94	0.69	0.93	0.916	0.74	0.66
800	0.97	0.43	0.94	0.70	0.93	0.903	0.74	0.67
900	0.97	0.44	0.94	0.70	0.93	0.896	0.75	0.67
1000	0.97	0.44	0.94	0.70	0.93	0.880	0.75	0.67
1100	0.97	0.41	0.94	0.69	0.93	0.858	0.74	0.66
1200	0.97	0.41	0.94	0.69	0.93	0.888	0.74	0.66
1300	0.97	0.45	0.94	0.71	0.93	0.899	0.75	0.69
1400	0.97	0.46	0.94	0.71	0.93	0.897	0.76	0.69
1500	0.97	0.46	0.94	0.71	0.93	0.895	0.76	0.69

EN İYİ öznitelik grubunun (100) dağılımı şu şekilde olduğu bulunmuştur; tanı: 82, ilaç: 4, prosedür: 14. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu prosedür ve ilaç öznitelikleri takip etmiştir.

4.2.3.5. Çok Katmanlı Algılayıcı sonuçları

Prostat kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Çok Katmanlı Algılayıcı modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

579 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.16. de gösterilmiştir. Burada 541 yaşayan hasta için F1 skoru 0.96 ve 38 ölen hasta için F1 skoru

0.39 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.68 ve Doğruluk sonucu 0.93 olarak bulunmuştur. Makro Ort. özelinde %66 Duyarlılığında %71 Kesinlik elde edilmiştir.

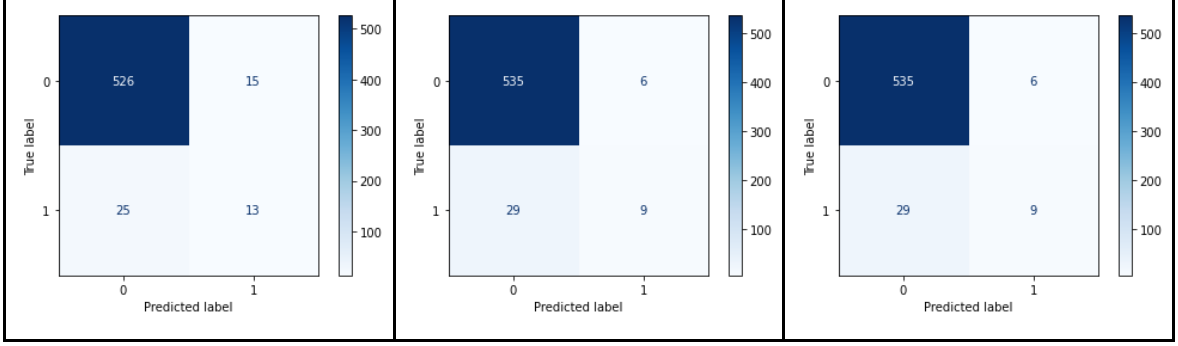
TEMEL öznitelikleri ile 541 yaşayan hastanın 526'sı doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 13'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 15'i hatalı bir şekilde ölen olarak, ölen hastaların ise 25'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 541 yaşayan hastanın 535'i doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 9'u doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 6'sı hatalı bir şekilde ölen olarak, ölen hastaların ise 29'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna yaklaşılmış ancak geçilememiştir.

EN İYİ öznitelikleri (100) ile 541 yaşayan hastanın 535'i doğru bir şekilde yaşayan olarak, 38 ölen hastanın ise 9'u doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 6'sı hatalı bir şekilde ölen olarak, ölen hastaların ise 29'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada prostat kanseri için belirlenmiş toplam özniteliklerin %1'i kullanılarak TEMEL sonucuna yaklaşılmış ancak geçilememiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.16. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.95	0.97	0.96
1	0.46	0.34	0.39
accuracy			0.93
macro avg	0.71	0.66	0.68
weighted avg	0.92	0.93	0.93



Şekil 4.16. Prostat kanseri Çok Katmanlı Algılayıcı için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.34 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.65 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %61 Duyarlılığında %77 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (100) sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.34 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.65 ve Doğruluk sonucu 0.94 olarak bulunmuştur. Makro Ort. özelinde %61 Duyarlılığında %77 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.585, İLK 100 için: 0.776 ve EN İYİ için: 0.776 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.15. üzerinde paylaşılmıştır.

Tablo 4.15. Prostat kanseri Çok Katmanlı Algılayıcı için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.96**	**0.39**	**0.93**	**0.68**	**0.93**	**0.585**	**0.71**	**0.66**
100	0.97	0.34	0.94	0.65	0.93	0.776	0.77	0.61
200	0.97	0.31	0.94	0.64	0.93	0.749	0.78	0.60
300	0.97	0.00	0.93	0.48	0.90	0.683	0.47	0.50
400	0.97	0.00	0.93	0.48	0.90	0.609	0.47	0.50
500	0.96	0.00	0.93	0.48	0.90	0.655	0.47	0.50
600	0.96	0.26	0.93	0.61	0.92	0.558	0.69	0.58
700	0.97	0.18	0.94	0.57	0.91	0.579	0.76	0.55
800	0.96	0.00	0.93	0.48	0.90	0.580	0.47	0.50
900	0.97	0.00	0.93	0.48	0.90	0.510	0.47	0.50
1000	0.97	0.21	0.93	0.59	0.92	0.686	0.72	0.56
1100	0.96	0.15	0.92	0.56	0.91	0.570	0.60	0.54
1200	0.96	0.00	0.92	0.48	0.90	0.586	0.47	0.49
1300	0.97	0.36	0.94	0.67	0.93	0.746	0.77	0.63
1400	0.97	0.00	0.93	0.48	0.90	0.478	0.47	0.50
1500	0.97	0.00	0.93	0.48	0.90	0.622	0.47	0.50

EN İYİ öznitelik grubunun (100) dağılımı şu şekilde olduğu bulunmuştur; tanı: 82, ilaç: 4, prosedür: 14. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu prosedür ve ilaç öznitelikleri takip etmiştir.

4.2.4. Mide Kanseri

Mide kanseri kohort verilerinin test kümesi üzerinde çalıştırılmış 5 modelin; sınıflandırma raporu, karışıklık matrisleri ve 100'erli gruplarla verilmiş öznitelik küme sonuçları bu bölümde gösterilmiş ve değerlendirilmiştir.

4.2.4.1. Lojistik Regresyon sonuçları

Mide kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Lojistik Regresyon modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

168 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.17. de gösterilmiştir. Burada 156 yaşayan hasta için F1 skoru 0.97 ve 12 ölen hasta için F1 skoru 0.48 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.72 ve Doğruluk sonucu 0.93 olarak bulunmuştur. Makro Ort. özelinde %70 Duyarlılığında %76 Kesinlik elde edilmiştir.

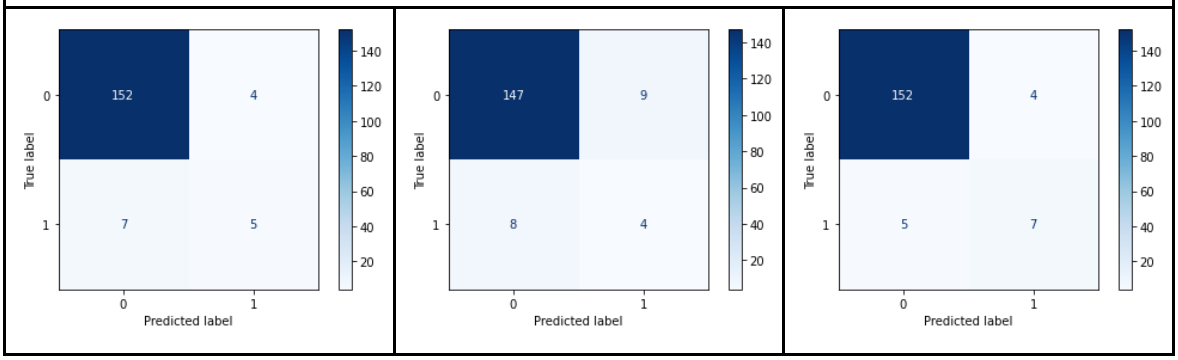
TEMEL öznitelikleri ile 156 yaşayan hastanın 152'si, doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 5'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 4'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 7'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 156 yaşayan hastanın 147'si doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 4'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 9'u hatalı bir şekilde ölen olarak, ölen hastaların ise 8'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna yaklaşılamamıştır.

EN İYİ öznitelikleri (600) ile 156 yaşayan hastanın 152'si doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 7'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 4'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 5'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada mide kanseri için belirlenmiş toplam özniteliklerin %12'si kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.17. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.96	0.97	0.97
1	0.56	0.42	0.48
accuracy			0.93
macro avg	0.76	0.70	0.72
weighted avg	0.93	0.93	0.93



Şekil 4.17. Mide kanseri Lojistik Regresyon için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.95, ölen için 0.32 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.63 ve Doğruluk sonucu 0.90 olarak bulunmuştur. Makro Ort. özelinde %64 Duyarlılığında %63 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (100) sınıflandırmaların F1 skoru yaşayan için 0.97, ölen için 0.61 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.79 ve Doğruluk sonucu 0.95 olarak bulunmuştur. Makro Ort. özelinde %78 Duyarlılığında %80 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.804, İLK 100 için: 0.653 ve EN İYİ için: 0.830 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.16. üzerinde paylaşılmıştır.

Tablo 4.16. Mide kanseri Lojistik Regresyon için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.97**	**0.48**	**0.93**	**0.72**	**0.93**	**0.804**	**0.76**	**0.70**
100	0.95	0.32	0.90	0.63	0.90	0.653	0.63	0.64
200	0.95	0.38	0.90	0.67	0.91	0.741	0.66	0.68
300	0.93	0.32	0.88	0.63	0.89	0.772	0.61	0.66
400	0.95	0.40	0.91	0.68	0.91	0.790	0.67	0.68
500	0.96	0.50	0.93	0.73	0.93	0.821	0.73	0.73
600	0.97	0.61	0.95	0.79	0.95	0.830	0.80	0.78
700	0.97	0.61	0.95	0.79	0.95	0.827	0.80	0.78
800	0.97	0.55	0.94	0.76	0.94	0.819	0.78	0.74
900	0.97	0.48	0.93	0.72	0.93	0.801	0.76	0.70
1000	0.97	0.48	0.93	0.72	0.93	0.803	0.76	0.70
1100	0.97	0.48	0.93	0.72	0.93	0.805	0.76	0.70
1200	0.97	0.48	0.93	0.72	0.93	0.803	0.76	0.70
1300	0.97	0.48	0.93	0.72	0.93	0.802	0.76	0.70
1400	0.97	0.48	0.93	0.72	0.93	0.802	0.76	0.70
1500	0.97	0.48	0.93	0.72	0.93	0.801	0.76	0.70

EN İYİ öznitelik grubunun (600) dağılımı şu şekilde olduğu bulunmuştur; tanı: 380, ilaç: 115, prosedür: 105. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.4.2. Karar Ağacı sonuçları

Mide kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Karar Ağacı modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

168 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.18. de gösterilmiştir. Burada 156 yaşayan hasta için F1 skoru 0.92 ve 12 ölen hasta için F1 skoru

0.37 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.64 ve Doğruluk sonucu 0.86 olarak bulunmuştur. Makro Ort. özelinde %73 Duyarlılığında %62 Kesinlik elde edilmiştir.

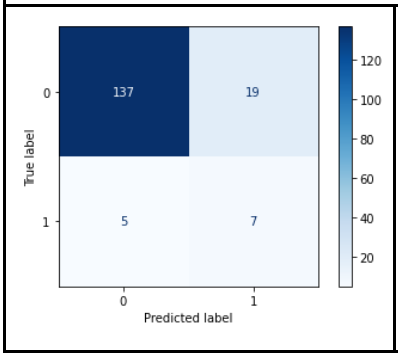
TEMEL öznitelikleri ile 156 yaşayan hastanın 137'si, doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 7'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 19'u hatalı bir şekilde ölen olarak, ölen hastaların ise 5'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

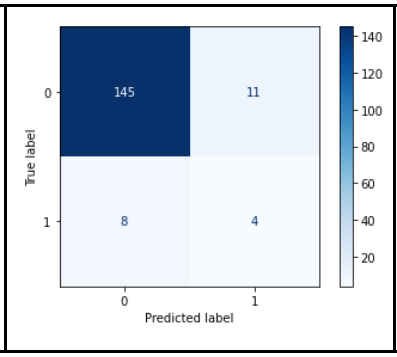
İLK 100 öznitelikleri ile 156 yaşayan hastanın 145'i doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 4'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 11'i hatalı bir şekilde ölen olarak, ölen hastaların ise 8'i hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna yaklaşılmış ancak geçilememiştir.

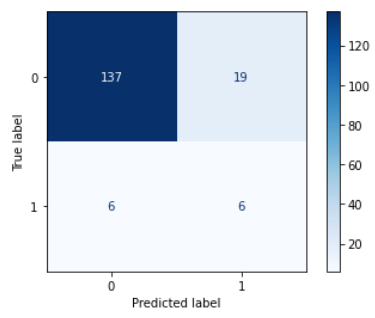
EN İYİ öznitelikleri (700) ile 156 yaşayan hastanın 137'si doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 6'sı doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 19'u hatalı bir şekilde ölen olarak, ölen hastaların ise 6'sı hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada mide kanseri için belirlenmiş toplam özniteliklerin %14'ü kullanılarak TEMEL sonucuna yaklaşılmış ancak geçilememiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.18. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.96	0.88	0.92
1	0.27	0.58	0.37
accuracy			0.86
macro avg	0.62	0.73	0.64
weighted avg	0.92	0.86	0.88







Şekil 4.18. Mide kanseri Karar Ağacı için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.94, ölen için 0.30 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.62 ve Doğruluk sonucu 0.89 olarak bulunmuştur. Makro Ort. özelinde %63 Duyarlılığında %61 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (700) sınıflandırmaların F1 skoru yaşayan için 0.92, ölen için 0.32 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.62 ve Doğruluk sonucu 0.85 olarak bulunmuştur. Makro Ort. özelinde %69 Duyarlılığında %60 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.730, İLK 100 için: 0.605 ve EN İYİ için: 0.689 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.17. üzerinde paylaşılmıştır.

Tablo 4.17. Mide kanseri Karar Ağacı için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.92**	**0.37**	**0.86**	**0.64**	**0.88**	**0.730**	**0.62**	**0.73**
100	0.94	0.30	0.89	0.62	0.89	0.605	0.61	0.63
200	0.93	0.27	0.87	0.60	0.88	0.621	0.58	0.62
300	0.91	0.28	0.85	0.60	0.87	0.647	0.58	0.65
400	0.93	0.28	0.88	0.60	0.88	0.624	0.59	0.62
500	0.93	0.27	0.87	0.60	0.88	0.621	0.58	0.62
600	0.92	0.25	0.86	0.59	0.87	0.615	0.57	0.62
700	0.92	0.32	0.85	0.62	0.87	0.689	0.60	0.69
800	0.91	0.26	0.83	0.58	0.86	0.641	0.57	0.64
900	0.91	0.32	0.85	0.61	0.87	0.685	0.59	0.69
1000	0.91	0.24	0.85	0.57	0.87	0.608	0.56	0.61
1100	0.93	0.21	0.86	0.57	0.87	0.580	0.56	0.58
1200	0.93	0.28	0.88	0.60	0.88	0.624	0.59	0.62
1300	0.92	0.24	0.85	0.58	0.87	0.612	0.57	0.61
1400	0.92	0.13	0.85	0.52	0.86	0.532	0.52	0.53
1500	0.94	0.29	0.88	0.61	0.89	0.628	0.60	0.63

EN İYİ öznitelik grubunun (700) dağılımı şu şekilde olduğu bulunmuştur; tanı: 448, ilaç: 132, prosedür: 120. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu ilaç ve prosedür öznitelikleri takip etmiştir.

4.2.4.3. Rastgele Orman sonuçları

Mide kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Rastgele Orman modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

168 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.19. de gösterilmiştir. Burada 156 yaşayan hasta için F1 skoru 0.96 ve 12 ölen hasta için F1 skoru

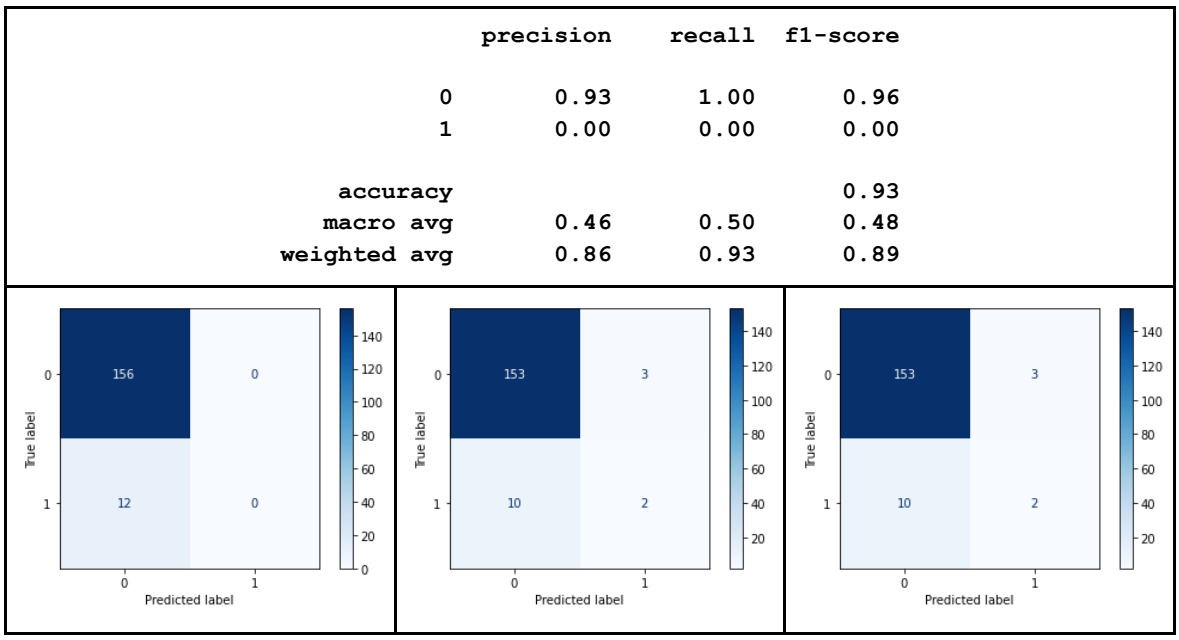
0.00 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.48 ve Doğruluk sonucu 0.93 olarak bulunmuştur. Makro Ort. özelinde %50 Duyarlılığında %46 Kesinlik elde edilmiştir.

TEMEL öznitelikleri ile 156 yaşayan hastanın 156'sı, doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 0'ı doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 0'ı hatalı bir şekilde ölen olarak, ölen hastaların ise 12'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 156 yaşayan hastanın 153'ü doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 2'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 3'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 10'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucu geçilmiştir.

EN İYİ öznitelikleri (100) ile 156 yaşayan hastanın 153'ü doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 2'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 3'ü hatalı bir şekilde ölen olarak, ölen hastaların ise 10'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada mide kanseri için belirlenmiş toplam özniteliklerin %2'si kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.19. üzerinde görselleştirilmiştir.



Şekil 4.19. Mide kanseri Rastgele Orman için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.96, ölen için 0.24 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.60 ve Doğruluk sonucu 0.92 olarak bulunmuştur. Makro Ort. özelinde %57 Duyarlılığında %67 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (100) sınıflandırmaların F1 skoru yaşayan için 0.96, ölen için 0.24 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.60 ve Doğruluk sonucu 0.92 olarak bulunmuştur. Makro Ort. özelinde %57 Duyarlılığında %67 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.848, İLK 100 için: 0.841 ve EN İYİ için: 0.841 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.18. üzerinde paylaşılmıştır.

Tablo 4.18. Mide kanseri Rastgele Orman için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.96**	**0.00**	**0.93**	**0.48**	**0.89**	**0.848**	**0.46**	**0.50**
100	0.96	0.24	0.92	0.60	0.91	0.841	0.67	0.57
200	0.95	0.00	0.90	0.47	0.88	0.875	0.46	0.48
300	0.96	0.00	0.92	0.48	0.89	0.887	0.46	0.49
400	0.95	0.00	0.91	0.48	0.89	0.846	0.46	0.49
500	0.95	0.00	0.91	0.48	0.89	0.871	0.46	0.49
600	0.95	0.00	0.91	0.48	0.89	0.872	0.46	0.49
700	0.96	0.00	0.92	0.48	0.89	0.864	0.46	0.50
800	0.95	0.00	0.91	0.48	0.89	0.851	0.46	0.49
900	0.95	0.00	0.91	0.48	0.89	0.843	0.46	0.49
1000	0.95	0.00	0.91	0.48	0.89	0.851	0.46	0.50
1100	0.96	0.00	0.92	0.48	0.89	0.859	0.46	0.49
1200	0.96	0.00	0.92	0.48	0.89	0.851	0.46	0.49
1300	0.96	0.00	0.92	0.48	0.89	0.851	0.46	0.49
1400	0.96	0.00	0.93	0.48	0.89	0.858	0.46	0.50
1500	0.96	0.00	0.92	0.48	0.89	0.845	0.46	0.50

EN İYİ öznitelik grubunun (100) dağılımı şu şekilde olduğu bulunmuştur; tanı: 77, ilaç: 8, prosedür: 15. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu prosedür ve ilaç öznitelikleri takip etmiştir.

4.2.4.4. Destek Vektör Makinesi sonuçları

Mide kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Destek Vektör Makinesi modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

168 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.20. de gösterilmiştir. Burada 156 yaşayan hasta için F1 skoru 0.95 ve 12 ölen hasta için F1 skoru 0.20 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.57 ve Doğruluk sonucu 0.90 olarak bulunmuştur. Makro Ort. özelinde %56 Duyarlılığında %59 Kesinlik elde edilmiştir.

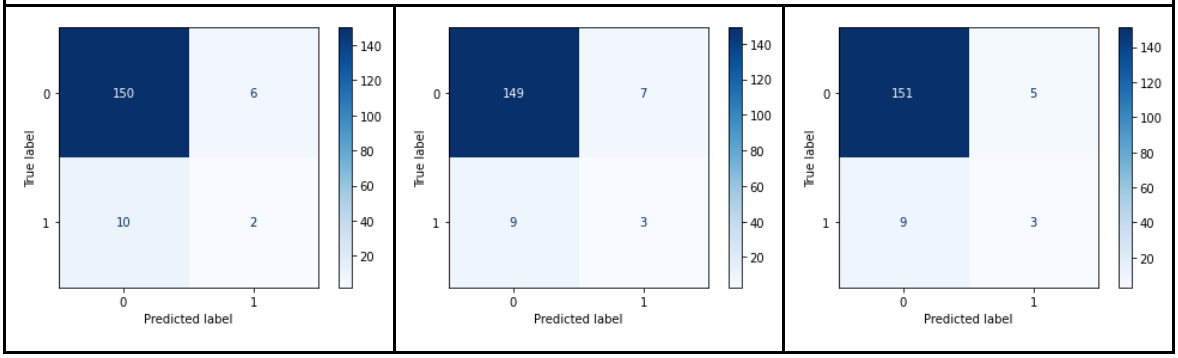
TEMEL öznitelikleri ile 156 yaşayan hastanın 150'si, doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 2'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 6'sı hatalı bir şekilde ölen olarak, ölen hastaların ise 10'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 156 yaşayan hastanın 149'u doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 3'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 7'si hatalı bir şekilde ölen olarak, ölen hastaların ise 9'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucu geçilmiştir.

EN İYİ öznitelikleri (200) ile 156 yaşayan hastanın 151'i doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 3'ü doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 5'i hatalı bir şekilde ölen olarak, ölen hastaların ise 9'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada mide kanseri için belirlenmiş toplam özniteliklerin %4'ü kullanılarak TEMEL sonucu geçilmiştir.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.20. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.94	0.96	0.95
1	0.25	0.17	0.20
accuracy			0.90
macro avg	0.59	0.56	0.57
weighted avg	0.89	0.90	0.90



Şekil 4.20. Mide kanseri Destek Vektör Makinesi için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.95, ölen için 0.27 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.61 ve Doğruluk sonucu 0.90 olarak bulunmuştur. Makro Ort. özelinde %60 Duyarlılığında %62 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (200) sınıflandırmaların F1 skoru yaşayan için 0.96, ölen için 0.30 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.63 ve Doğruluk sonucu 0.92 olarak bulunmuştur. Makro Ort. özelinde %61 Duyarlılığında %66 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.855, İLK 100 için: 0.683 ve EN İYİ için: 0.814 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.19. üzerinde paylaşılmıştır.

Tablo 4.19. Mide kanseri Destek Vektör Makinesi için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.95**	**0.20**	**0.90**	**0.57**	**0.90**	**0.855**	**0.59**	**0.56**
100	0.95	0.27	0.90	0.61	0.90	0.683	0.62	0.60
200	0.96	0.30	0.92	0.63	0.91	0.814	0.66	0.61
300	0.94	0.25	0.89	0.60	0.89	0.826	0.60	0.60
400	0.95	0.29	0.91	0.62	0.90	0.849	0.64	0.61
500	0.95	0.21	0.91	0.58	0.90	0.810	0.61	0.57
600	0.96	0.30	0.92	0.63	0.91	0.853	0.66	0.61
700	0.95	0.21	0.91	0.58	0.90	0.840	0.61	0.57
800	0.95	0.20	0.90	0.57	0.90	0.826	0.59	0.56
900	0.95	0.19	0.90	0.57	0.89	0.847	0.58	0.56
1000	0.96	0.24	0.92	0.60	0.91	0.852	0.67	0.57
1100	0.95	0.21	0.91	0.58	0.90	0.869	0.61	0.57
1200	0.95	0.19	0.90	0.57	0.89	0.865	0.58	0.56
1300	0.95	0.19	0.90	0.57	0.89	0.850	0.58	0.56
1400	0.95	0.19	0.90	0.57	0.89	0.853	0.58	0.56
1500	0.95	0.21	0.91	0.58	0.90	0.856	0.61	0.57

EN İYİ öznitelik grubunun (200) dağılımı şu şekilde olduğu bulunmuştur; tanı: 147, ilaç: 22, prosedür: 31. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu prosedür ve ilaç öznitelikleri takip etmiştir.

4.2.4.5. Çok Katmanlı Algılayıcı sonuçları

Mide kanseri kohortundaki hastalar yaşayan ve ölen olarak işaretlenmiş ve Çok Katmanlı Algılayıcı modeli ile yapılmış sınıflandırma performansları elde edilmiştir. Sonuçlar TEMEL, İLK 100 ve EN İYİ öznitelik grupları için değerlendirilmiş ve 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları paylaşılmıştır.

168 test kümesinden hasta ile TEMEL öznitelik için sınıflandırma raporu Şekil 4.21. de gösterilmiştir. Burada 156 yaşayan hasta için F1 skoru 0.96 ve 12 ölen hasta için F1 skoru

0.42 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.69 ve Doğruluk sonucu 0.92 olarak bulunmuştur. Makro Ort. özelinde %69 Duyarlılığında %69 Kesinlik elde edilmiştir.

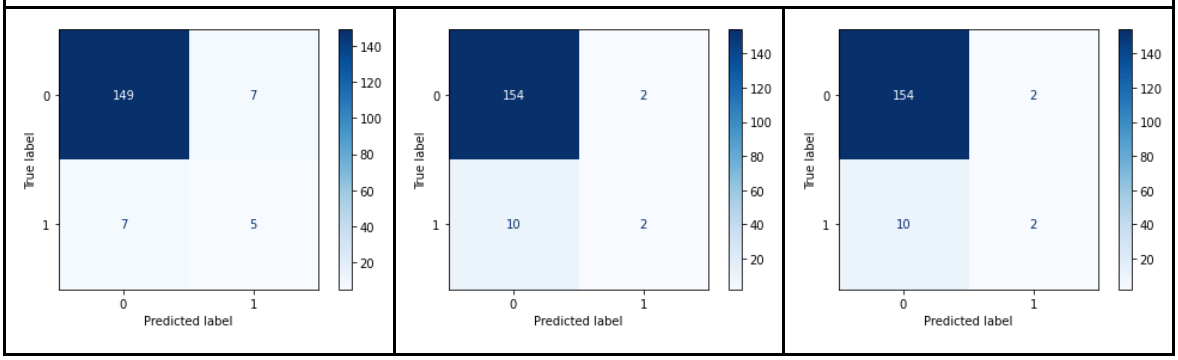
TEMEL öznitelikleri ile 156 yaşayan hastanın 149'u, doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 5'i doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 7's, hatalı bir şekilde ölen olarak, ölen hastaların ise 7'si hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Temel nokta olarak belirlenen bu sonuçları geçen her sonuç başarılı olarak nitelendirilmiştir.

İLK 100 öznitelikleri ile 156 yaşayan hastanın 154'ü doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 2'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 2'si hatalı bir şekilde ölen olarak, ölen hastaların ise 10'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada sadece 100 öznitelik ile TEMEL sonucuna yaklaşılamamıştır.

EN İYİ öznitelikleri (100) ile 156 yaşayan hastanın 154'ü doğru bir şekilde yaşayan olarak, 12 ölen hastanın ise 2'si doğru bir şekilde ölen olarak sınıflandırılmıştır. Yaşayan hastaların 2'si hatalı bir şekilde ölen olarak, ölen hastaların ise 10'u hatalı bir şekilde yaşayan olarak sınıflandırılmıştır. Burada mide kanseri için belirlenmiş toplam özniteliklerin %2'si kullanılarak TEMEL sonucuna yaklaşılamamıştır.

Bu test kümesi için denenmiş TEMEL, İLK 100 ve EN İYİ öznitelik gruplarının karışıklık matris sonucu Şekil 4.21. üzerinde görselleştirilmiştir.

	precision	recall	f1-score
0	0.96	0.96	0.96
1	0.42	0.42	0.42
accuracy			0.92
macro avg	0.69	0.69	0.69
weighted avg	0.92	0.92	0.92



Şekil 4.21. Mide kanseri Çok Katmanlı Algılayıcı için üstte TEMEL sonucun sınıflandırma raporu, altta soldan sağa TEMEL, İLK 100 ve EN İYİ sonuçlarının karışıklık matrisleri

İLK 100 öznitelik ile sınıflandırmaların F1 skoru yaşayan için 0.96, ölen için 0.25 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.61 ve Doğruluk sonucu 0.93 olarak bulunmuştur. Makro Ort. özelinde %58 Duyarlılığında %72 Kesinlik elde edilmiştir.

EN İYİ öznitelik ile (100) sınıflandırmaların F1 skoru yaşayan için 0.96, ölen için 0.25 çıkmıştır. F1 skoru özelinde Makro Ort. sonucu 0.61 ve Doğruluk sonucu 0.93 olarak bulunmuştur. Makro Ort. özelinde %58 Duyarlılığında %72 Kesinlik elde edilmiştir.

Sınıfların ne derece iyi ayrılabildiği TEMEL için: 0.692, İLK 100 için: 0.505 ve EN İYİ için: 0.505 AUC-ROC sonuçları ile gösterilmiştir. 100 den 1500 e kadar olan öznitelik gruplarının tam sonuçları Tablo 4.20. üzerinde paylaşılmıştır.

Tablo 4.20. Mide kanseri Çok Katmanlı Algılayıcı için farklı öznitelik gruplarının sınıflandırma sonucu

	F1-Skoru						Makro Ort.	
Öznitelik Sayısı	Yaşayan(0)	Ölen(1)	Doğruluk	Makro Ort.	Ağırlıklı Ort.	AUC-ROC	Kesinlik	Duyarlılık
TEMEL	**0.96**	**0.42**	**0.92**	**0.69**	**0.92**	**0.692**	**0.69**	**0.69**
100	0.96	0.25	0.93	0.61	0.91	0.505	0.72	0.58
200	0.94	0.23	0.88	0.58	0.89	0.557	0.58	0.59
300	0.95	0.00	0.90	0.47	0.88	0.557	0.46	0.48
400	0.96	0.00	0.93	0.48	0.89	0.353	0.46	0.50
500	0.96	0.00	0.93	0.48	0.89	0.506	0.46	0.50
600	0.96	0.00	0.93	0.48	0.89	0.500	0.46	0.50
700	0.96	0.00	0.93	0.48	0.89	0.432	0.46	0.50
800	0.96	0.00	0.93	0.48	0.89	0.400	0.46	0.50
900	0.96	0.00	0.92	0.48	0.89	0.595	0.46	0.50
1000	0.96	0.00	0.93	0.48	0.89	0.423	0.46	0.50
1100	0.95	0.11	0.90	0.53	0.89	0.426	0.54	0.52
1200	0.96	0.00	0.92	0.48	0.89	0.418	0.46	0.50
1300	0.96	0.00	0.92	0.48	0.89	0.392	0.46	0.49
1400	0.96	0.00	0.93	0.48	0.89	0.433	0.46	0.50
1500	0.96	0.00	0.93	0.48	0.89	0.382	0.46	0.50

EN İYİ öznitelik grubunun (100) dağılımı şu şekilde olduğu bulunmuştur; tanı: 77, ilaç: 8, prosedür: 15. Belirtilen kanser grubu için en çok tanı özniteliği ilişkili olduğu görülmüş ve bunu prosedür ve ilaç öznitelikleri takip etmiştir.

4.3. Genel Sonuçlar

Genel sonuçlar F1 Makro Ortalama ve AUC-ROC karşılaştırması olarak ikiye ayrılmıştır. Aşağıdaki tablolar için satırlar, ilgili kanser türüne göre gruplandırılmış olarak kullanılan modelleri göstermektedir. Sütunlar olarak üç tür sonuç kümesi vardır. "TEMEL (TÜM) Öznitelikler", o model için tüm öznitelikler ile birlikte temel performans sonuçlarını gösterir. “EN İYİ X Öznitelik”, temel sonuca kıyasla en iyi performans sonuçlarını gösterir. İlgili sonucu veren kullanılmış öznitelik miktarı X değeri de gösterilmiştir. Son olarak, “İLK 100 Öznitelik”, olası minimum öznitelik kümesi olan ilk 100 özniteliğin sonucunu gösterir. Tüm renkli hücreler, diğer modellerle karşılaştırıldığında, o öznitelik grubunda belirtilen kanser türü için en iyi yerel sonuç olduklarını gösterir.

EN İYİ X Öznitelik için dağılımın aşağıdaki şekilde olduğu bulunmuştur:

- Meme kanserinde 1900 öznitelik arasından; tanı: 1238, ilaç: 411, prosedür: 251
- Akciğer kanserinde 1000 öznitelik arasından; tanı: 715, ilaç: 153, prosedür: 132
- Prostat kanserinde 700 öznitelik arasından; tanı: 487, ilaç: 119, prosedür: 94
- Mide kanserinde 600 öznitelik arasından; tanı: 380, ilaç: 115, prosedür: 105

İLK 100 Öznitelik için dağılımın aşağıdaki şekilde olduğu bulunmuştur:

- Meme kanserinde; tanı: 73, ilaç: 15, prosedür: 12
- Akciğer kanserinde; tanı: 86, ilaç: 2, prosedür: 12
- Prostat kanserinde; tanı: 82, ilaç: 4, prosedür: 14
- Mide kanserinde; tanı: 77, ilaç: 8, prosedür: 15

İLK 100 Öznitelik listesinin detayları EK 1 de bulunabilir.

4.3.1. Genel Karşılaştırma (Makro F1)

Tablo 4.21. de görüldüğü gibi TEMEL öznitelikleri kullanılarak meme, akciğer ve prostat kanserinde en iyi sonuçları Karar Ağacı yöntemi vermiştir. Mide ve akciğer kanseri için Lojistik Regresyon en iyi sonuçları vermiştir. Tüm kanser türleri arasından prostat kanseri 0.82 ile en yüksek skoru vermiştir.

Tüm kanser türleri için İLK 100 sonuçları genel olarak tüm yöntemler için TEMEL sonuçlarına yakın çıkmıştır. İLK 100 özelinde Rastgele Orman ve Çok Katmanlı Algılayıcı yöntemlerinin sonuçları, meme ve akciğer kanseri için TEMEL sonuçlarını geçmiştir. Aynı şekilde Rastgele Orman ve Destek Vektör Makinesi yöntemlerinin sonuçları, prostat ve mide kanseri için TEMEL sonuçlarını geçmiştir. İLK 100 özniteliklerinin sonuçlarının TEMEL sonuçlarına yaklaşık %90 oranında yakın olduğu görülmüştür (Tablo 4.21.).

Tablo 4.21. TEMEL ve İLK 100 öznitelik için F1 Makro skorları

		TEMEL (TÜM) Öznitelikler	İLK 100 Öznitelik
Kanser	**Model**	**F1 Makro Ort.**	**F1 Makro Ort.**
Meme	Lojistik Regresyon	0.70	0.67
	Karar Ağacı	0.76	0.63
	Rastgele Orman	0.53	0.61
	Destek Vektör Makinesi	0.63	0.62
	Çok Katmanlı Algılayıcı	0.60	0.62
Akciğer	Lojistik Regresyon	0.71	0.68
	Karar Ağacı	0.71	0.67
	Rastgele Orman	0.49	0.63
	Destek Vektör Makinesi	0.67	0.66
	Çok Katmanlı Algılayıcı	0.62	0.64
Prostat	Lojistik Regresyon	0.76	0.74
	Karar Ağacı	0.82	0.76
	Rastgele Orman	0.53	0.67
	Destek Vektör Makinesi	0.70	0.75
	Çok Katmanlı Algılayıcı	0.68	0.65
Mide	Lojistik Regresyon	0.72	0.63
	Karar Ağacı	0.64	0.62
	Rastgele Orman	0.48	0.60
	Destek Vektör Makinesi	0.57	0.61
	Çok Katmanlı Algılayıcı	0.69	0.61

EN İYİ X öznitelik, hesaplama yükü ve bellek kaynaklarının sınırlı durumlarda kullanılabilir. EN İYİ X öznitelik kullanılarak mortalite tahmin sonuçları Tablo 4.22. de görülebilir.

- Meme kanseri için; Karar Ağacı modeli ile tüm özniteliklerin yaklaşık %20'si kullanılarak alınan skor TEMEL sonucuna yakın çıkmıştır. Lojistik Regresyon ile tüm özniteliklerin %18'i kullanılarak TEMEL sonuç ile aynı skor elde edilmiştir.

- Akciğer kanseri için; Lojistik Regresyon ile tüm özniteliklerin %10'u kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır. Karar Ağacı ile tüm özniteliklerin yaklaşık %12'si kullanılarak alınan skor TEMEL sonucuna yakın çıkmıştır.
- Prostat kanseri için; Lojistik Regresyon ile tüm özniteliklerin %8'i kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır. Karar Ağacı ve Çok Katmanlı Algılayıcı ile sırasıyla tüm özniteliklerin %14'ü ve %1'i kullanılarak alınan skorlar TEMEL sonucuna yakın çıkmıştır.
- Mide kanseri için; Lojistik Regresyon ile tüm özniteliklerin %12'si kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır. Karar Ağacı ve Çok Katmanlı Algılayıcı ile sırasıyla tüm özniteliklerin %14'ü ve %2'si kullanılarak alınan skorlar TEMEL sonucuna yakın çıkmıştır.

Tablo 4.22. EN İYİ X öznitelik için F1 Makro skorları

		EN İYİ X Öznitelik	
Kanser	**Model**	**X Değeri**	**F1 Makro Ort.**
Meme	Lojistik Regresyon	1500	0.70
	Karar Ağacı	1900	0.74
	Rastgele Orman	100	0.61
	Destek Vektör Makinesi	400	0.72
	Çok Katmanlı Algılayıcı	100	0.62
Akciğer	Lojistik Regresyon	1000	0.73
	Karar Ağacı	1400	0.70
	Rastgele Orman	200	0.66
	Destek Vektör Makinesi	600	0.72
	Çok Katmanlı Algılayıcı	1100	0.68
Prostat	Lojistik Regresyon	700	0.82
	Karar Ağacı	1000	0.79
	Rastgele Orman	100	0.67
	Destek Vektör Makinesi	100	0.75
	Çok Katmanlı Algılayıcı	100	0.65
Mide	Lojistik Regresyon	600	0.79
	Karar Ağacı	700	0.62
	Rastgele Orman	100	0.60
	Destek Vektör Makinesi	200	0.63
	Çok Katmanlı Algılayıcı	100	0.61

4.3.2. Genel Karşılaştırma (AUC-ROC)

Tüm kanser türleri için İLK 100 sonuçları genel olarak tüm yöntemler için TEMEL sonuçlarına yakın çıkmıştır. İLK 100 özelinde Rastgele Orman ve Çok Katmanlı Algılayıcı yöntemlerinin sonuçları, meme ve akciğer kanseri için TEMEL sonuçlarını geçmiştir. Aynı

şekilde Rastgele Orman ve Destek Vektör Makinesi yöntemlerinin sonuçları, prostat ve mide kanseri için TEMEL sonuçlarını geçmiştir. İLK 100 özniteliklerinin sonuçlarının TEMEL sonuçlarına yaklaşık %90 oranında yakın olduğu görülmüştür (Tablo 4.23.).

Tablo 4.23. de görüldüğü gibi TEMEL öznitelikleri kullanılarak meme ve mide kanserinde en iyi sonuçları Destek Vektör Makinesi yöntemi vermiştir. Akciğer ve prostat kanseri için ise Rastgele Orman en iyi sonuçları vermiştir. Tüm kanser türleri arasından prostat kanseri 0.94 ile en yüksek skoru vermiştir.

Tüm kanser türleri için İLK 100 sonuçları genel olarak tüm yöntemler için TEMEL sonuçlarına oldukça yakın çıkmıştır. İLK 100 özelinde Lojistik Regresyon yönteminin sonucu, meme kanseri için TEMEL sonucunu geçmiştir. Aynı şekilde Lojistik Regresyon ve Çok Katmanlı Perceptron yöntemlerinin sonuçları, akciğer ve prostat kanseri için TEMEL sonuçlarını geçmiştir. İLK 100 özniteliklerinin sonuçlarının TEMEL sonuçlarına yaklaşık %98 oranında yakın olduğu görülmüştür (Tablo 4.23.).

Tablo 4.23. TEMEL ve İLK 100 öznitelik için AUC-ROC skorları

		TEMEL (TÜM) Öznitelikler	İLK 100 Öznitelik
Kanser	**Model**	**AUC-ROC**	**AUC-ROC**
Meme	Lojistik Regresyon	0.75	0.85
	Karar Ağacı	0.77	0.65
	Rastgele Orman	0.89	0.90
	Destek Vektör Makinesi	0.92	0.79
	Çok Katmanlı Algılayıcı	0.78	0.73
Akciğer	Lojistik Regresyon	0.77	0.86
	Karar Ağacı	0.70	0.67
	Rastgele Orman	0.88	0.88
	Destek Vektör Makinesi	0.82	0.78
	Çok Katmanlı Algılayıcı	0.68	0.72
Prostat	Lojistik Regresyon	0.74	0.93
	Karar Ağacı	0.79	0.77
	Rastgele Orman	0.94	0.93
	Destek Vektör Makinesi	0.84	0.78
	Çok Katmanlı Algılayıcı	0.58	0.77
Mide	Lojistik Regresyon	0.80	0.65
	Karar Ağacı	0.73	0.60
	Rastgele Orman	0.84	0.84
	Destek Vektör Makinesi	0.85	0.68
	Çok Katmanlı Algılayıcı	0.69	0.50

EN İYİ X öznitelikleri kullanılarak mortalite tahmininde AUC-ROC sonuçları Tablo 4.24. de görülebilir.

- Meme kanseri için; Rastgele Orman modeli ile tüm özniteliklerin yaklaşık %8'i kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır. Çok Katmanlı Algılayıcı ile tüm özniteliklerin yaklaşık %5'i kullanılarak alınan skor TEMEL sonucuna yakın çıkmıştır. Karar Ağacı ile tüm özniteliklerin %23'ü kullanılarak

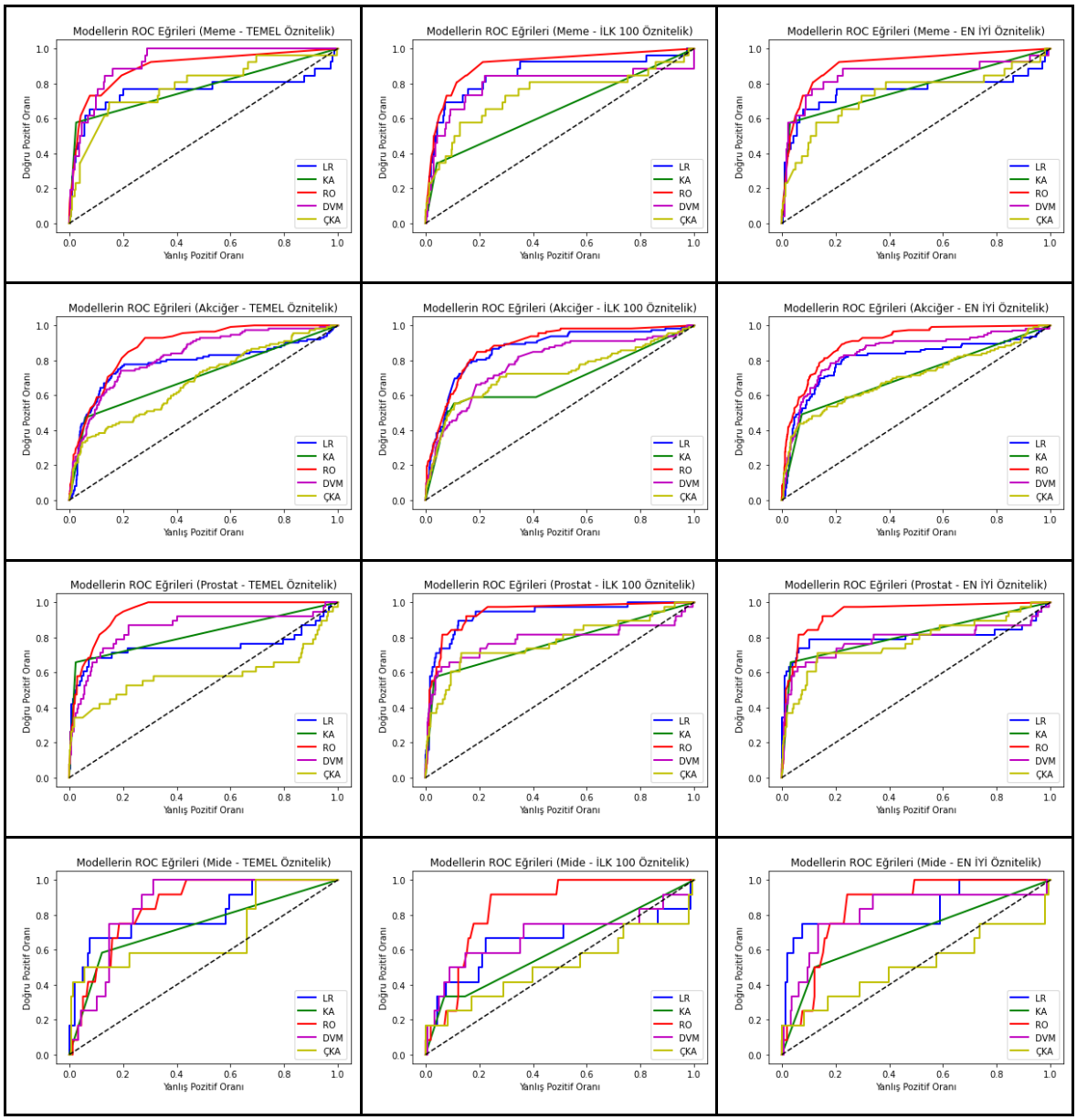
alınan skor TEMEL ile aynı elde edilmiştir. Lojistik Regresyon ve Destek Vektör Makinesi ile sırasıyla tüm özniteliklerin %1'i ve %18'i kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır.

- Akciğer kanseri için; Rastgele Orman ile tüm özniteliklerin %5'i kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır. Karar Ağacı ile tüm özniteliklerin %14'ü kullanılarak alınan skor TEMEL ile aynı elde edilmiştir. Lojistik Regresyon, Destek Vektör Makinesi ve Çok Katmanlı Algılayıcı ile sırasıyla tüm özniteliklerin %1, %6 ve %2'si kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır.
- Prostat kanseri için; Rastgele Orman ile tüm özniteliklerin yaklaşık %3'ü kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır. Lojistik Regresyon, Karar Ağacı, Destek Vektör Makinesi ve Çok Katmanlı Algılayıcı ile sırasıyla tüm özniteliklerin %1, %12, %4 ve %1'i kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır.
- Mide kanseri için; Rastgele Orman ile tüm özniteliklerin %6'sı kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır. Karar Ağacı ve Çok Katmanlı Algılayıcı ile sırasıyla tüm özniteliklerin %14'ü ve %18'i kullanılarak alınan skor TEMEL sonucuna yakın çıkmıştır. Lojistik Regresyon ve Destek Vektör Makinesi ile sırasıyla tüm özniteliklerin %12'si ve %22'si kullanılarak TEMEL sonuçtan daha iyi performans alınmıştır.

Tablo 4.24. EN İYİ X öznitelik için AUC-ROC skorları

		EN İYİ X Öznitelik	
Kanser	**Model**	**X Değeri**	**AUC-ROC**
Meme	Lojistik Regresyon	100	0.85
	Karar Ağacı	1900	0.77
	Rastgele Orman	700	0.94
	Destek Vektör Makinesi	1500	0.94
	Çok Katmanlı Algılayıcı	400	0.76
Akciğer	Lojistik Regresyon	100	0.86
	Karar Ağacı	1400	0.70
	Rastgele Orman	500	0.91
	Destek Vektör Makinesi	600	0.84
	Çok Katmanlı Algılayıcı	200	0.72
Prostat	Lojistik Regresyon	100	0.93
	Karar Ağacı	1000	0.81
	Rastgele Orman	300	0.96
	Destek Vektör Makinesi	400	0.93
	Çok Katmanlı Algılayıcı	100	0.77
Mide	Lojistik Regresyon	600	0.83
	Karar Ağacı	700	0.68
	Rastgele Orman	300	0.88
	Destek Vektör Makinesi	1100	0.86
	Çok Katmanlı Algılayıcı	900	0.59

AUC-ROC puanlarına ek olarak, mortalite tahmininde TEMEL, İLK 100 ve EN İYİ özniteliklerinin nasıl davrandığını göstermek için ROC eğrileri çizilmiş ve Şekil 4.22. de gösterilmiştir.



Şekil 4.22. Modellerin ROC eğrileri

5. SONUÇ VE TARTIŞMA

Kanser, dünyanın en tehlikeli hastalığıdır. Hastaların mortalite oranını azaltmak için kanserin erken evrede tespiti çok önemlidir. Makine öğrenimi yaklaşımları manuel karar vermekten daha hızlı olsa da, özniteliklerin sayısı arttıkça hesaplama süresi ve modelin ihtiyaç duyduğu kaynaklar da genişler. Burada ele alınan temel sorun; doktorlara yardımcı olmak adına çeşitli kanser hastalarında teşhis sonrası hastane içi mortalite tahmini için, tahmin oranını mümkün olan en yüksekte tutacak gerekli en az öznitelik kümesini bulmaktır. Güncel bir veri kümesi olan MIMIC-IV üzerinde gerçekleştirilen bu çalışma ile anonimleştirilmiş gerçek hasta verileri ile çalışılmıştır. Ayrıca MIMIC-IV veri kümesinde bu amaca yönelik başka bir çalışmaya rastlanmamıştır.

Çeşitli makine öğrenmesi yöntemleri ve hastaların tanı, ilaç ve prosedür öznitelikleri kullanılarak bu yöntemlerin karşılaştırmalı analizi yapılmıştır. Makine öğrenme yöntemleri; Lojistik Regresyon, Karar Ağacı, Rastgele Orman, Destek Vektör Makinesi ve Çok Katmanlı Algılayıcıdır. Öznitelik çıkarımında, model eğitimine harcanacak hesaplama maliyetini ve zamanını azaltmak için hasta verileri tek bitlik gösterim ile temsil edilmiştir. Lojistik Regresyon ile en önemli öznitelikler belirlenmiş ve seçilmiştir. Bu öznitelikler, tamamı kullanıldığında alınan sonuçlara benzer veya tercihen daha iyi sonuçları almak amacıyla modellere 100'ün katları ile verilmiştir.

Sınırlı öznitelik kümeleri ile alınan sonuçlar, karşılaştırma amaçlı alınan başlangıç skorlarını geçmiştir. F1 Makro Ortalama skorları; Meme için 0.74, akciğer için 0.73, prostat için 0.82, mide için 0.79 olarak bulunmuştur. AUC-ROC skorları; Meme için 0.94, akciğer için 0.91, prostat için 0.96, mide için 0.88 olarak bulunmuştur.

Bu öznitelik kümesinin, prostat kanseri için diğer kanser türlerine göre daha uygun olduğu ve daha çok genelleme olanağına sahip olduğu görülmektedir. Başarım sırasında bunu; mide, meme ve akciğer kanseri izlemiştir.

Bu öznitelik kümesi ve bu kanser türleri için mortalite sınıflandırmasında diğerlerine kıyasla en başarılı model Lojistik Regresyon olarak bulunmuştur. Bunu Karar Ağacı, Destek Vektör Makinesi ve Çok Katmanlı Algılayıcı izlemiştir. Rastgele Orman sınıflandırıcısı aralarında en düşük skorları vermiştir.

Önerilen yaklaşımın çeşitli katkıları vardır. İlk olarak, bu yaklaşımın daha az öznitelik kullanarak mevcut tüm öznitelikleri kullanmakla aynı tahmin performansını koruduğu gösterilmiştir. İkinci olarak, öznitelik kümeleri ikili gösterim değerlerine dönüştürülerek,

öznitelik vektörlerinin bellek ayak izi küçültülmüştür. Üçüncüsü, bu tez çalışması MIMIC-IV veri kümesinde kanser mortalitesini araştıran ilk çalışmadır. Son olarak, bu çalışmada mortaliteyi tahmin etmede hangi makine öğrenimi modellerinin hangi kanser türüyle birlikte iyi çalıştığı belirlenmiştir. EN İYİ X Öznitelik gruplarına göre ortalama %67 oranında tanı özniteliği mortalite tahmininde daha baskın olduğu ortaya konulmuştur. Bu ilişkiyi sırasıyla yaklaşık %18 ve %15 oranlarıyla ilaç ve prosedür öznitelikleri takip etmiştir. İLK 100 Öznitelik gruplarına göre ise, ortalama %80 oranında tanı özniteliği mortalite tahmininde daha baskın olduğu ortaya konulmuştur. Bu ilişkiyi sırasıyla yaklaşık %13 ve %7 oranlarıyla prosedür ve ilaç öznitelikleri takip etmiştir.

Kullanılan kanser verileri dengesiz olduğundan, modeller sınıf dağılımındaki örnek miktarlarına ve muhtemelen yanlış etiketlenmiş gürültü verilerine karşı oldukça hassas sonuçlar vermiştir. Buna rağmen sonuçlardan da anlaşılacağı gibi, buradaki yaklaşım sınırlı miktardaki veri ile ilişkili öznitelik kümelerini ve verimli makine öğrenme algoritmalarını kullanarak başarılı sonuçlar almıştır. Gelecekteki çalışmalarda; yaş, cinsiyet ve laboratuvar değerleri gibi geleneksel öznitelikler eklenebilir ve aynı prosedürler tekrar takip edilebilir. Ayrıca, diğer benzer açık erişim veri kümeleri ile gerçek veriler de, buradaki yaklaşımı doğrulamak veya karşılaştırmak için kullanılabilir. Son olarak, daha fazla hasta verisi oluştuğunda, derin öğrenme yaklaşımları [59] gelecek çalışmalar için düşünülebilir.

KAYNAKLAR

[1] “Cancer,” World Health Organization. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer. [Accessed: 13-May-2022].

[2] “Worldwide cancer data: World cancer research fund international,” WCRF International, 14-Apr-2022. [Online]. Available: https://www.wcrf.org/cancer-trends/world wide-cancer-data/. [Accessed: 13-May-2022].

[3] “Breast cancer statistics: World cancer research fund international,” WCRF International, 14-Apr-2022. [Online]. Available: https://www.wcrf.org/cancer-trends/breast -cancer-statistics/. [Accessed: 13-May-2022].

[4] “Lung cancer statistics: World cancer research fund international,” WCRF International, 14-Apr-2022. [Online]. Available: https://www.wcrf.org/cancer-trends/lung-cancer-statistics /. [Accessed: 13-May-2022].

[5] “Prostate cancer statistics: World cancer research fund international,” WCRF International, 14-Apr-2022. [Online]. Available: https://www.wcrf.org/cancer-trends/ prostate-cancer-statistics/. [Accessed: 13-May-2022].

[6] “Stomach cancer statistics,” WCRF International, 14-Apr-2022. [Online]. Available: https://www.wcrf.org/cancer-trends/stomach-cancer-statistics/. [Accessed: 13-May-2022].

[7] Y. Xie, W.-Y. Meng, R.-Z. Li, Y.-W. Wang, X. Qian, C. Chan, Z.-F. Yu, X.-X. Fan, H.-D. Pan, C. Xie, Q.-B. Wu, P.-Y. Yan, L. Liu, Y.-J. Tang, X.-J. Yao, M.-F. Wang, and E. L.-H. Leung, “Early lung cancer diagnostic biomarker discovery by machine learning methods,” Translational Oncology, vol. 14, no. 1, p. 100907, 2021.

[8] S. S. Raoof, M. A. Jabbar, and S. A. Fathima, “Lung cancer prediction using machine learning: A comprehensive approach,” 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020.

[9] E. CENGIL and A. CINAR, “A deep learning based approach to lung cancer identification,” 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018.

[10] J. Alam, S. Alam, and A. Hossan, “Multi-stage lung cancer detection and prediction using multi-class SVM classifier,” 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018.

[11] A. Iyer, H. Vyshnavi A M, and K. Namboori P K, “Deep convolution network based prediction model for medical diagnosis of lung cancer - a deep pharmacogenomic approach : Deep diagnosis for Lung Cancer,” 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC), 2018.

[12] T. Patel and V. Nayak, “Hybrid approach for feature extraction of Lung Cancer Detection,” 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018.

[13] Q. Wu and W. Zhao, “Small-cell lung cancer detection using a supervised machine learning algorithm,” 2017 International Symposium on Computer Science and Intelligent Controls (ISCSIC), 2017.

[14] A. Dekker, C. Dehing-Oberije, D. D. Ruysscher, P. Lambin, K. Komati, G. Fung, S. Yu, A. Hope, W. D. Neve, and Y. Lievens, “Survival prediction in lung cancer treated with radiotherapy: Bayesian networks vs. support vector machines in handling missing data,” 2009 International Conference on Machine Learning and Applications, 2009.

[15] M. Shalini and S. Radhika, “Machine learning techniques for prediction from various breast cancer datasets,” 2020 Sixth International Conference on Bio Signals, Images, and Instrumentation (ICBSII), 2020.

[16] T. Thomas, N. Pradhan, and V. S. Dhaka, “Comparative analysis to predict breast cancer using machine learning algorithms: A survey,” 2020 International Conference on Inventive Computation Technologies (ICICT), 2020.

[17] M. I. Showrov, M. T. Islam, M. D. Hossain, and M. S. Ahmed, “Performance comparison of three classifiers for the classification of Breast Cancer Dataset,” 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 2019.

[18] Naveen, R. K. Sharma, and A. Ramachandran Nair, “Efficient Breast Cancer Prediction Using Ensemble Machine Learning Models,” 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 2019.

[19] V. Mishra, Y. Singh, and S. Kumar Rath, “Breast cancer detection from thermograms using feature extraction and Machine Learning Techniques,” 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019.

[20] E. A. Bayrak, P. Kirci, and T. Ensari, “Comparison of machine learning methods for breast cancer diagnosis,” 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), 2019.

[21] A. Bharat, N. Pooja, and R. A. Reddy, “Using machine learning algorithms for breast cancer risk prediction and diagnosis,” 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), 2018.

[22] N. Khuriwal and N. Mishra, “Breast cancer diagnosis using Adaptive Voting Ensemble Machine Learning Algorithm,” 2018 IEEMA Engineer Infinite Conference (eTechNxT), 2018.

[23] N. Kolay and P. Erdogmus, “The classification of breast cancer with Machine Learning Techniques,” 2016 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 2016.

[24] B. M. Gayathri and C. P. Sumathi, “Feature selection using linear discriminant analysis for Breast Cancer Dataset,” 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018.

[25] K. Revett, S. T. de Magalhaes, and H. M. Santos, “Data mining a prostate cancer dataset using rough sets,” 2006 3rd International IEEE Conference Intelligent Systems, 2006.

[26] V. Danilatou, D. Antonakaki, C. Tzagkarakis, A. Kanterakis, V. Katos, and T. Kostoulas, “Automated mortality prediction in critically-ill patients with thrombosis using machine learning,” 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), 2020.

[27] S. Afrose, W. Song, C. B. Nemeroff, C. Lu, and D. (D. Yao, “Subpopulation-specific machine learning prognosis for underrepresented patients with double prioritized bias correction,” 2021.

[28] G. H. Lee and S.-Y. Shin, “Federated learning on clinical benchmark data: Performance assessment,” Journal of Medical Internet Research, vol. 22, no. 10, 2020.

[29] I. Hammoud, P. Prasanna, I. V. Ramakrishnan, A. Singer, M. Henry, and H. Thode, “EventScore: An automated real-time early warning score for clinical events,” arXiv.org, 14-Feb-2021. [Online]. Available: https://arxiv.org/abs/2102.05958. [Accessed: 06-May-2022].

[30] C. M. Sauer, J. Dong, L. A. Celi, and D. Ramazzotti, “Improved survival of cancer patients admitted to the Intensive Care Unit between 2002 and 2011 at a U.S. teaching hospital,” Cancer Research and Treatment, vol. 51, no. 3, pp. 973–981, 2019.

[31] A. A. Magna, H. Allende-Cid, C. Taramasco, C. Becerra, and R. L. Figueroa, “Application of machine learning and word embeddings in the classification of cancer diagnosis using patient anamnesis,” IEEE Access, vol. 8, pp. 106198–106213, 2020.

[32] H. Wang, Y. Li, S. A. Khan, and Y. Luo, “Prediction of breast cancer distant recurrence using natural language processing and knowledge-guided convolutional neural network,” Artificial Intelligence in Medicine, vol. 110, p. 101977, 2020.

[33] Z. Zeng, L. Yao, A. Roy, X. Li, S. Espino, S. E. Clare, S. A. Khan, and Y. Luo, "Identifying breast cancer distant recurrences from electronic health records using Machine Learning," Journal of Healthcare Informatics Research, vol. 3, no. 3, pp. 283–299, 2019.

[34] Z. Nowroozilarki, A. Pakbin, J. Royalty, D. K. K. Lee, and B. J. Mortazavi, "Real-time mortality prediction using MIMIC-IV ICU data via boosted Nonparametric Hazards," *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2021.

[35] C. Meng, L. Trinh, N. Xu, and Y. Liu, "Mimic-if: Interpretability and fairness evaluation of deep learning models on Mimic-IV Dataset," 2021.

[36] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L. A. Celi, and R. Mark, "Mimic-IV," MIMIC-IV v1.0, 16-Mar-2021. [Online]. Available: https://physionet.org/content/mimiciv/1.0/. [Accessed: 13-Apr-2022].

[37] G. Miao, Z. Li, L. Chen, W. Li, G. Lan, Q. Chen, Z. Luo, R. Liu, and X. Zhao, "A novel nomogram for predicting morbidity risk in patients with secondary malignant neoplasm of bone and bone marrow: An analysis based on the large mimic-III clinical database," International Journal of General Medicine, vol. Volume 15, pp. 3255–3264, 2022.

[38] M. E. O'Rourke, "Decision making and prostate cancer treatment selection: A Review," Seminars in Oncology Nursing, vol. 17, no. 2, pp. 108–117, 2001.

[39] K. E. Osann, "Lung Cancer in Women: The Importance of Smoking, Family History of Cancer, and Medical History of Respiratory Disease," CANCER RESEARCH, vol. 51, no. 18, pp. 4893–4897, 1991.

[40] J. F. Piccirillo, "Importance of comorbidity in head and neck cancer," The Laryngoscope, vol. 110, no. 4, pp. 593–602, 2000.

[41] R. Rafique, S. M. R. Islam, and J. U. Kazi, "Machine learning in the prediction of cancer therapy," Computational and Structural Biotechnology Journal, vol. 19, pp. 4003–4017, 2021.

[42] B. L. Brady, M. Lucci, K. Wilson, K. M. Fox, J. Wojtynek, C. Cooper, H. Varker, C. L. Chebili, and I. Dokubo, “Chemotherapy-induced peripheral neuropathy in metastatic breast cancer patients initiating intravenous paclitaxel/NAB-paclitaxel,” The American Journal of Managed Care, vol. 27, no. 1, pp. 37–43, Dec. 2020.

[43] C. Lin, R. Clark, P. Tu, H. B. Bosworth, and L. L. Zullig, “Breast cancer oral anti-cancer medication adherence: A systematic review of psychosocial motivators and barriers,” Breast Cancer Research and Treatment, vol. 165, no. 2, pp. 247–260, 2017.

[44] X. Deng and Y. Nakamura, “Cancer precision medicine: From cancer screening to drug selection and personalized immunotherapy,” Trends in Pharmacological Sciences, vol. 38, no. 1, pp. 15–24, 2017.

[45] L. K. Saarelainen, J. P. Turner, S. Shakib, N. Singhal, J. Hogan-Doran, R. Prowse, S. Johns, J. Lees, and J. S. Bell, “Potentially inappropriate medication use in older people with cancer: Prevalence and correlates,” Journal of Geriatric Oncology, vol. 5, no. 4, pp. 439–446, 2014.

[46] A. Ali, Y. P. Song, S. Mehta, H. Mistry, R. Conroy, C. Coyle, J. Logue, A. Tran, J. Wylie, T. Janjua, L. Joseph, J. Joseph, and A. Choudhury, “Palliative radiation therapy in bladder cancer—importance of patient selection: A retrospective multicenter study,” International Journal of Radiation Oncology*Biology*Physics, vol. 105, no. 2, pp. 389–393, 2019.

[47] N. Choudhury and Y. Nakamura, “Importance of immunopharmacogenomics in cancer treatment: Patient selection and monitoring for immune checkpoint antibodies,” Cancer Science, vol. 107, no. 2, pp. 107–115, 2016.

[48] M. A. Schonberg, E. R. Marcantonio, D. Li, R. A. Silliman, L. Ngo, and E. P. McCarthy, “Breast cancer among the oldest old: Tumor characteristics, treatment choices, and survival,” Journal of Clinical Oncology, vol. 28, no. 12, pp. 2038–2045, 2010.

[49] R. J. Simes, “Treatment selection for cancer patients: Application of statistical decision theory to the treatment of advanced ovarian cancer,” Journal of Chronic Diseases, vol. 38, no. 2, pp. 171–186, 1985.

[50] A. GERON, Hands-on machine learning with scikit-learn, Keras, and tensorflow: Concepts, tools and techniques to build Intelligent Systems. Beijing ; Boston ; Farnham etc.: O'Reilly, 2019.

[51] R. O. Duda, D. G. Stork, and P. E. Hart, Pattern classification and scene analysis. New York: Wiley, 2000.

[52] K. P. Murphy, Machine learning: A probabilistic perspective. Cambridge, MA: MIT Press, 2021.

[53] C. M. BISHOP, Pattern recognition and machine learning. SPRINGER-VERLAG NEW YORK, 2016.

[54] D. Needell, R. Saab, and T. Woolf, “Simple classification using binary data,” The Journal of Machine Learning Research, vol. 19, no. 1, pp. 2487–2516, 2018.

[55] E. Fitkov-Norris, S. Vahid, and C. Hand, “Evaluating the impact of categorical data encoding and scaling on neural network classification performance: The case of repeat consumption of identical cultural goods,” Engineering Applications of Neural Networks, pp. 343–352, 2012.

[56] Z. Khandezamin, M. Naderan, and M. J. Rashti, “Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier,” Journal of Biomedical Informatics, vol. 111, p. 103591, Nov. 2020.

[57] X.-Y. Liu, S.-B. Wu, W.-Q. Zeng, Z.-J. Yuan, and H.-B. Xu, “Logsum + L2 penalized logistic regression model for biomarker selection and cancer classification,” Scientific Reports, vol. 10, no. 1, 2020.

[58] H.-H. Huang, X.-Y. Liu, and Y. Liang, “Feature selection and cancer classification via sparse logistic regression with the hybrid L1/2 +2 regularization,” PLOS ONE, vol. 11, no. 5, 2016.

[59] Açıcı, K., Sümer, E. & Beyaz, S. Comparison of different machine learning approaches to detect femoral neck fractures in x-ray images. Health Technol. 11, 643–653 (2021).

[60] “BigQuery: Cloud Data Warehouse,” Google. [Online]. Available: https://cloud.google.com/bigquery. [Accessed: 13-May-2022].

EKLER

Meme Kanseri İLK 100 Öznitelik	Akciğer Kanseri İLK 100 Öznitelik	Prostat Kanseri İLK 100 Öznitelik	Mide Kanseri İLK 100 Öznitelik
diag-v667: Encounter for palliative care diag-79902: Hypoxemia diag-4280: Congestive heart failure, unspecified diag-3485: Cerebral edema diag-2762: Acidosis diag-78959: Other ascites pro-9604: Insertion of endotracheal tube diag-42731: Atrial fibrillation pro-9671: Continuous invasive mechanical ventilation for less than 96 consecutive hours diag-v861: Estrogen receptor negative status [ER-] pro-3897: Central venous catheter placement with guidance diag-2761: Hyposmolality and/or hyponatremia diag-v4986: Do not resuscitate status diag-v8741: Personal history of antineoplastic chemotherapy diag-2440: Postsurgical hypothyroidism diag-4168: Other chronic pulmonary heart diseases pro-9229: Other radiotherapeutic procedure diag-5789: Hemorrhage of gastrointestinal tract, unspecified diag-27669: Other fluid overload diag-4011: Benign essential hypertension diag-51881: Acute respiratory failure diag-570: Acute and subacute necrosis of liver diag-5845: Acute kidney failure with lesion of tubular necrosis diag-29420: Dementia, unspecified, without behavioral disturbance med-morphine sulfate (oral solution) 2 mg/ml: diag-78052: Insomnia, unspecified diag-e9320: Adrenal cortical steroids causing adverse effects in therapeutic use	diag-v667: Encounter for palliative care diag-51881: Acute respiratory failure diag-1976: Secondary malignant neoplasm of retroperitoneum and peritoneum diag-e9331: Antineoplastic and immunosuppressive drugs causing adverse effects in therapeutic use diag-5109: Empyema without mention of fistula diag-z515: Encounter for palliative care diag-4275: Cardiac arrest pro-3323: Other bronchoscopy diag-36570: Glaucoma stage, unspecified pro-3201: Endoscopic excision or destruction of lesion or tissue of bronchus diag-43491: Cerebral artery occlusion, unspecified with cerebral infarction pro-3326: Closed [percutaneous] [needle] biopsy of lung diag-56409: Other constipation diag-1972: Secondary malignant neoplasm of pleura diag-5070: Pneumonitis due to inhalation of food or vomitus diag-v1255: Personal history of pulmonary embolism diag-591: Hydronephrosis diag-45342: Acute venous embolism and thrombosis of deep vessels of distal lower extremity pro-4131: Biopsy of bone marrow diag-36510: Open-angle glaucoma, unspecified diag-78552: Septic shock pro-3406: Thoracoscopic drainage of pleural cavity diag-5845: Acute kidney failure with lesion of tubular necrosis diag-z66: Do not resuscitate pro-4011: Biopsy of lymphatic structure diag-5849: Acute kidney	diag-7310: Osteitis deformans without mention of bone tumor pro-9604: Insertion of endotracheal tube med-morphine infusion – comfort care guidelines: diag-51881: Acute respiratory failure diag-v667: Encounter for palliative care diag-99592: Severe sepsis diag-2875: Thrombocytopenia, unspecified pro-9925: Injection or infusion of cancer chemotherapeutic substance diag-7823: Edema diag-28803: Drug induced neutropenia diag-v4501: Cardiac pacemaker in situ diag-1977: Malignant neoplasm of liver, secondary diag-v4986: Do not resuscitate status diag-486: Pneumonia, organism unspecified diag-71535: Osteoarthrosis, localized, not specified whether primary or secondary, pelvic region and thigh diag-28522: Anemia in neoplastic disease pro-6011: Closed [percutaneous] [needle] biopsy of prostate diag-z515: Encounter for palliative care diag-0389: Unspecified septicemia diag-2853: Antineoplastic chemotherapy induced anemia diag-1713: Malignant neoplasm of connective and other soft tissue of lower limb, including hip diag-1970: Secondary malignant neoplasm of lung diag-570: Acute and subacute necrosis of liver pro-5011: Closed (percutaneous) [needle] biopsy of liver diag-78552: Septic shock diag-z66: Do not resuscitate	diag-2762: Acidosis diag-1976: Secondary malignant neoplasm of retroperitoneum and peritoneum diag-v667: Encounter for palliative care med-morphine infusion – comfort care guidelines: diag-5070: Pneumonitis due to inhalation of food or vomitus diag-v4581: Aortocoronary bypass status diag-5119: Unspecified pleural effusion diag-v1003: Personal history of malignant neoplasm of esophagus diag-60000: Hypertrophy (benign) of prostate without urinary obstruction and other lower urinary tract symptom (LUTS) diag-261: Nutritional marasmus diag-56400: Constipation, unspecified diag-1977: Malignant neoplasm of liver, secondary pro-9671: Continuous invasive mechanical ventilation for less than 96 consecutive hours diag-1961: Secondary and unspecified malignant neoplasm of intrathoracic lymph nodes diag-99592: Severe sepsis pro-3893: Venous catheterization, not elsewhere classified diag-79092: Abnormal coagulation profile diag-45341: Acute venous embolism and thrombosis of deep vessels of proximal lower extremity pro-3897: Central venous catheter placement with guidance diag-1505: Malignant neoplasm of lower third of esophagus diag-5849: Acute kidney failure, unspecified diag-5761: Cholangitis diag-27651: Dehydration

diag-4580: Orthostatic hypotension
diag-5533: Diaphragmatic hernia without mention of obstruction or gangrene
med-cefazolin:
diag-1985: Secondary malignant neoplasm of bone and bone marrow
diag-e9331: Antineoplastic and immunosuppressive drugs causing adverse effects in therapeutic use
diag-1984: Secondary malignant neoplasm of other parts of nervous system
diag-78829: Other specified retention of urine
diag-v4586: Bariatric surgery status
pro-8879: Other diagnostic ultrasound
diag-78701: Nausea with vomiting
diag-36900: Profound impairment, both eyes, impairment level not further specified
diag-486: Pneumonia, organism unspecified
diag-99592: Severe sepsis
diag-3694: Legal blindness, as defined in U.S.A.
diag-1978: Secondary malignant neoplasm of other digestive organs and spleen
pro-4513: Other endoscopy of small intestine
diag-7850: Tachycardia, unspecified
diag-6826: Cellulitis and abscess of leg, except foot
diag-70703: Pressure ulcer, lower back
diag-51882: Other pulmonary insufficiency, not elsewhere classified
diag-7455: Ostium secundum type atrial septal defect
diag-56400: Constipation, unspecified
diag-2930: Delirium due to conditions classified elsewhere
diag-36250: Macular degeneration (senile), unspecified
diag-73311: Pathologic fracture of humerus
diag-34590: Epilepsy, unspecified, without mention of intractable epilepsy
pro-3404: Insertion of intercostal catheter for drainage
diag-v160: Family history of

failure, unspecified
diag-2768: Hypopotassemia
diag-43889: Other late effects of cerebrovascular disease
diag-78062: Postprocedural fever
diag-5130: Abscess of lung
diag-20280: Other malignant lymphomas, unspecified site, extranodal and solid organ sites
diag-7802: Syncope and collapse
diag-v4986: Do not resuscitate status
diag-20410: Chronic lymphoid leukemia, without mention of having achieved remission
diag-34590: Epilepsy, unspecified, without mention of intractable epilepsy
diag-2753: Disorders of phosphorus metabolism
med-pilocarpine 4%:
pro-9904: Transfusion of packed cells
diag-1124: Candidiasis of lung
diag-59970: Hematuria, unspecified
diag-41519: Other pulmonary embolism and infarction
diag-49320: Chronic obstructive asthma, unspecified
diag-32723: Obstructive sleep apnea (adult)(pediatric)
diag-28489: Other specified aplastic anemias
diag-20500: Acute myeloid leukemia, without mention of having achieved remission
diag-7850: Tachycardia, unspecified
pro-387: Interruption of the vena cava
diag-2875: Thrombocytopenia, unspecified
diag-33394: Restless legs syndrome (RLS)
diag-7866: Swelling, mass, or lump in chest
med-acetaminophen w/codeine:
diag-2720: Pure hypercholesterolemia
diag-47811: Nasal mucositis (ulcerative)
diag-70704: Pressure ulcer, hip
diag-42821: Acute systolic heart failure
diag-42789: Other specified cardiac dysrhythmias
diag-2822: Anemias due to disorders of glutathione

diag-1890: Malignant neoplasm of kidney, except pelvis
diag-2536: Other disorders of neurohypophysis
diag-1991: Other malignant neoplasm without specification of site
diag-27652: Hypovolemia
pro-9672: Continuous invasive mechanical ventilation for 96 consecutive hours or more
diag-41071: Subendocardial infarction, initial episode of care
pro-4011: Biopsy of lymphatic structure
diag-4589: Hypotension, unspecified
diag-33394: Restless legs syndrome (RLS)
diag-2760: Hyperosmolality and/or hypernatremia
diag-1889: Malignant neoplasm of bladder, part unspecified
diag-5728: Other sequelae of chronic liver disease
diag-28529: Anemia of other chronic disease
diag-7821: Rash and other nonspecific skin eruption
diag-07032: Chronic viral hepatitis B without mention of hepatic coma without mention of hepatitis delta
diag-1961: Secondary and unspecified malignant neoplasm of intrathoracic lymph nodes
diag-v1087: Personal history of malignant neoplasm of thyroid
diag-2740: Gouty arthropathy
diag-44024: Atherosclerosis of native arteries of the extremities with gangrene
med-glycopyrrolate:
diag-42823: Acute on chronic systolic heart failure
diag-4241: Aortic valve disorders
med-cepastat (phenol) lozenge:
diag-v1254: Personal history of transient ischemic attack (TIA), and cerebral infarction without residual deficits
diag-2721: Pure hyperglyceridemia
diag-73313: Pathologic fracture of vertebrae
diag-78550: Shock, unspecified
diag-40390: Hypertensive

diag-1962: Secondary and unspecified malignant neoplasm of intra-abdominal lymph nodes
diag-1974: Secondary malignant neoplasm of small intestine including duodenum
diag-19889: Secondary malignant neoplasm of other specified sites
pro-4516: Esophagogastroduodenoscopy [EGD] with closed biopsy
diag-v1083: Personal history of other malignant neoplasm of skin
pro-9604: Insertion of endotracheal tube
diag-v153: Personal history of irradiation, presenting hazards to health
diag-1978: Secondary malignant neoplasm of other digestive organs and spleen
diag-7994: Cachexia
diag-56409: Other constipation
diag-5363: Gastroparesis
diag-78552: Septic shock
diag-4280: Congestive heart failure, unspecified
diag-5362: Persistent vomiting
diag-591: Hydronephrosis
diag-1975: Secondary malignant neoplasm of large intestine and rectum
diag-v1046: Personal history of malignant neoplasm of prostate
diag-4019: Unspecified essential hypertension
diag-5715: Cirrhosis of liver without mention of alcohol
diag-4430: Raynaud's syndrome
diag-45382: Acute venous embolism and thrombosis of deep veins of upper extremity
diag-78909: Abdominal pain, other specified site
diag-1986: Secondary malignant neoplasm of ovary
diag-56723: Spontaneous bacterial peritonitis
diag-v5861: Long-term (current) use of anticoagulants
pro-5198: Other percutaneous procedures on biliary tract
diag-2800: Iron deficiency anemia secondary to blood loss (chronic)
diag-78720: Dysphagia, unspecified
diag-5789: Hemorrhage of gastrointestinal tract,

malignant neoplasm of gastrointestinal tract
diag-v1582: Personal history of tobacco use
diag-1749: Malignant neoplasm of breast (female), unspecified
med-mannitol:
pro-8843: Arteriography of pulmonary arteries
med-loperamide:
diag-7837: Adult failure to thrive
med-albumin 25% (12.5g / 50ml):
diag-v153: Personal history of irradiation, presenting hazards to health
diag-2409: Goiter, unspecified
diag-07070: Unspecified viral hepatitis C without hepatic coma
diag-2875: Thrombocytopenia, unspecified
diag-v1251: Personal history of venous thrombosis and embolism
med-polyethylene glycol:
pro-9915: Parenteral infusion of concentrated nutritional substances
diag-78650: Chest pain, unspecified
diag-42833: Acute on chronic diastolic heart failure
diag-51181: Malignant pleural effusion
diag-79029: Other abnormal glucose
diag-5849: Acute kidney failure, unspecified
diag-7876: Incontinence of feces
med-guaifenesin er:
med-vitamin b complex:
med-magnesium sulfate replacement (critical care and oncology):
diag-3569: Unspecified hereditary and idiopathic peripheral neuropathy
diag-1960: Secondary and unspecified malignant neoplasm of lymph nodes of head, face, and neck
med-morphine sulfate (oral soln.):
diag-z515: Encounter for palliative care
diag-z853: Personal history of malignant neoplasm of breast
diag-0389: Unspecified septicemia
med-potassium chloride:
diag-78079: Other malaise and metabolism
diag-30473: Combinations of opioid type drug with any other drug dependence, in remission
diag-3331: Essential and other specified forms of tremor
diag-0312: Disseminated due to other mycobacteria
diag-v1046: Personal history of malignant neoplasm of prostate
diag-78791: Diarrhea
diag-2766: Fluid overload disorder
diag-2869: Other and unspecified coagulation defects
diag-2762: Acidosis
diag-2760: Hyperosmolality and/or hypernatremia
diag-3594: Toxic myopathy
pro-4513: Other endoscopy of small intestine
diag-29562: Schizophrenic disorders, residual type, chronic
diag-41401: Coronary atherosclerosis of native coronary artery
diag-51882: Other pulmonary insufficiency, not elsewhere classified
diag-1971: Secondary malignant neoplasm of mediastinum
diag-33189: Other cerebral degeneration
diag-56210: Diverticulosis of colon (without mention of hemorrhage)
diag-7812: Abnormality of gait
diag-1960: Secondary and unspecified malignant neoplasm of lymph nodes of head, face, and neck
diag-v103: Personal history of malignant neoplasm of breast
diag-59971: Gross hematuria
diag-9092: Late effect of radiation
diag-1173: Aspergillosis
diag-28529: Anemia of other chronic disease
diag-51181: Malignant pleural effusion
pro-403: Regional lymph node excision
diag-042: Human immunodeficiency virus [HIV] disease
diag-71941: Pain in joint, shoulder region
diag-5781: Blood in stool
chronic kidney disease, unspecified, with chronic kidney disease stage I through stage IV, or unspecified
diag-36573: Severe stage glaucoma
diag-99591: Sepsis
pro-3322: Fiber-optic bronchoscopy
pro-3897: Central venous catheter placement with guidance
diag-v1082: Personal history of malignant melanoma of skin
diag-4240: Mitral valve disorders
diag-25040: Diabetes with renal manifestations, type II or unspecified type, not stated as uncontrolled
diag-78820: Retention of urine, unspecified
diag-2762: Acidosis
diag-78061: Fever presenting with conditions classified elsewhere
diag-3383: Neoplasm related pain (acute) (chronic)
diag-2113: Benign neoplasm of colon
diag-e9426: Other antihypertensive agents causing adverse effects in therapeutic use
diag-3659: Unspecified glaucoma
diag-04111: Methicillin susceptible Staphylococcus aureus in conditions classified elsewhere and of unspecified site
diag-v462: Other dependence on machines, supplemental oxygen
pro-3722: Left heart cardiac catheterization
diag-e8889: Unspecified fall
diag-99529: Unspecified adverse effect of other drug, medicinal and biological substance
diag-2948: Other persistent mental disorders due to conditions classified elsewhere
diag-5733: Hepatitis, unspecified
diag-42822: Chronic systolic heart failure
pro-3961: Extracorporeal circulation auxiliary to open heart surgery
diag-7242: Lumbago
diag-v153: Personal history of irradiation, presenting hazards unspecified
diag-59971: Gross hematuria
pro-9915: Parenteral infusion of concentrated nutritional substances
diag-41519: Other pulmonary embolism and infarction
diag-5768: Other specified disorders of biliary tract
diag-1970: Secondary malignant neoplasm of lung
pro-8751: Percutaneous hepatic cholangiogram
med-octreotide acetate:
diag-7140: Rheumatoid arthritis
diag-78559: Other shock without mention of trauma
diag-2767: Hyperpotassemia
diag-v160: Family history of malignant neoplasm of gastrointestinal tract
diag-07032: Chronic viral hepatitis B without mention of hepatic coma without mention of hepatitis delta
diag-78951: Malignant ascites
diag-1508: Malignant neoplasm of other specified part of esophagus
pro-4639: Other enterostomy
diag-70703: Pressure ulcer, lower back
diag-2749: Gout, unspecified
diag-5990: Urinary tract infection, site not specified
med-magnesium sulfate:
diag-7823: Edema
diag-1987: Secondary malignant neoplasm of adrenal gland
diag-5609: Unspecified intestinal obstruction
pro-4432: Percutaneous [endoscopic] gastrojejunostomy
diag-73300: Osteoporosis, unspecified
diag-0389: Unspecified septicemia
med-propofol:
diag-311: Depressive disorder, not elsewhere classified
diag-52801: Mucositis (ulcerative) due to antineoplastic therapy
diag-51881: Acute respiratory failure
diag-5762: Obstruction of bile duct
diag-27800: Obesity, unspecified
diag-v1588: History of fall
diag-28860: Leukocytosis, unspecified

fatigue diag-7840: Headache pro-8523: Subtotal mastectomy diag-v161: Family history of malignant neoplasm of trachea, bronchus, and lung diag-25000: Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled pro-7855: Internal fixation of bone without fracture reduction, femur pro-8595: Insertion of breast tissue expander med-duloxetine: diag-42832: Chronic diastolic heart failure med-chlorpheniramine-hydrocodone: med-tizanidine: med-famotidine: diag-2689: Unspecified vitamin D deficiency diag-z66: Do not resuscitate diag-70722: Pressure ulcer, stage II	diag-38612: Vestibular neuronitis diag-78551: Cardiogenic shock diag-v1254: Personal history of transient ischemic attack (TIA), and cerebral infarction without residual deficits pro-3249: Other lobectomy of lung diag-2851: Acute posthemorrhagic anemia diag-44020: Atherosclerosis of native arteries of the extremities, unspecified diag-7837: Adult failure to thrive diag-7823: Edema diag-25080: Diabetes with other specified manifestations, type II or unspecified type, not stated as uncontrolled diag-3559: Mononeuritis of unspecified site diag-28850: Leukocytopenia, unspecified pro-0392: Injection of other agent into spinal canal diag-0414: Escherichia coli [E. coli] infection in conditions classified elsewhere and of unspecified site diag-2749: Gout, unspecified	to health diag-60000: Hypertrophy (benign) of prostate without urinary obstruction and other lower urinary tract symptom (LUTS) diag-9971: Cardiac complications, not elsewhere classified diag-e9413: Sympatholytics [antiadrenergics] causing adverse effects in therapeutic use pro-605: Radical prostatectomy diag-1628: Malignant neoplasm of other parts of bronchus or lung diag-7804: Dizziness and giddiness diag-41401: Coronary atherosclerosis of native coronary artery med-ketorolac: diag-v8531: Body Mass Index 31.0-31.9, adult diag-29570: Schizoaffective disorder, unspecified diag-43820: Late effects of cerebrovascular disease, hemiplegia affecting unspecified side diag-g935: Compression of brain diag-4239: Unspecified disease of pericardium diag-3051: Tobacco use disorder pro-30283b1: Transfusion of Nonautologous 4-Factor Prothrombin Complex Concentrate into Vein, Percutaneous Approach pro-9910: Injection or infusion of thrombolytic agent diag-2724: Other and unspecified hyperlipidemia pro-8856: Coronary arteriography using two catheters diag-2866: Defibrination syndrome diag-e9478: Other drugs and medicinal substances causing adverse effects in therapeutic use diag-v4582: Percutaneous transluminal coronary angioplasty status	med-ipratropium bromide neb: med-hydromorphone infusion – comfort care guidelines: diag-e9308: Other specified antibiotics causing adverse effects in therapeutic use pro-3324: Closed [endoscopic] biopsy of bronchus med-multivitamins: pro-9955: Prophylactic administration of vaccine against other diseases pro-9995: Stretching of foreskin med-ranitidine: diag-v5869: Long-term (current) use of other medications diag-5728: Other sequelae of chronic liver disease diag-e8788: Other specified surgical operations and procedures causing abnormal patient reaction, or later complication, without mention of misadventure at time of operation diag-53789: Other specified disorders of stomach and duodenum pro-5011: Closed (percutaneous) [needle] biopsy of liver pro-3895: Venous catheterization for renal dialysis diag-2639: Unspecified protein-calorie malnutrition

EK 1: Kanser grupları için seçilen İLK 100 öznitelik detayı