Fırat Üniversitesi Müh. Bil. Dergisi
33(2), 547-565, 2021

Araştırma Makalesi
https://doi.org/10.35234/fumbd.881228

Kanser Teşhisinde Protein Haritalama Tekniklerinin Başarımlarının Derin Öğrenme Kullanılarak Karşılaştırılması

Talha Burak ALAKUŞ[1*], İbrahim TÜRKOĞLU[2]
[1] Yazılım Mühendisliği Bölümü, Mühendislik Fakültesi, Kırklareli Üniversitesi, Kırklareli, Türkiye
[2] Yazılım Mühendisliği Bölümü, Teknoloji Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye
[*1] talhaburakalakus@klu.edu.tr, [2] iturkoglu@firat.edu.tr

(Geliş/Received: 16/02/2021; Kabul/Accepted: 03/06/2021)

Öz: Kanser, dünya çapında çoğu insanın ölmesine neden olan ve birçok farklı alt tiplerden oluşan heterojen bir hastalıktır. Bir kanser türünün erken teşhisi ve prognozu, hastaların sonraki klinik takibini kolaylaştırabildiği için kanser araştırmalarında bir gereklilik haline gelmiştir. Bunun için en çok kullanılan yöntemlerden birisi histolojik incelemedir. Ancak bu yöntemde çok sayıda gözlemciler arası değişkenlik bulunmakta, bu ise inceleme sürecinin uzun olmasına ve zaman almasına neden olmaktadır. Bu dezavantajın önüne geçmek için araştırmacılar hesaplama-tabanlı yaklaşımlara yönelmişler ve kanserli proteinlerin belirlenmesi için protein-protein etkileşimleri, protein etkileşim ağları ve moleküler parmak izleri yöntemlerinden yararlanmaktadırlar. Bu yöntemler arasında, çeşitli çalışmalar genomik bilgilerden de kanserli hücrelerin tespit edilebildiğini göstermiştir. Kansere ait genlerin dizilimlerine göre belirli kanser türlerinin belirlenebildiği ve bu süreçte yapay öğrenme tabanlı yaklaşımların etkili olduğu görülmüştür. Bu çalışmada, derin öğrenme algoritmalarından birisi olan tekrarlayıcı sinir ağı mimarisi kullanılmış ve insana ait mesane, kolon ve prostat kanserlerinin, protein dizilimlerine göre sınıflandırılması yapılmıştır. Çalışmanın ana amacı protein haritalama tekniklerinin, kanserli genleri sınıflandırmadaki performanslarının kıyaslanmasıdır. Çalışma, verilerin elde edilmesi, protein dizilimlerinin sayısallaştırılması, derin öğrenme model uygulamasının geliştirilmesi ve protein haritalama tekniklerinin başarımının karşılaştırılması olmak üzere dört aşamadan meydana gelmektedir. Protein dizilimlerini sayısallaştırmak için AESNN1, hidrofobiklik, tam sayı, Miyazawa enerjileri ve rastgele kodlama yöntemleri ele alınmıştır. Çalışmanın sonunda, mesane kanseri için en yüksek doğruluk değeri %87.15 ile AESNN1 haritalama yöntemiyle, kolon kanseri ve prostat kanseri için ise en yüksek doğruluk değeri sırasıyla %94.40 ve %75.45 olarak Miyazawa enerjileri ve rastgele kodlama protein haritalama yöntemi ile elde edilmiştir. Bu çalışma ile yapay öğrenme ve protein haritalama tekniklerinin, kanserli protein dizilimlerinin belirlenmesinde etkili olduğu gözlemlenmiştir.

Anahtar kelimeler: Kanser, biyoenformatik, derin öğrenme, proteinler

Comparing the Performance of Protein Mapping Techniques in Cancer Diagnosis Using Deep Learning

Abstract: Cancer is a heterogeneous disease made up of many different subtypes that cause most people to die worldwide. Early diagnosis and prognosis of a cancer type has become a necessity in cancer research as it can facilitate subsequent clinical follow-up of patients. One of the most used methods for this is histological examination. However, there is a large number of inter-observer variability in this method, which causes the examination process to be long and time consuming. To avoid this disadvantage, researchers have turned to computation-based approaches and utilize protein-protein interactions, protein interaction networks, and molecular fingerprints to identify cancerous proteins. Among these methods, various studies have shown that cancerous cells can also be detected from genomic information. It has been observed that certain types of cancer can be determined according to the sequences of genes belonging to cancer, and artificial learning-based approaches are effective in this process. In this study, recurrent neural network architecture, one of the deep learning algorithms, was used and the classification of human bladder, colon and prostate cancers according to their protein sequences was made. The study consists of four stages: data acquisition, mapping of protein sequences, development of deep learning model application, and comparison of the performance of protein mapping techniques. AESNN1, hydrophobicity, integer, Miyazawa energies and random coding methods were discussed to map protein sequences. At the end of the study, the highest accuracy value for bladder cancer was obtained with AESNN1 mapping method with 87.15%, and the highest accuracy value for colon cancer and prostate cancer was obtained with Miyazawa energies and random coding protein mapping method as 94.40% and 75.45%, respectively. With this study, it has been observed that artificial learning and protein mapping techniques are effective in determining cancerous protein sequences.

Key words: Cancer, bioinformatics, deep learning, proteins

[*] Sorumlu yazar: talhaburakalakus@klu.edu.tr. Yazarların ORCID Numarası: [1] 0000-0003-3136-3341, [2] 0000-0003-4938-4167

1. Giriş

Kanser, dünya çapında insan sağlığını etkileyen ve tehdit eden en yaygın hastalıklardan birisidir. Göğüs, deri, akciğer, kolon, prostat olmak üzere 100.000'den fazla kanser türü bulunmaktadır [1]. Kanser genellikle tümör baskılayıcı genler, hücre döngüsü düzenleyiciler ve proto-onkojenler üzerinde ortaya çıkan zararlı mutasyonlardan meydana gelmektedir [2]. Kanser tedavilerinde en önemli unsur kanserin erken teşhis edilmesidir. Bu durumda beklenen hayat süresi ve hayat kalitesi artmaktadır. Kanserin erken teşhisinde kan testleri, röntgen, manyetik rezonans görüntüleme, endoskopi, genetik görüntüleme, histolojik inceleme gibi çeşitli sayıda yöntemler bulunmaktadır. Ancak bu yöntemlerin insan gücü ve laboratuvar ekipmanı gerektirmesi, uzman yorumuna ihtiyaç duyması, hem teşhis işlemin zaman almasına hem de maliyetli olmasına neden olmaktadır [3,4]. Bu dezavantajlardan dolayı, araştırmacılar kanser teşhisinin erken tahmini için yeni stratejiler geliştirmektedirler. Tıp alanında yeni teknolojilerin ortaya çıkmasıyla, büyük miktarda kanser verisi toplanabilmekte ve tıbbi araştırmalarda etkili bir şekilde kullanılabilmektedir [5].

Biyoenformatik çalışmalarında, hesaplama-tabanlı yaklaşımların ve yapay öğrenme yöntemlerinin başarılı olması, kansere neden olan proteinlerin bu yöntemlerle analiz edilmesinin önünü açmıştır. Makine öğrenmesi ve derin öğrenme mimarileri, kansere neden olan proteinlerin analizinde kullanılmakta ve erken teşhis çalışmalarında sıklıkla değerlendirilmektedir [6,8]. Protein dizilimlerinin yapay öğrenme teknikleri ile analiz edilebilmesi için dizilimlerin sayısallaştırılması gerekmektedir. Bilindiği üzere protein dizilimleri 20 adet amino asit kodlarından (harflerden) meydana gelmektedir. Ham protein verilerinin yapay öğrenme yöntemleri ile değerlendirilmesi mümkün değildir [9]. Protein dizilimlerinin sayısallaştırılmasıyla gerçekleştirilen çalışmalar hesaplama-tabanlı olarak ifade edilmektedir. Literatürde belirli sayıda protein haritalama yöntemleri mevcut olmakla beraber bu yöntemler belirli kategorilere ayrıştırılmıştır. Tablo 1'de çeşitli protein haritalama yöntemleri ve bunların kategorileri verilmiştir.

Tablo 1. Protein haritalama yöntemleri ve ait olduğu kategoriler

Kategori	Protein Haritalama Yöntemleri
Fizikokimyasal-tabanlı	Hidrofobiklik Atchley Faktörleri
Evrişim-tabanlı	PAM250 PSSM
Yapı-tabanlı	Miyazawa Enerjileri
Makine öğrenmesi-tabanlı	AESNN1 ProtVec

Protein haritalama yöntemleri proteinlerin arasındaki etkileşimleri belirlemek [10], ilaç-hedef etkileşimlerini belirlemek [11], protein ailelerini sınıflandırmak [12], protein fonksiyonlarını tahmin etmek [13], protein ikincil yapılarını analiz etmek [14] gibi çeşitli çalışmalarda sıklıkla kullanılmaktadır. Çalışmalardan anlaşıldığı üzere kullanılan protein haritalama yöntemleri ve yapay öğrenme tekniklerinin tahmin ve teşhis işlemlerinde belirleyici olduğu gözlemlenmiştir

Bu çalışmada, belirli protein haritalama yöntemleri kullanılarak, çeşitli kanserlere ait proteinlerin sınıflandırılması yapılmıştır. Çalışmadaki ana amaç protein haritalama tekniklerinin kanser içeren genleri sınıflandırmadaki başarımlarının kıyaslanmasıdır. Çalışma dört aşamadan meydana gelmektedir. Birinci aşamada mesane, kolon ve prostat kanserine neden olan protein dizilimleri elde edilmiştir. İkinci aşamada elde edilen protein dizilimleri sayısal ifadelere dönüştürülmüş ve bu işlem için AESNN1, hidrofobiklik, tam sayı, Miyazawa enerjileri ve rastgele kodlama olmak üzere beş farklı yöntem kullanılmıştır. Sayısal haritalama işleminin ardından tekrarlayıcı sinir ağı modeli tasarlanmış ve bu model ile sınıflandırma işlemi gerçekleştirilmiştir. Çalışmanın son kısmında ise kullanılan protein haritalama tekniklerinin başarımları kıyaslanmış ve değerlendirme ölçütleri olarak ise doğruluk, AUC (Area Under Curve – Eğri Altında Kalan Alan), F1-skor ve kesinlik skorları kullanılmıştır.

Çalışmanın öne çıkan kısımları şu şekilde ifade edilebilir:

- Bu çalışma ile kanser içeren proteinlerin derin öğrenme mimarisi ile analizi yapılarak, protein haritalama yöntemlerinin başarımları kıyaslanmıştır.
- Kullanılan protein haritalama tekniklerinin kanser içeren proteinleri belirlemede etkili bir role sahip olduğu gözlemlenmiştir.
- Derin öğrenme algoritmasının kullanılmasıyla hem bu alanda hem de diğer biyoenformatik alanındaki başarımları pekiştirilmiştir.

Çalışmanın diğer kısımları şu şekilde organize edilmiştir. İkinci kısımda bu alanda yapılmış olan diğer çalışmalar irdelenmiş ve kullanılan yöntem, veri seti ve başarım ölçütleri hakkında bilgiler verilmiştir. Üçüncü kısımda bu çalışmada kullanılan veri setinden bahsedilerek, değerlendirilen protein haritalama yöntemleri ele alınmıştır. Dördüncü kısımda tasarlanan tekrarlayıcı sinir ağından bahsedilmiş ve parametreleri liste halinde verilmiştir. Ayrıca bu kısımda protein haritalama yöntemlerinin başarım ölçütleri hesaplanarak kıyaslama işlemi de gerçekleştirilmiştir. Bunlara ek olarak bu kısımda çalışmanın sonuçları irdelenmiş ve tartışma yapılarak avantaj ve dezavantajlardan bahsedilmiştir. Son kısımda ise çalışmanın etkisi, literatüre katkısı ve gelecek çalışmalar üzerinde durulmuştur.

2. İlgili Çalışmalar

Bu kısımda literatürde kanser içeren proteinlerin sınıflandırılmasına yönelik gerçekleştirilen yapay öğrenmeye dayalı çalışmalardan bahsedilmiştir. Belirli çalışmalar incelenmiş, kullanılan yapay öğrenme, protein haritalama ve veriler hakkında bilgiler verilmiş ve çalışmaların başarım ölçütleri ele alınmıştır. [1] numaralı çalışmada, araştırmacılar on iki farklı türdeki kanser tipini genom bilgisi ve derin öğrenme modeli kullanarak tanımlamışlardır. Çalışmada kanser türlerini belirleyebilmek için genom üzerindeki nokta mutasyonlardan yararlanmışlardır. Araştırmacılar sağlıklı örnekleri IGSR (The Internatioanl Genome Sample Resource) veri setinden elde ederken, tümör içeren örnekleri ise TCGA (The Cancer Genome Atlas) veri setinden toplamışlardır. Çalışmada mesane, göğüs, kolon, beyin tümörü, böbrek, gliyom, akciğer, yumurtalık, prostat, mide, pankreas ve rahim kanserleri üzerinde durulmuştur. Toplamda 1.991 adet sağlıklı veri kullanılırken, 6.083 adet kanser içeren proteinler değerlendirilmiştir. Derin öğrenme modeli olarak ise derin sinir ağları tasarlanmıştır. Protein dizilimleri bir-sıcak yöntemi ile sayısallaştırılmış ve sınıflandırılmaya hazır hale getirilmiştir. Çalışmanın sonunda ortalama %87.42 oranında doğruluk elde edilmiştir. Başka bir çalışmada araştırmacılar protein-protein etkileşimlerine dayanarak kanserle ilişkili genlerin karşılaştırmalı analizini yapmışlardır [15]. Çalışmada mesane, kolon, böbrek ve tiroit kanserlerine ait veriler kullanılmıştır. Mesane için 3.704 adet farklı gen ifadeleri kullanılırken, 2.081 adet protein etkileşimi ele alınmıştır. Kolon kanseri için 3.515 adet farklı gen ifadesi ve 1.944 adet protein etkileşim verisi kullanılmıştır. Bu oran böbrek kanseri ve tiroit kanseri için azalmış ve sırasıyla 1.257 ve 1.380 adet farklı gen ifadesi değerlendirilmiştir. Protein etkileşimi için ise sırasıyla 728 ve 715 adet veri kullanılmıştır. Farklı gen ifadeleri GEO (Gene Expression Omnibus) veri setinden elde edilmiştir. Proteinlere ait etkileşimler ise DIP (Database of Interaction Proteins), IntAct, BIND (Biomolecular Interaction Network Database), HPRD (Human Protein Reference Database) ve MINT (Molecular Interaction Database) veri setlerinden toplanmıştır. Çalışmanın sonunda önerilen yöntemin etkili olduğu belirlenmiştir. [16] numaralı çalışmada araştırmacılar kanser içeren proteinleri belirleyebilmek için protein etkileşimlerinden, alan frekansından ve alan etkileşim verilerinden yararlanmışlardır. Çalışmanın ilk aşamasında kanser içeren proteinler çeşitli veri setlerinden elde edilmiştir. Bu amaç için kullanılan veri setleri TAG (Tumor Associated Gene), MSKCC (Memorial Sloan-Kettering Cancer Center), Tayvan Yang Ming Üniversitesi, OMIM (Online Medelian Inheritance in Man) ve HLungDB (Human Lung Database)'dir. Bunun yanı sıra protein etkileşim verileri ise BioGrid ve Swiss-Prot veri setlerinden elde edilmiştir. Sınıflandırma işlemi için çeşitli makine öğrenmesi algoritmaları kullanılmış ve başarımları doğruluk, belirlilik, hassasiyet, F1-skor, AUC gibi başarım ölçütleri ile belirlenmiştir. En iyi başarım akciğer kanseri ile elde edilmiş ve %89.4 oranında doğruluk hesaplanmıştır. [17] numaralı çalışmada evrişimsel sinir ağları kullanılarak akciğer kanserine neden olan proteinlerin sınıflandırılması yapılmış ve bu amaç için protein etkileşim ağından ve gen ifadesi bilgilerinden yararlanılmıştır. Çalışmada protein etkileşim verileri HINT (The Homologous Interactions) veri setinden elde edilmiş ve toplamda 47.358 adet protein etkileşim verileri kullanılmıştır. Gen ifadesi bilgisi ise GEO veri seti araçlığı ile toplanmış ve sadece akciğer kanserine neden olan proteinler değerlendirilmiştir. Bu veri seti ile 152 adet kanser içermeyen ve 487 adet kanser içeren örnekler kullanılmıştır. Sınıflandırma işlemi için evrişimsel sinir ağı tasarlanmış ve sınıflandırıcının başarımı doğruluk, kesinlik, özgüllük ve duyarlılık ölçütleri ile belirlenmiştir. Çalışmanın sonunda %83.15 oranında doğruluk değeri elde edilmiştir.

[18] numaralı çalışmada makine öğrenmesi algoritmaları kullanılarak, protein özelliklerine dayalı akciğer tümörü türlerinin tahmini yapılmıştır. Akciğer tümörüne neden olan üç farklı tümör türü üzerinde durulmuş ve bu tümör türlerine ait proteinler kullanılmıştır. Çalışmada sınıflandırıcı olarak DVM (Destek Vektör Makineleri) kullanılmış ve farklı parametreler üzerinden sınıflandırıcının performansı kıyaslanmıştır. Performanslar Kappa değeri ve doğruluk ile belirlenmiş ve en iyi sonuçlar doğrusal DVM ile elde edilmiştir. Bu sınıflandırıcı ile en yüksek %67.42 oranında doğruluk ve %40.85 oranında Kappa değeri elde edilmiştir. [35] numaralı çalışmada araştırmacılar derin öğrenme kullanarak gen ifadelerine dayalı kanser sınıflandırması gerçekleştirmişlerdir. Çalışmada otuz iki farklı kanser türüne ait gen ifadeleri elde edilmiş ve toplamda çalışma için 10.267 adet gen dizilimi kullanılmıştır. Sınıflandırma işlemi için araştırmacılar iki-boyutlu evrişimsel sinir ağı geliştirmişlerdir.

Önerilen derin öğrenme modelinin başarımı kesinlik, hassaslık, doğruluk ve F1-skor ile belirlenmiştir. Bunlara ek olarak ise RO (Rastgele Orman), DVM, kNN (k En Yakın Komşu) sınıflandırıcıları ile de karşılaştırılmıştır. Çalışmanın sonunda en yüksek doğruluk skoru önerilen yöntem ile elde edilmiş ve %95.65 oranında bir değer hesaplanmıştır. Başka bir çalışmada ise derin öğrenme modeli ile kanser moleküllerinin alt tipleri sınıflandırılmıştır [36]. Çalışmada sadece kolorektal kanserine ait veriler kullanılmış ve bu kansere ait mikro dizi ve RNA dizilimleri sınıflandırmaya tabi tutulmuştur. Ardından dizilimler GSEA (Gene Set Enrichment Analysis – Gen Kümesi Zenginleştirme Analizi) uygulaması ile analiz edilerek, bu dizilimlerin fonksiyonel spektrum özellikleri elde edilmiştir. Daha sonra bu özellikler YSA (Yapay Sinir Ağları) modeli geliştirilerek sınıflandırılmış ve kolorektal kanserine ait moleküllerin alt tipleri belirlenmiştir. Çalışmanın son kısmında ise geliştirilmiş olan YSA modeli RO, DVM ve doğrusal regresyon sınıflandırıcıları ile de kıyaslanmıştır. Önerilen YSA sınıflandırıcısının performansı doğruluk, ortalama hassaslık ve ortalama özgüllük değerleri ile belirlenmiştir. Çalışmanın sonunda yaklaşık %92 oranında bir doğruluk sonucu elde edilmiştir.

[37] numaralı çalışmada araştırmacılar meme tümörü morfolojisini ve gen ifadelerini kullanarak sınıflandırma yapmışlar ve bunun için derin öğrenme modeli kullanmışlardır. Çalışmada toplamda yirmi üç kanser hastasından 30.612 adet gen ifadesi elde edilmiş ve bu genlerin ayrımı yapılmıştır. Gen ifadeleri görüntülere dönüştürülmüş ve 224x224 boyutunda görüntüler elde edilmiştir. Ardından evrişimsel sinir ağları modeli tasarlanmış ve gen ifadelerinin sınıflandırılması gerçekleştirilmiştir. Önerilen yöntemin performansı korelasyon ölçütü ile belirlenmiş ve %48 oranında bir başarım elde edilmiştir. Başka bir çalışmada ise araştırmacılar yüksek-boyutlu genomik verileri kullanarak alt kanser türlerinin teşhisini derin öğrenme yöntemi ile gerçekleştirmişlerdir [38]. Çalışmada meme kanserine yönelik bir uygulama gerçekleştirilmiştir. 1989 adet kanserli ve 144 adet normal dokular kullanılarak toplamda 25.160 adet gen elde edilmiştir. Gen ifadeleri PAM250 matrisi kullanılarak sayısallaştırılmıştır. Sınıflandırma işlemi için YSA modeli tasarlanmış ve önerilen modelin performansı doğruluk skoru ile ölçülmüştür. Çalışmanın sonunda önerilen yöntem ile ortalama %96.5 oranında bir başarım elde edilmiştir.

3. Materyal ve Yöntemler

3.1. Protein verileri

Bu kısımda kullanılan protein verileri ve protein haritalama teknikleri ele alınmıştır. Çalışmada mesane, kolon ve prostat kanserine neden olan proteinler kullanılmıştır. Bu proteinlere ait genel bilgiler Tablo 2'de verilmiştir.

Tablo 2. Farklı kanser türleri için belirlenmiş olan normal ve hastalıklı dokuların gen serileri

Kanser Türü	Gen Serisi	Normal Gen Bilgileri	Hastalıklı Gen Bilgileri
Mesane	GSE746	GSM180991, GSM180992, GSM180993	GSM180994, GSM180995, GSM180996, GSM180997, GSM180998, GSM180999, GSM181000, GSM181001, GSM181002
Kolon	GSE4107	GSM93938, GSM93939, GSM93941, GSM93943, GSM93944, GSM93946, GSM93948, GSM93950, GSM93952, GSM93954	GSM93789, GSM93920, GSM93921, GSM93922, GSM93923, GSM93924, GSM93925, GSM93926, GSM93927, GSM93928, GSM93929, GSM93932
Prostat	GSE3325	GSM74875, GSM74876, GSM74877, GSM74878, GSM74879, GSM74880	GSM74881, GSM74882, GSM74883, GSM74884, GSM74885, GSM74886, GSM74887, GSM74888, GSM74889, GSM74890, GSM748991, GSM74892, GSM74893

Tablo 2'de her bir kanser türü için verilen gen serileri mikroçipler üzerinde bulunan serilerdir. Bu serilerin ham bir şekilde kullanılması mümkün değildir. Bunun en büyük nedeni mikroçip üzerinde bulunan bu seriler protein dizilimleri içermemektedir. Her bir seri çeşitli gen ifadelerinden oluşmaktadır. Örnek verilecek olursa kolon kanserinin gen serisi GSE4107'dir. Bu seriye ait toplamda 10 adet normal gen bilgisi bulunurken, 12 adet hastalıklı gen bilgisi bulunmaktadır. Gen bilgileri de tıpkı gen serileri gibi mikroçipler üzerinde saklanır ve erişilebilir bir protein dizilimleri bulunmamaktadır. Bu gen bilgilerinde protein dizilimlerini ifade edebilmek için, gen bilgilerinin farklı gen ifadeleri ile analiz edilmesi gerekmektedir. Her bir gen serisinin farklı gen ifadelerini elde edebilmek için çalışmada GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) programı kullanılmıştır. Tablo 3'te her bir kansere ait gen serisinin farklı gen ifadeleri sayısı verilmiştir.

Tablo 3. Mesane, kolon ve prostat kanserine ait farklı gen ifadeleri

Kanser Türü	Farklı Gen İfadeleri (Normal)	Farklı Gen İfadeleri (Hastalıklı)	Toplam Farklı Gen İfadeleri
Mesane	41.960	944	42.904
Kolon	16.257	26.114	42.641
Prostat	18.283	24.621	42.904

Tablo 3'ten de anlaşıldığı üzere, mesane kanseri için toplamda 42.904 adet, kolon kanseri için 42.641 adet ve prostat kanseri için ise 42.904 adet farklı gen ifadesi belirlenmiştir. Farklı gen ifadelerinin sayısı her bir kanser türü için aynı olmadığından, veri seti dengesiz bir durumdadır. Veri setinin dengeli olabilmesi için tüm kanser türleri için farklı gen ifadesi sayısı 944'e kadar indirilmiştir. İfadelerin bu değere indirilmesinin nedeni en düşük gen ifadesi değerinin 944 olmasıdır. Bu sayede her bir kanser türü için eşit sayıda farklı gen ifadeleri kullanılmıştır. Ardından bu farklı gen ifadelerinin protein dizilimleri elde edilmiş ve çalışmada kullanılmıştır. Protein dizilimleri UniProt (https://www.uniprot.org/) veri setinden elde edilmiştir.

3.2. Protein haritalama yöntemleri

Protein dizilimlerinin yapay öğrenme teknikleri ile analiz edilebilmesi için dizilimlerin haritalanması gerekmektedir. Bu amaç için literatüde çeşitli sayıda yöntemler bulunmaktadır [34]. Bu çalışmada da belirli protein haritalama teknikleri kullanılmış ve proteinler sayısal ifadelere dönüştürülmüştür. Bu kısımda kullanılan protein haritalama teknikleri hakkında bilgiler verilmiştir.

3.2.1. AESNN1 protein haritalama yöntemi

Bu protein haritalama yönteminde, proteinlere ait sayısal ifadeler, proteinlerin yapı bilgilerinden elde edilmiştir [20]. Her bir amino asit tipi 1 boyutlu vektör ile ifade edilmektedir. AESNN1 haritalama yöntemi makine öğrenmesi tabanlı bir yöntemdir. Tablo 4'te AESNN1 protein haritalama yönteminin sayısal değerleri verilmiştir.

Tablo 4. Amino asitlerin AESNN1 değerleri

Amino Asit Kodu	Sayısal İfade	Amino Asit Kodu	Sayısal İfade
A	-0.99	L	-0.92
R	0.28	K	-0.63
N	0.77	M	-0.80
D	0.74	F	0.87
C	0.34	P	-0.99
Q	0.12	S	0.99
E	0.59	T	0.42
G	-0.79	W	-0.13
H	0.08	Y	0.59
I	-0.77	V	-0.99

Tablo 4'teki değerlere göre $S(n) = [ALKRNM]$ şeklinde bir protein dizisi $S(n) = [-0.99 - 0.92 - 0.63\ 0.28\ 0.77 - 0.80]$ biçiminde ifade edilir.

3.2.2. Hidrofobiklik protein haritalama yöntemi

Hidrofobiklik protein haritalama yönteminde, amino asit dizilimleri proteinlerin kimyasal bilgilerine dayanarak sayısallaştırılmaktadır [21]. Protein dizilimlerinin polipeptit zincirlerindeki hidrofilik ve hidrofobik özelliklerinden yararlanmaktadır. Bundan dolayı fizikokimyasal-tabanlı bir yöntem olarak değerlendirilmektedir. Tablo 5'te hidrofobiklik protein haritalama yönteminin sayısal değerleri verilmiştir.

Tablo 5. Amino asitlerin hidrofobiklik değerleri

Amino Asit Kodu	Sayısal İfade	Amino Asit Kodu	Sayısal İfade
A	1.8	L	3.8
R	-4.5	K	-3.9
N	-3.5	M	1.9
D	-3.5	F	2.8
C	2.5	P	-1.6
Q	-3.5	S	-0.8
E	-3.5	T	-0.7
G	-0.4	W	-0.9
H	-3.2	Y	-1.3
I	4.5	V	4.2

Tablo 5'teki değerlere göre $S(n) = [ALKRNM]$ şeklinde bir protein dizisi $S(n) = [1.8\ 3.8 - 3.9 - 4.5 - 3.5\ 1.9]$ biçiminde ifade edilir.

3.2.3. Tam sayı protein haritalama yöntemi

Karakter tabanlı bir protein haritalama yöntemidir. Belirli bir protein bilgisine ihtiyaç duyulmamaktadır. Amino asit kodları alfabetik sıraya göre yerleştirilir ve ilk sırada bulunan amino asit değerine 1 verilir ve bu değer amino asit kodlarının sırasına göre bir bir arttırılarak devam edilir. Tablo 6'da tam sayı protein haritalama yönteminin sayısal değerleri verilmiştir.

Tablo 6. Amino asitlerin tam sayı değerleri

Amino Asit Kodu	Sayısal İfade	Amino Asit Kodu	Sayısal İfade
A	1	L	11
R	2	K	12
N	3	M	13
D	4	F	14
C	5	P	15
Q	6	S	16
E	7	T	17
G	8	W	18
H	9	Y	19
I	10	V	20

Tablo 6'daki değerlere göre $S(n) = [ALKRNM]$ şeklinde bir protein dizisi $S(n) = [1\ 11\ 12\ 2\ 3\ 13]$ biçiminde ifade edilir.

3.2.4. Miyazawa enerjileri protein haritalama yöntemi

Miyazawa enerjileri yapı-tabanlı bir protein haritalama yöntemi olup, protein dizilimlerindeki kalıntılar arasındaki enerjileri belirlemek için önerilmiştir [22]. Bu yöntemde protein dizilimlerindeki temas enerjileri regresyon katsayıları ile elde edilmiştir. Temas enerjilerinin hesaplanmasından sonra protein dizilimlerinin doğrulama enerji değerleri dikkate alınmıştır. Tablo 7'de amino asit kodlarının Miyazawa enerji değerleri verilmiştir.

Tablo 7. Amino asitlerin Miyazawa enerji değerleri

Amino Asit Kodu	Sayısal İfade	Amino Asit Kodu	Sayısal İfade
A	-2.51	L	-5.79
R	-1.39	K	0.13
N	-1.59	M	-6.06
D	-0.96	F	-6.85
C	-5.44	P	-0.18
Q	-0.89	S	-1.48
E	-1.18	T	-1.72
G	-2.17	W	-5.42
H	-2.78	Y	-3.55
I	-6.22	V	-4.94

Tablo 7'deki değerlere göre $S(n) = [ALKRNM]$ şeklinde bir protein dizisi $S(n) = [-2.51 - 5.79\ 0.13 - 1.39 - 1.59 - 6.06]$ biçiminde ifade edilir.

3.2.5. Rastgele kodlama protein haritalama yöntemi

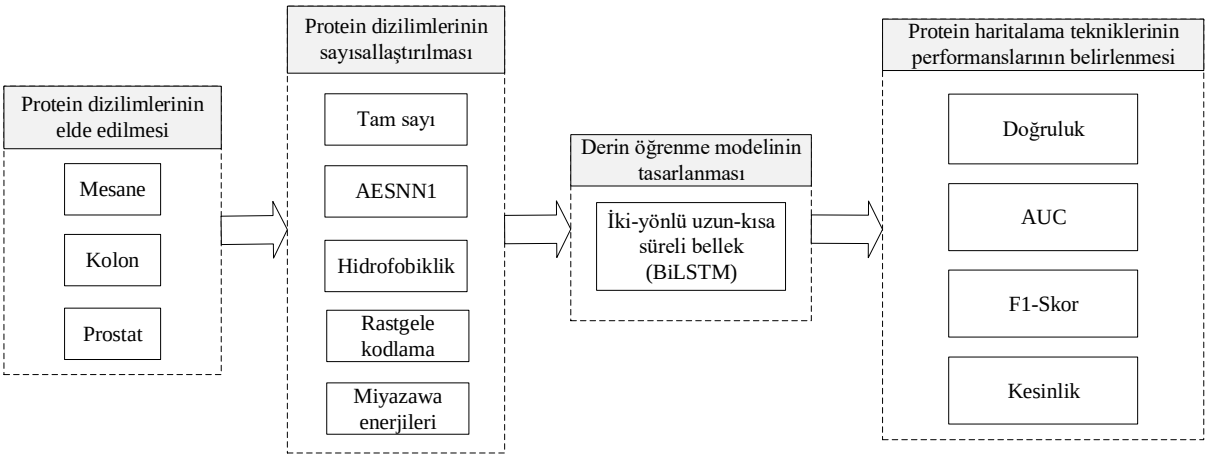
Rastgele kodlama protein haritalama yönteminde, amino asit dizilimlerine rastgele sayılar verilmektedir. Tıpkı tam sayı protein haritalama tekniğinde olduğu gibi belirli bir bilgiye ihtiyaç duyulmamaktadır. Bu çalışmada 1'den 20'ye kadar olan sayılar belirlenmiş ve bu sayılar amino asit kodlarına rastgele bir şekilde atanmıştır. Tablo 8'de amino asit kodlarına atanmış rastgele sayılar verilmiştir.

Tablo 8. Amino asitlerin rastgele kodlama değerleri

Amino Asit Kodu	Sayısal İfade	Amino Asit Kodu	Sayısal İfade
A	9	L	11
R	7	K	4
N	15	M	12
D	6	F	17
C	19	P	14
Q	5	S	3
E	18	T	8
G	3	W	20
H	1	Y	13
I	10	V	2

Tablo 8'deki değerlere göre $S(n) = [ALKRNM]$ şeklinde bir protein dizisi $S(n) = [9\ 11\ 4\ 7\ 15\ 12]$ biçiminde ifade edilir.

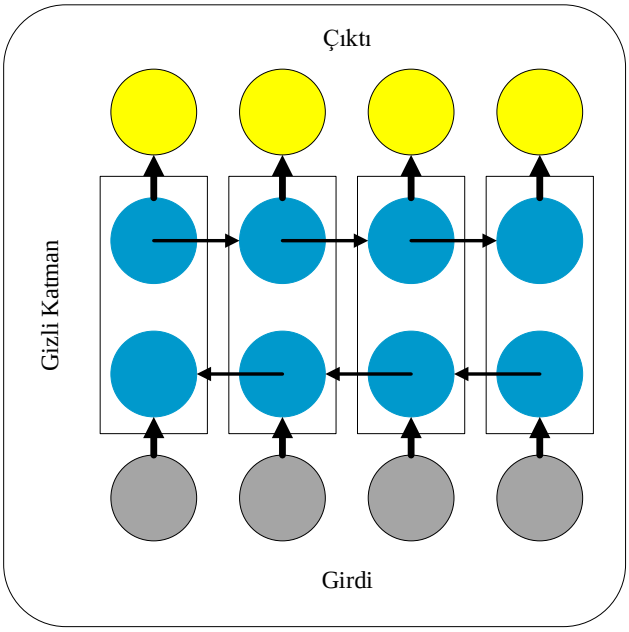
Protein dizilimleri sayısallaştırıldıktan sonra tekrarlayıcı sinir ağı modeli ile sınıflandırılmış ve kanserli genler tahmin edilmiştir. Tahmin işleminin ardından protein haritalama tekniklerinin başarımları doğruluk, F1-skor, kesinlik ve AUC değerlendirme ölçütleri ile belirlenmiştir. Şekil 1'de çalışmanın akış diyagramı verilmiştir.



Şekil 1. Çalışmanın akış diyagramı

4. Uygulama Sonuçları ve Tartışma

Bu kısımda çalışmada kullanılan çift-yönlü uzun-kısa süreli bellek ve parametreleri hakkında bilgiler verilmiştir. Bunlara ek olarak uygulama sonuçları irdelenmiş ve protein haritalama tekniklerinin kıyaslaması yapılarak tartışılmıştır. Derin öğrenme bir makine öğrenmesi türü olup, günümüzde etkin bir şekilde kullanılmaktadır. Günümüzde popüler olmasının en büyük nedenlerinden birisi verilerin artık daha hızlı ve kolay bir şekilde elde edilebilmesi ve bu verileri analiz edebilecek donanım ihtiyaçlarının karşılanabilmesidir [23]. Bunlara ek olarak karmaşık ve büyük veri setlerinde başarılı olması da bu anlamda bir etken olmaktadır. Derin öğrenmenin makine öğrenmesine göre en büyük avantajı özellik çıkarım aşamasındaki işlemlerdir. Makine öğrenmesi tabanlı yaklaşımlarda anahtar özellikler el ile çıkarılırken, derin öğrenmede adaptif bir yaklaşım bulunmaktadır [24]. Veri sayısının çok olması ya da veri setinin karmaşık olması özelliklerin el ile çıkarılmasını zorlaştırmakta ve hatta zaman almaktadır [25]. Bu gibi avantajlar derin öğrenmenin hemen hemen her alanda kullanılmasını sağlamıştır. Derin öğrenme ile biyomedikal [26,27], biyoenformatik [28], nesne tanıma [29], robotik [30], enerji [31] alanlarda çalışmalar bulunmaktadır. Bu çalışmada da derin öğrenme modellerinden birisi olan tekrarlayıcı sinir ağları kullanılmış ve çift-yönlü uzun-kısa süreli bellek (BiLSTM) ağı tasarlanmıştır. Uzun-kısa süreli bellek daha çok zaman ve metin serisi gibi seriler olmak üzere sıralı yapılarda etkili olmaktadır [32]. Uzun-kısa süreli bellek yapısında geçmişten gelen bilgiler kullanılmaktadır. Çift-yönlü uzun-kısa süreli bellek ise aynı anda iki uzun-kısa süreli belleğin eğitilmesi mantığına dayanmaktadır. Bu yapı sayesinde derin öğrenme ağı sadece geçmişteki bilgileri değil gelecek hakkında bilgileri de bünyesinde tutmaktadır. BiLSTM mimarisinde ileri ve geri yöndeki hesaplamalar aynı anda çalıştırılır ve iki yönde de yapılan hesaplamalar sonucu ulaşılan bilgiler birleştirilerek çıktı elde edilir. Bu nedenden dolayı iki yöndeki bilgilerin kullanılması sıralı verilerin ve zaman serilerinin işlenmesinde avantaj sağlamaktadır. İki adet LSTM ünitesinden oluşmaktadır. Bunlardan bir tanesi girdiyi alıp ileriye doğru işlerken, diğeri ise geriye doğru bilgiyi işlemektedir. Şekil 2'de BiLSTM mimarisinin genel bir yapısı verilmiştir.

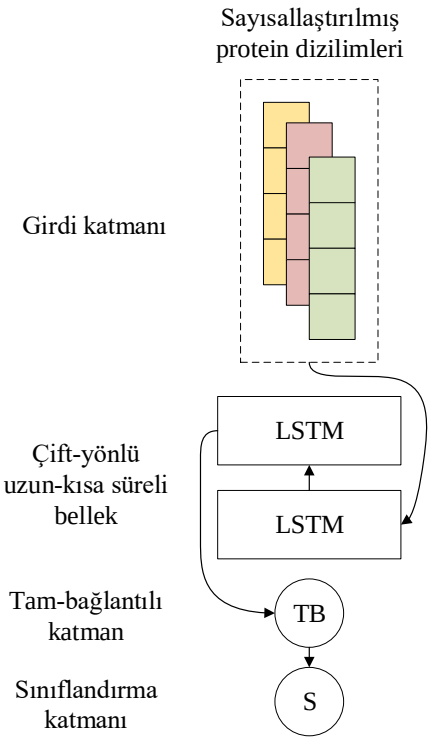


Şekil 2. BiLSTM yapısı

Bu çalışmada tasarlanmış olan BiLSTM ağının parametreleri şu şekilde özetlenebilir;

- Girdi katmanında her bir protein haritalama tekniği ile sayısallaştırılmış olan protein dizilimleri kullanılmıştır.
- İkinci katmanda BiLSTM kullanılmış ve bu amaç için 32 adet ünite değerlendirilmiştir. Aktivasyon fonksiyonu olarak SELU kullanılmıştır.
- Ardından çıktılar düzleştirme işlemine tabi tutulmuş ve veriler 1-boyutlu hale getirilmiştir.
- Daha sonra tüm verilerin aynı aralıkta olmasını sağlayabilmek için normalleştirme işlemi yapılmış ve yığın normalleştirilmesi kullanılmıştır.
- Son olarak bir adet tam-bağlantılı katman kullanılmış ve bu kısım için 128 adet nöron ele alınmıştır.
- Son katmanda ise sınıflandırma yapılmış ve toplamda her bir kanser türü için 2 adet durum (hastalıklı ve normal) olduğu için bu katmanda 2 adet nöron kullanılmıştır. Aktivasyon fonksiyonu olarak ise Softmax hesaplamasından yararlanılmıştır.
- Modelin kaybını belirleyebilmek için kategorik çapraz-entropi kullanılmıştır.
- En iyileme işlemi için SGD (Stokastik Gradyan İnişi) kullanılmıştır.
- Eğitim işlemi 250 iterasyon ile yapılmıştır.
- Modeli test etmek için ise eğitim ve test verileri ayrılmıştır. Eğitmek için tüm verilerin %80'i kullanılmış ve geri kalan %20'si ile de test işlemi yapılmıştır.
- Çalışmada belirlenmiş olan bütün parametreler deneme-yanılma yaklaşımı ile belirlenmiş ve en iyi sonuçları veren parametreler değerlendirilmiştir.

Şekil 3'te tasarlanmış olan BiLSTM ağ yapısı verilmiştir.



Şekil 3. Tasarlanan BiLSTM ağı

Sınıflandırma işleminin ardından protein haritalama fonksiyonlarının başarımları doğruluk, F1-skor, kesinlik ve AUC skorları ile belirlenmiştir. Tablo 9-11 arasında mesane, kolon ve prostat kanserine ait sınıflandırma işleminin her bir protein haritalama tekniğine göre sonuçları verilmiştir.

Tablo 9. Mesane kanserine göre protein haritalama tekniklerinin başarım ölçütleri

Protein Haritalama Yöntemleri	**Mesane**			
	Doğruluk (%)	**AUC (%)**	**F1-Skor (%)**	**Kesinlik (%)**
AESNN1	87.15	86	87.08	87.25
Hidrofobiklik	82.31	84	81.89	82.15
Tam sayı	83.84	84	83.95	84.02
Miyazawa Enerjileri	78.57	79	78.46	78.72
Rastgele kodlama	75.93	76	76.15	76.18

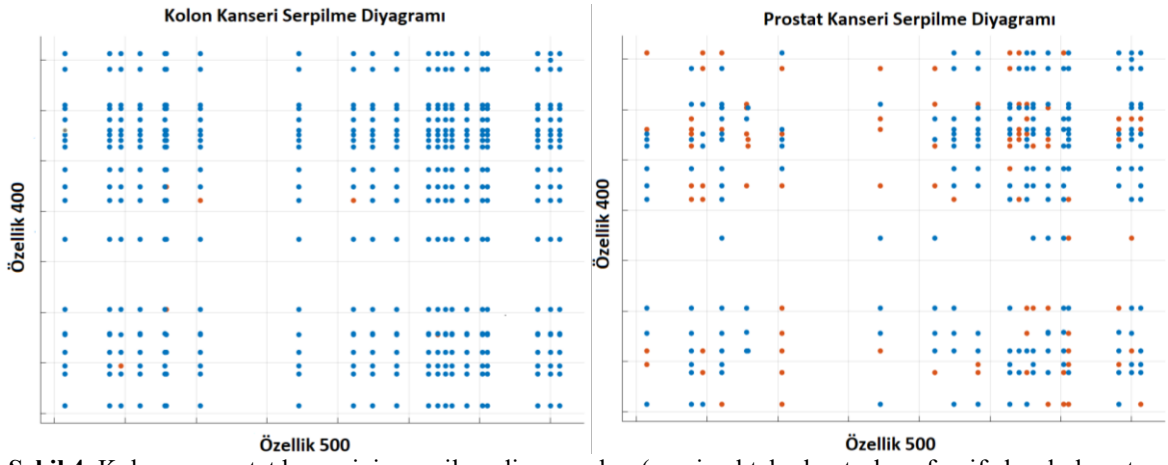
Tablo 10. Kolon kanserine göre protein haritalama tekniklerinin başarım ölçütleri

Protein Haritalama Yöntemleri	**Kolon**			
	Doğruluk(%)	**AUC(%)**	**F1-Skor(%)**	**Kesinlik(%)**
AESNN1	92.45	95	92.46	92.43
Hidrofobiklik	94.35	97	94.39	94.35
Tam sayı	93.13	99	93.06	93.12
Miyazawa Enerjileri	94.40	95	94.39	94.41
Rastgele kodlama	91.87	93	91.90	91.87

Tablo 11. Prostat kanserine göre protein haritalama tekniklerinin başarım ölçütleri

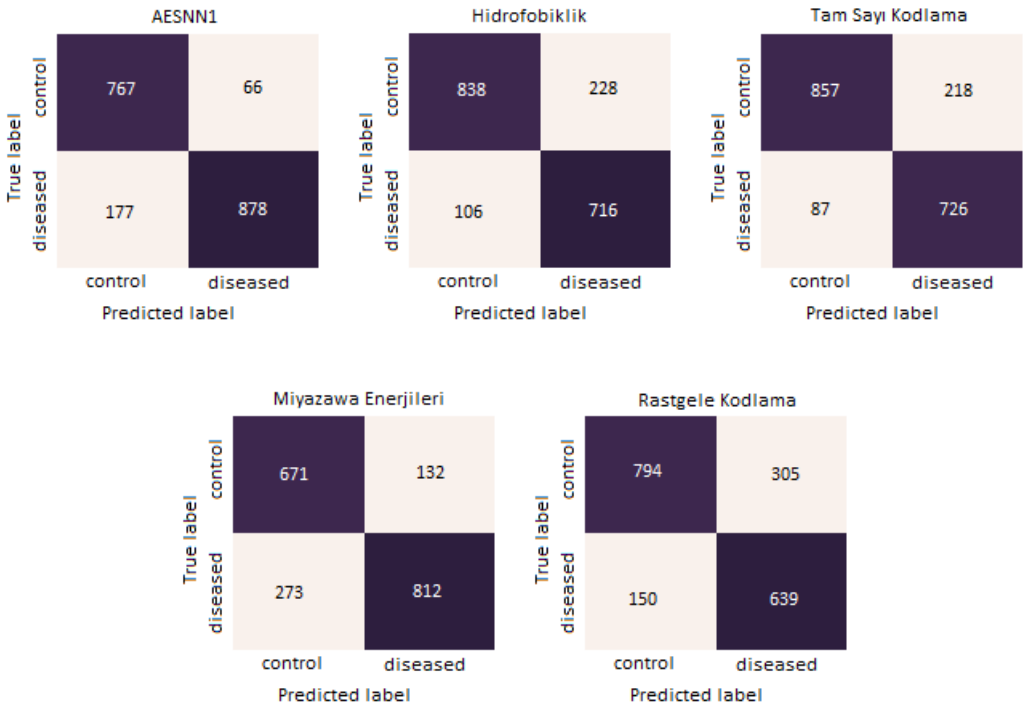
Protein Haritalama Yöntemleri	Prostat			
	Doğruluk(%)	AUC(%)	F1-Skor(%)	Kesinlik(%)
AESNN1	75.17	0.72	75.00	75.10
Hidrofobiklik	75.26	0.81	75.12	75.23
Tam sayı	75.45	0.70	75.46	75.28
Miyazawa Enerjileri	75.15	0.74	75.17	75.21
Rastgele kodlama	75.45	0.78	75.70	75.48

Tablo 9-11'deki sonuçlar incelendiğinde en iyi sonuçların kolon kanserinin sınıflandırılmasından elde edildiği görülmektedir. Tüm protein haritalama yöntemleri %90'ın üzerinde bir doğruluk, AUC, F1-skor ve kesinlik performansı göstermiştir. En iyi doğruluk skoru kolon kanseri için Miyazawa enerjileri ile elde edilmiş ve %94.40 oranında doğruluğa ulaşılmıştır. Ancak, en iyi AUC skoru ise tam sayı protein haritalama yöntemi ile elde edilmiş ve %99 oranında bir değer elde edilmiştir. Bunun yanı sıra, en başarısız olan protein haritalama yöntemi rastgele kodlamadır. Bu yöntem ile en düşük doğruluk ve AUC skorları elde edilmiştir. Sonuçlar genel olarak incelendiğinde ikinci en iyi sonuçlar mesane kanseri için elde edilmiştir. Kolon kanserindeki sonuçların aksine, mesane kanserinin sınıflandırılmasında en başarılı yöntemin AESNN1 olduğu görülmektedir. Bu yöntem ile %87.15 oranında doğruluk ve %86 oranında AUC skoru elde edilmiştir. Tıpkı kolon kanserinde olduğu gibi, en etkisiz sınıflandırma işlemini rastgele kodlama yöntemi yapmış ve sırasıyla %75.93 ve %76 oranında doğruluk ve AUC skoruna ulaşmıştır. Kolon ve mesane kanserinin aksine, en kötü başarımlar prostat kanserinde elde edilmiştir. Bu kanser türünde tüm protein haritalama teknikleri etkisiz olmuştur. Prostat kanserinin sınıflandırılmasında en başarılı olan yöntemler tam sayı ve rastgele yöntemleridir. Bu yöntemlerin doğruluk skoru %75.45 olarak hesaplanmıştır. En iyi AUC skoru ise hidrofobiklik yöntemi ile elde edilmiştir. Mesane kanserinin ayrımında en iyi sonucun AESNN1 yöntemi ile elde edilmesinin nedeni bu yöntemin hem makine öğrenmesi hem de yapısal tabanlı olmasından kaynaklanıyor olabilir. Çünkü mesane kanserinin, proteinlerin yapısı ile bağlantılı olduğu bilinmektedir [33]. Miyazawa enerjileri de yapısal-tabanlı bir yöntem olup AESNN1 kadar başarılı olmamıştır. Bu durumun nedeni, bu yöntemin yapısal bilgiden ziyade protein dizilimlerindeki temas enerjilerinden bilgi elde ediyor olması olabilir. Kolon kanseri için bütün proteinler başarılı sonuçlar üretmiş ve tüm yöntemler %90'ın üzerinde doğruluk ve AUC skorlarına ulaşmıştır. Bunun tam tersi bir çıkarım prostat için yapılabilir. Prostat kanserinin ayrımında tüm yöntemlerin etkisiz ve doğruluk skorlarının %80'in altında olduğu görülmektedir. En başarılı sonuçların kolon kanserinden ve en başarısız sonuçların prostat kanserinden elde edilmesinin nedeni verilerin dağılımdan kaynaklanıyor olabilir. Şekil 4'te kolon ve prostat kanserinin veri dağılımını gösteren serpilme diyagramı verilmiştir.

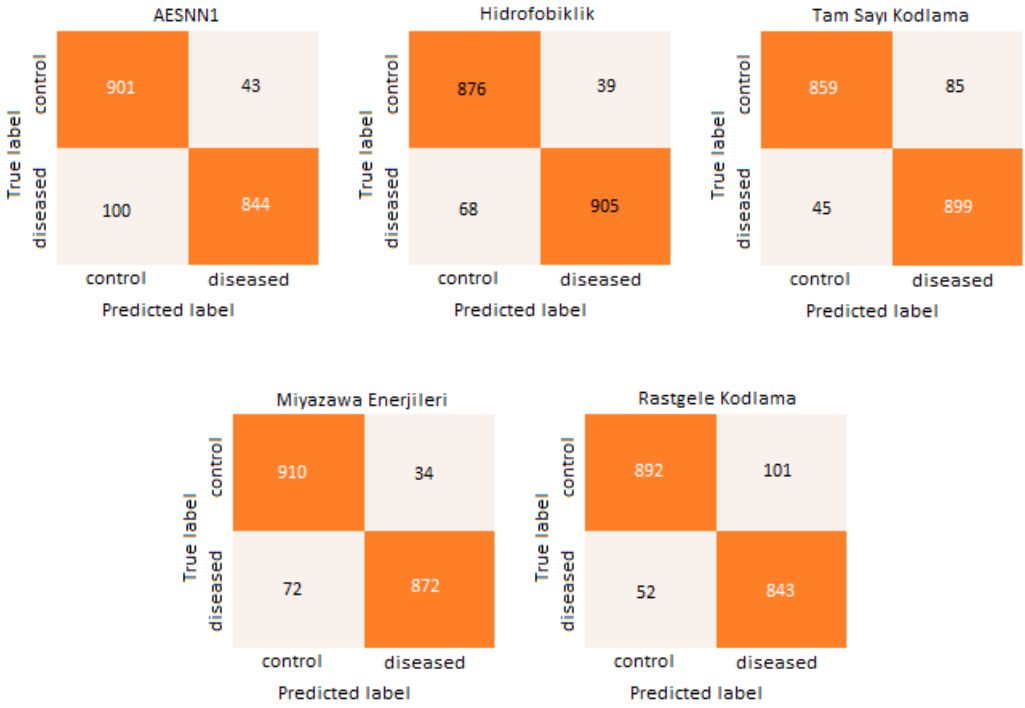


Şekil 4. Kolon ve prostat kanserinin serpilme diyagramları (mavi noktalar kontrol sınıfını ifade ederken, turuncu noktalar hastalıklı sınıfı ifade etmektedir)

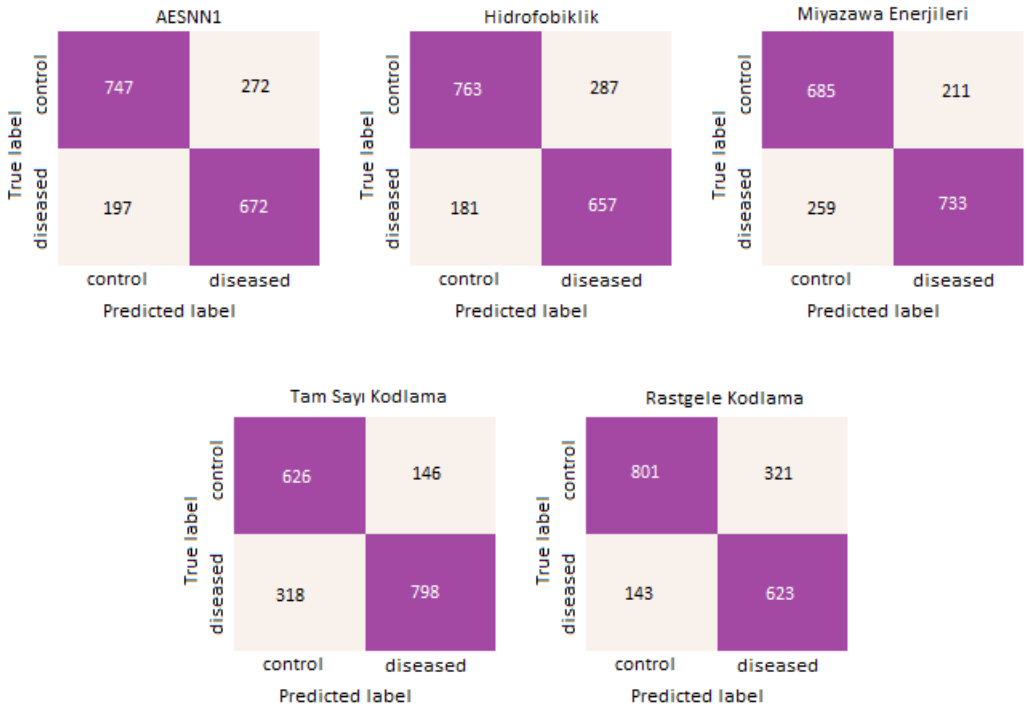
Şekil 4'ten de anlaşılacağı üzere kolon kanserinin dağılımı prostat kanserinin veri dağılımına göre daha ayrıktır. Bu da sınıflandırma işleminin daha yüksek olmasına ve sınıflar arasındaki ayrımın daha net yapılabilmesini sağlamaktadır. Tablo 9'da verilen genel sonuçlar bu durumu onaylamaktadır. Şekil 5-7'de protein haritalama tekniklerinin her bir kanser türü için göstermiş olduğu karmaşıklık matris değerleri verilmiştir.



Şekil 5: Mesane kanserine göre protein haritalama tekniklerinin karmaşıklık matrisleri

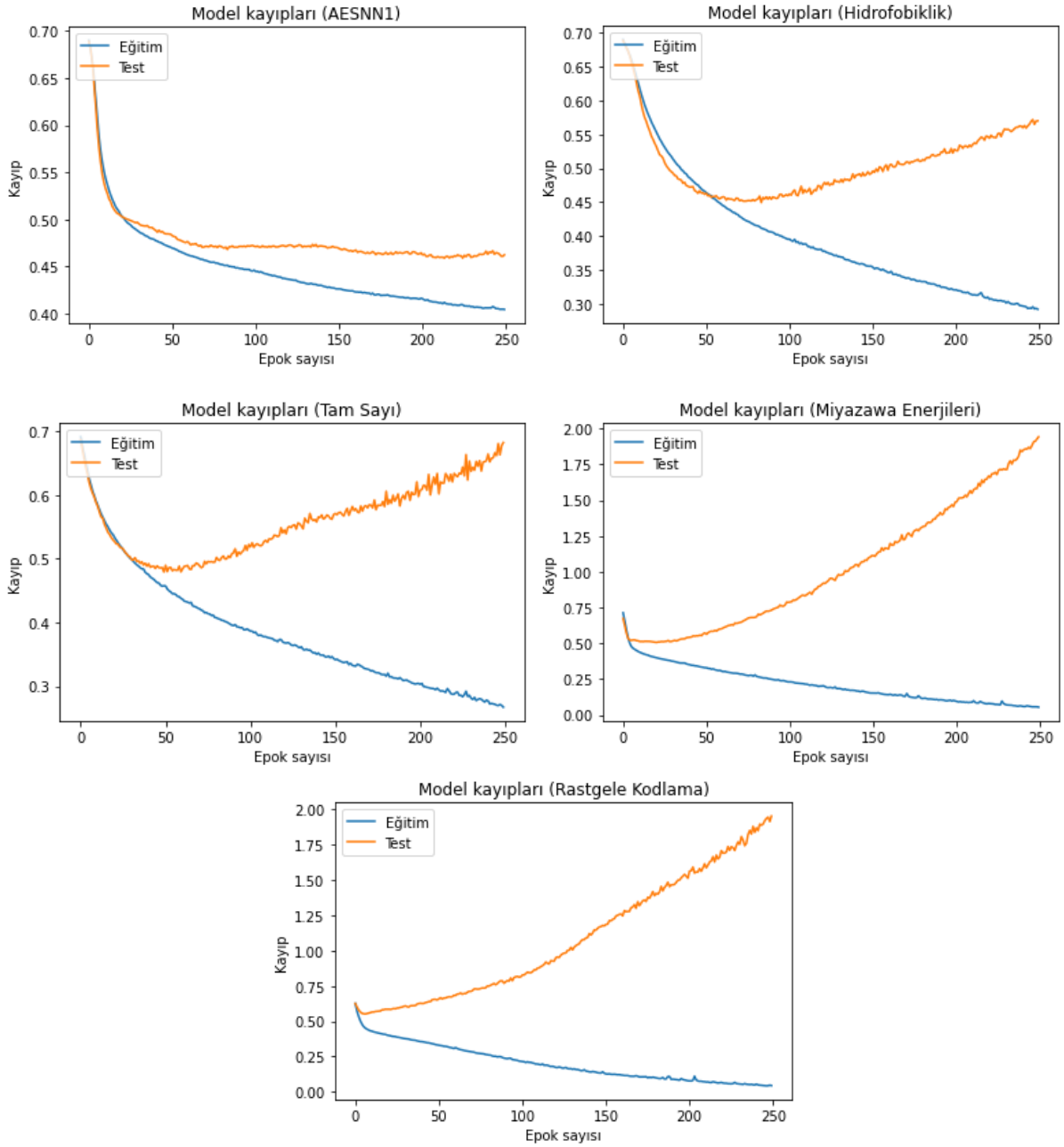


Şekil 6: Kolon kanserine göre protein haritalama tekniklerinin karmaşıklık matrisleri

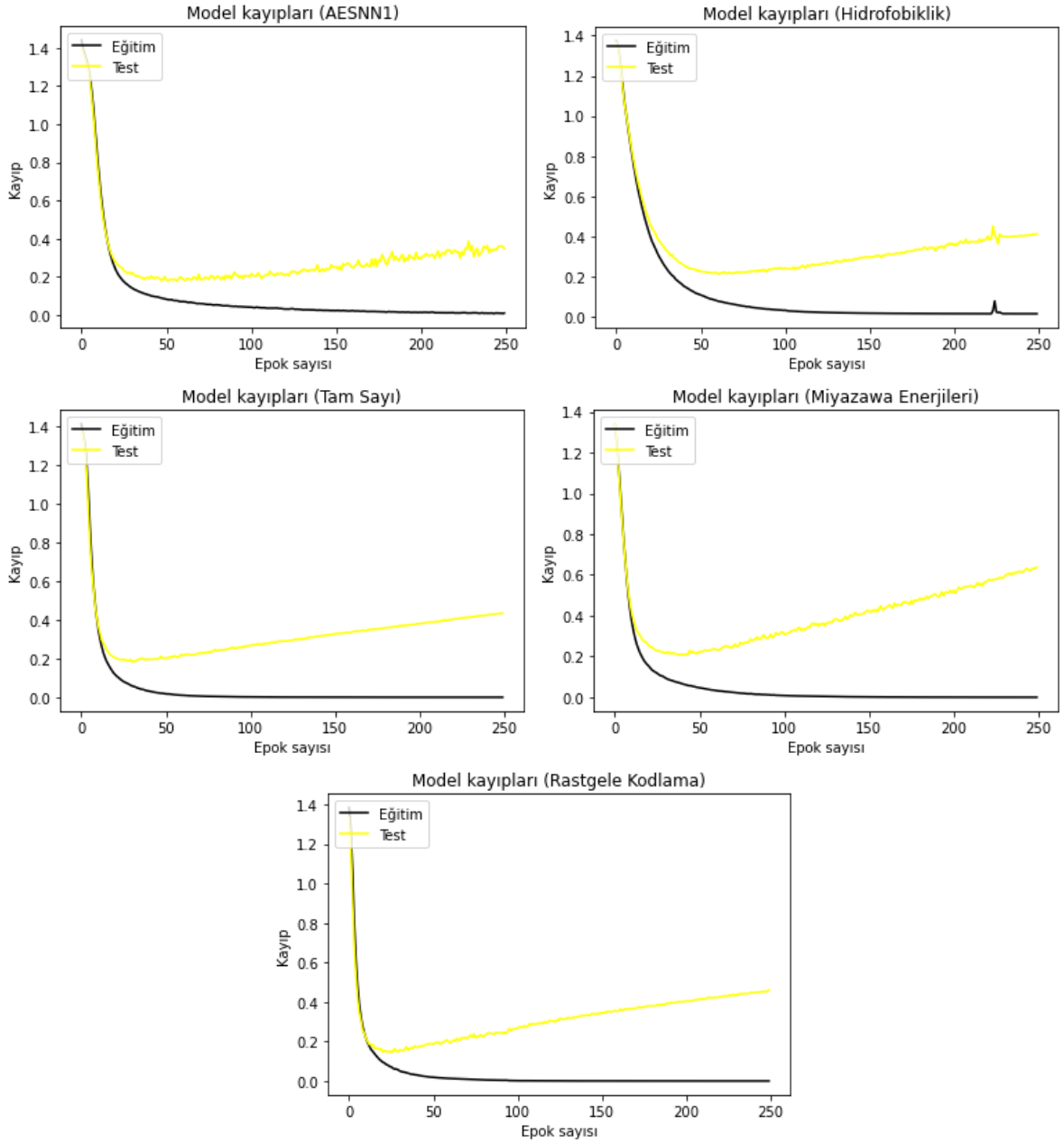


Şekil 7: Prostat kanserine göre protein haritalama tekniklerinin karmaşıklık matrisleri

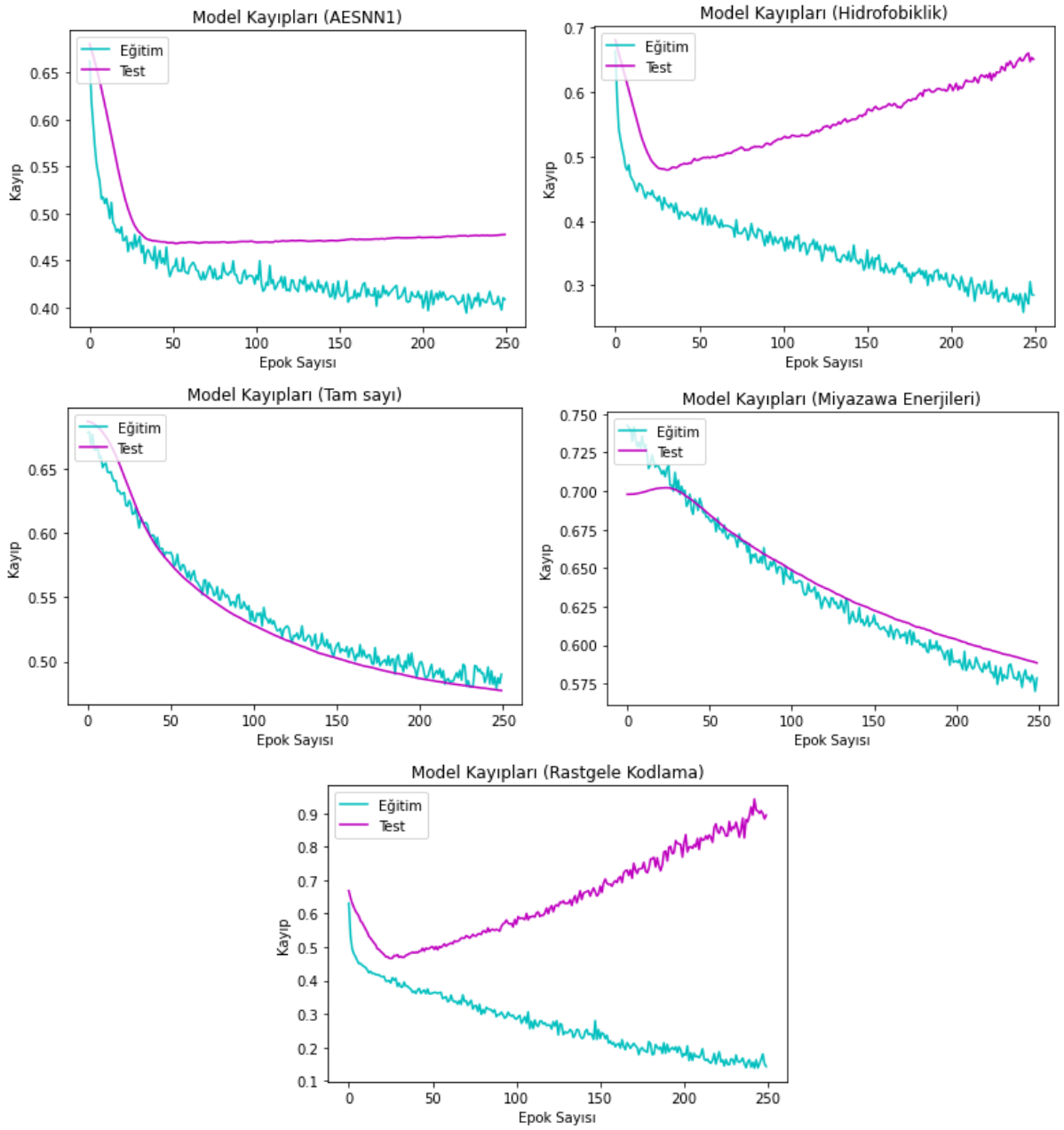
Şekil 8-10'da protein haritalama tekniklerinin kanser türleri için göstermiş olduğu eğitim ve test kayıpları verilmiştir.



Şekil 8: Mesane kanserine göre protein haritalama tekniklerinin kayıp grafikleri



Şekil 9: Kolon kanserine göre protein haritalama tekniklerinin kayıp grafikleri



Şekil 10: Prostat kanserine göre protein haritalama tekniklerinin kayıp grafikleri

Çalışmanın avantajları şu şekilde ifade edilebilir;

- Bu çalışma ile protein haritalama tekniklerinin kanser içeren proteinleri belirlemede başarılı olduğu gözlemlenmiştir.
- Uygulama sonuçlarına göre hesaplama tabanlı yaklaşımların bu alanda kullanılabileceği belirlenmiştir. Literatürde bulunan mevcut çalışmalar bu sonucu desteklemektedir.
- Protein haritalama tekniklerine ek olarak, derin öğrenme algoritmalarının çoğu biyoenformatik çalışmalarında da olduğu gibi, bu alanda da etkili ve başarılı olmuştur.

Çalışmanın dezavantajları şu şekilde sıralanabilir;

- Literatürdeki diğer çalışmalarla kıyaslandığında, bu çalışmada kullanılan veriler daha azdır. Bu durumun nedeni donanım ihtiyacımızın yetersiz olmasından kaynaklanmaktadır. Veri sayısının arttırılmasıyla bu sonuçlarda iyileşme olabilir ya da daha kötü performans sonuçları elde edilebilir. Bu durumun ileriki çalışmalarda değerlendirilmesi gerekmektedir.
- Farklı bir derin öğrenme algoritması ile sonuçlar farklı bir şekilde değerlendirilebilir. Çalışmanın çeşitliliğini ve başarımını koruması için, diğer derin öğrenme algoritmaları ile de değerlendirilmesi gerekmektedir.
- Bu çalışmada kullanılan protein haritalama teknikleri dışında farklı protein haritalama tekniklerinin kullanılması çalışmanın doğruluğu ve güvenilirliği açısından önem arz etmektedir. Diğer protein haritalama teknikeri ile de çalışmanın değerlendirilmesi bu bakımdan önem arz etmektedir.

Önerilen çalışmanın literatüre katkıları aşağıda maddeler halinde verilmiştir;

- Bilgisayar tabanlı bir yaklaşımın kullanılmasıyla, gen ifadelerine dayalı işlemler hizalama programlarına göre daha hızlı gerçekleştiği için, kanserli genlerin belirlenmesi ve tahmin edilmesi daha hızlı olmaktadır. Bu da kanser gibi önemli hastalıkların teşhisinde ve doğru bir tedavinin uygulanmasında büyük önem arz etmektedir.
- Çalışmada kullanılan protein haritalama tekniklerinin etkili olduğu gözlemlenmiştir. Bu sayede protein haritalama tekniklerinin bu alanda uygulanabilmesinin önü açılmıştır. Farklı türde protein haritalama teknikleri de kullanılarak bu başarı pekiştirilebilir.
- Yeni nesil hizalama haritalarındaki dizilim sorunları protein haritalama teknikleri ile çözülebilmektedir. Bu durum protein haritalama tekniklerinin daha sık kullanılabilmesine yol açabilecektir.

5. Sonuç

Bu çalışmada kanser içeren proteinlerin derin öğrenme ile analizi yapılarak, protein haritalama tekniklerinin performansları karşılaştırılmıştır. Mesane, kolon ve prostat kanserine ait veriler değerlendirilmiştir. Her bir kanser türü için farklı gen ifadeleri elde edilmiş ve her biri hem kontrol hem de hastalıklı olmak üzere iki ayrı şekilde etiketlendirilmiştir. Çalışmada AESNN1, hidrofobiklik, tam sayı, rastgele kodlama ve Miyazawa enerjileri olmak üzere beş farklı protein haritalama teknikleri ele alınmıştır. Derin öğrenme modeli olarak tekrarlayıcı sinir ağı modeli olan çift-yönlü uzun-kısa süreli bellek tasarlanmış ve protein haritalama tekniklerinin başarımları, doğruluk ve AUC değerlendirme ölçütleri belirlenmiştir. Mesane kanseri için en iyi doğruluk ve AUC skoru AESNN1 yöntemi ile elde edilmiş ve sırasıyla %87.15 ve %86 gibi değerlere ulaşılmıştır. Kolon kanseri, üç kanser türü içerisinde en etkili sınıflandırılan kanser olmuştur. Bu kanser türünde bütün protein haritalama teknikleri %90'ın üzerinde doğruluk ve AUC skorlarına erişmiştir. Kolon kanserinde en iyi doğruluk skor Miyazawa enerjileri yöntemi ile hesaplanmış ve sonuç %94.40 olarak bulunmuştur. Bunun yanı sıra en iyi AUC skoru ise tam sayı yöntemi ile elde edilmiştir. Protein haritalama yöntemlerinin en etkisiz olduğu kanser türü prostat olmuştur. Prostat kanserinin ayrımında bütün protein haritalama teknikleri %80'in altında bir başarım sergilemişlerdir. Prostat kanseri için tam sayı ve rastgele kodlama en iyi doğruluk skoruna erişmiştir. Bunlara ek olarak en yüksek AUC skoru ise hidrofobiklik ile hesaplanmıştır. Genel sonuçlardan da anlaşılacağı üzere, protein haritalama teknikleri ve yapay öğrenme içeren hibrit bir modelin kanser içeren proteinleri belirlemede etkili bir yöntem olduğu ve haritalama tekniklerine göre değişiklik gösterdiği belirlenmiştir. Bu durumda doğru protein haritalama tekniğini seçmek önem arz etmektedir. İleride gerçekleştirilecek çalışmalarda, kanser içeren proteinler, protein-protein etkileşimlerinden ve etkileşim ağından yararlanılarak analiz edilebilir ve bu çalışmadaki sonuçlar ile karşılaştırılabilir.

Kaynaklar

[1] Sun Y, Sitao Z, Ma K, Liu W, Yue Y, Hu G, Lu H, Chen W. Identification of 12 cancer types through genome deep learning. Scientific Reports 2019; 9: 1-9.
[2] Baykara O. Kanser tedavisinde güncel yaklaşımlar. Balıkesir Sağlık Bilimleri Dergisi 2016; 5(3): 155-165.
[3] Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Informatics 2006; 3: 59-78.

[4] Dhahri H, Maghayreh EA, Mahmood A, Elkilani W, Nagi MF. Automated breast cancer diagnosis based on machine learning algorithms. Journal of Healthcare Engineering 2019; 1-12.
[5] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis, DI. Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal 2015; 13: 8-17.
[6] Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ. Deep learning for health informatics. IEEE Journal of Biomedical and Health Informatics 2017; 21(1): 4-21.
[7] Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nature Reviews Genetics 2015; 16: 321-322.
[8] Gevaert O, Smet FD, Timmerman D, Moreau Y, Moor BD. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. Bioinformatics 2006; 22(14): 184-190.
[9] Jurtz VI, Johansen AR, Nielsen M, Armenteros JJA, Nielsen H, Sonderby CK, Winther O, Sonderby SK. An introduction to deep learning on biological sequence data: examples and solutions. Bioinformatics 2015; 33(22): 3685-3690.
[10] Chen KH, Wang TF, Hu YJ. Protein-protein interaction prediction using a hybrid feature representation and a stacked generalization scheme. BMC Bioinformatics 2019; 20: 1-17.
[11] Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y, Lu H. Deep-learning-based drug-target interaction prediction. Journal of Proteome Research 2017; 16(4): 1401-1409.
[12] Seo S, Oh M, Pak Y, Kim S. DeepFam: deep learning based alignment-free method for protein family modeling and prediction. Bioinformatics 2018; 34(13): 254-262.
[13] Lv Z, Ao C, Zou Q. Protein function prediction: from tradition classifier to deep learning. Proteomics 2019; 19(14): 1-3.
[14] Wang S,Peng J, Ma J, Xu J. Protein secondary structure prediction using deep convolutional neural fields. Scientific Reports 2016; 6: 1-11.
[15] Guda P, Chittur SV, Guda C. Comparative analysis of protein-protein interactions in cancer-associated genes. Genomics, Proteomics & Bioinformatics 2009; 7(1-2): 25-36.
[16] Huang CH, Peng HS, Ng KL. Prediction of cancer proteins by integrating protein interaction, domain frequency, and domain interaction data using machine learning algorithms. BioMed Research Journal 2015; 2015: 1-15.
[17] Matsubara T, Ochiai T, Hayashida M, Akutsu T, Nacher JC. Convolutional neural network approach to lung cancer classification integrating protein interaction network and gene expression profiles. Journal of Bioinformatics and Computational Biology 2019; 17(3): 1-11.
[18] Hosseinzadeh F, KayvanJoo AH, Ebrahimi M, Goliaei B. Prediction of lung tumor types based on protein attributes by machine learning algorithms. Springer Plus 2013; 2(238): 1-14.
[19] Chang JW, Ding Y, Qamar MT, Shen Y, Gao J, Chen LL. A deep learning model based on sparse auto-encoder for prioritizing cancer-related genes and drug-target combinations. Carcinogenesis 2019; 40(5): 624-632.
[20] Lin K, May ACW, Taylor WR. Amino acid encoding schemes from protein structure alignments: multi-dimensional vectors to describe residue types. Journal of Theoretical Biology 2002; 216(3): 361-365.
[21] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 1982; 157(1): 105-132.
[22] Miyazawa S, Jernigan RL. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 1985; 18(3): 534-552.
[23] Goodfellow I, Bengio Y, Courville A. Derin Öğrenme, 1. Basım, Eryaman, Ankara: Buzdağı Yayınevi, 2018.
[24] Santur Y. Derin öğrenme ve aşağı örnekleme yaklaşımları kullanılarak duygu sınıflandırma performansının iyileştirilmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi 2020; 32(2): 561-570.
[25] Alakus TB, Turkoglu I. Prediction of protein-protein interaction with LSTM deep learning model. In: 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies; 11-13 October 2019; Ankara Turkey: IEEE. Pp. 1-5.
[26] Budak Ü. SegNet mimarisi ile bilgisayarlı tomografi görüntülerinden karaciğer bölgesinin bölütlenmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi 2019; 31(1): 215-222.
[27] Toraman S, Türkoğlu İ. Derin öğrenme ile FTIR sinyallerinden kolon kanseri riskinin belirlenmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi 2018; 30(2): 115-120.
[28] Das B, Turkoglu I. A novel numerical mapping method based on Entropy for digitazing DNA sequences. Neural Computing and Applications 2018; 29: 207-215.
[29] Daş R, Polat B, Tuna G. Derin öğrenme ile resim ve videolarda nesnelerin tanınması ve takibi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi 2019; 31(2): 571-581.
[30] Bingöl MS, Kaymak Ç, Uçar A. Derin öğrenme kullanarak otonom araçların insan sürüşünden öğrenmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi 2019; 31(1): 177-185.
[31] Çayır A, Yenidoğan I, Dağ H. Konutların günlükelektrik güç tüketimi tahmini için uygun model seçimi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi 2018; 30(3): 15-21.
[32] Alpay Ö. LSTM mimarisi kullanarak USD/TRY fiyat tahmini. Avrupa Bilim ve Teknoloji Dergisi 2020; Özel sayı: 452-456.
[33] Brunner A, Tzankov A. The role of structural extracellular matrix proteins in urothelial bladder cancer (review). Biomark Insights 2007; 2: 418-427.
[34] Jing X, Dong Q, Hong D, Lu R. Amino acid encoding methods for protein sequences: a comprehensive review and assessment. IEEE/ACM Transactions on Computational Biology and Bioinformatics 2020; 17: 1918-1931.

[35] Guia JM, Devaraj M, Leung CK. DeepGx: deep learning using gene expression for cancer classification. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 27-30 August 2019; Vancouver, Canada. pp. 914-920.
[36] Gao F, Wang W, Tan M, Zhu L, Zhang Y, Fessler E, Vermeulen L, Wang X. DeepCC: a novel deep learning-based framework for cancer molecular subtype classification. Oncogenesis 2019;8: 1-12.
[37] He B, Bergenstrahle L, Stenbeck L, Abid A, Andersson A, Bork A, Maaskola J, Lundeberg J, Zou J. Integrating spatial gene expression and breast tumour morphology via deep learning. Nature Biomedical Engineering 2020; 4: 827-834.
[38] Chen R, Yang L, Goodison S, Sun Y. Deep-learning approach to identifying cancer subtypes using high-dimensional genomic data. Bioinformatics 2020; 36: 1476-1483.