YALOVA ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

YARDIMCI T HÜCRELERİ / BÜYÜK DOKU UYGUNLUK KOMPLEKSİ MOLEKÜLLERİ BAĞLANMA YERLERİNİN TESPİTİNDE ÖZNİTELİK KODLAMA YÖNTEMLERİ GELİŞTİRİLMESİ

YÜKSEK LİSANS TEZİ

İlknur ÇINAR EFE

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

HAZİRAN 2016

YALOVA ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

YARDIMCI T HÜCRELERİ / BÜYÜK DOKU UYGUNLUK KOMPLEKSİ MOLEKÜLLERİ BAĞLANMA YERLERİNİN TESPİTİNDE ÖZNİTELİK KODLAMA YÖNTEMLERİ GELİŞTİRİLMESİ

YÜKSEK LİSANS TEZİ

İlknur ÇINAR EFE
115105018

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Doç.Dr.Murat GÖK

HAZİRAN 2016

YALOVA ÜniversitesiFen Bilimleri Enstitüsü'nün 115105018 numaralı Yüksek Lisans Öğrencisi **İlknur ÇINAR EFE**, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı **"YARDIMCI T HÜCRELERİ / BÜYÜK DOKU UYGUNLUK KOMPLEKSİ MOLEKÜLLERİ BAĞLANMA YERLERİNİN TESPİTİNDE ÖZNİTELİK KODLAMA YÖNTEMLERİ GELİŞTİRİLMESİ "** başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Doç.Dr. Murat GÖK** .....
Yalova Üniversitesi

Jüri Üyeleri : **Doç.Dr.Murat GÖK** .....
Yalova Üniversitesi

Doç.Dr. Müfit ÇETİN .....
Yalova Üniversitesi

Doç.Dr.Murat ÇAKIROĞLU .....
Sakarya Üniversitesi

Teslim Tarihi : 02.05.2016

Savunma Tarihi : 02.06.2016

Oğlum Ömer Çınar'a,

ÖNSÖZ

Tez konumun belirlenmesinde ve yürütülmesinde ki tüm aşamalarda yanımda olan, yüksek lisans eğitimim boyunca ilminden faydalandığım, insani ve ahlaki değerleri ile de örnek edindiğim, yanında çalışmaktan onur duyduğum tez danışmanım, Saygıdeğer Hocam Sayın Doç. Dr. Murat GÖK'e teşekkürlerimi sunarım.

Bütün hayatım boyunca yanımda olan, bana güvenen, yoğun zamanlı çalışmalarımı anlayışla karşılayan, maddi ve manevi desteklerini hep hissettiğim, başta annem ve babam Nermin-Nusrettin ÇINAR'a, abime ve kardeşime, en zor anlarımda yardımlarını hiçbir zaman esirgemeyen arkadaşım Uğur TURHAL'a, üzerimde emeği olan bütün aileme ve arkadaşlarıma sonsuz teşekkürlerimi sunarım.

Mayıs 2016

İlknur ÇINAR EFE
(Bilgisayar ve Kontrol Öğretmeni)

İÇİNDEKİLER

Sayfa

KISALTMALAR

BDUK : Büyük Doku Uygunluk Kompleksi
IEDB : Immune Epitope Database
İLA : İnsan Lökosit Antijeni
NK : Doğal Öldürücüler
THR : T Hücre Reseptörü
ASH : Antijen Sunan Hücreler
Th : T yardımcı hücreler
Tc : Öldürücü T hücreleri
BKY : Birimdik Kodlama Yöntemi
AKKY : Ağırlık ve Konum Öznitelik Kodlama Yöntemi
BloFTKY : Blosum 50-Fizikokimyasal Özellik Tabanlı Öznitelik Kodlama Yöntemi
BloAKKY : Blosum 50-Ağırlık ve Konum Tabanlı Öznitelik Kodlama Yöntemi
ÇDT : Çapraz Doğrulama Tekniği
DVM : Destek Vektör Makineleri
RO : Rastgele Orman
OOB : Out Of Bag
DA : Doğru Artı
YA : Yanlış Artı
YE : Yanlış Eksi
DE : Doğru Eksi
Na : Not Applicable
NuN : Not a Number
nM : Nano Molar
OV : Orta Veriseti
GV : Güçlü Veriseti
TV : Tam Veriseti
MKK : Matthews Korelasyon Katsayısı
$\mathbf{a_i}$: Ağırlık
$\mathbf{k_i}$: Konum

ÇİZELGE LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

YARDIMCI T HÜCRELERİ / BÜYÜK DOKU UYGUNLUK KOMPLEKSİ MOLEKÜLLERİ BAĞLANMA YERLERİNİN TESPİTİNDE ÖZNİTELİK KODLAMA YÖNTEMLERİ GELİŞTİRİLMESİ

ÖZET

T hücreleri, bağışıklık yanıtı oluşumunda önemli bir yere sahiptir. İnsan vücudu her an dışarıdan gelen ve sürekli değişiklik gösteren çok sayıda mikroorganizma ile karşı karşıya kalır. Bağışıklık sisteminin harekete geçmesine zararlı mikroorganizmalara ait olan antijen proteinleri neden olur. Uzun antijen proteinleri antijen sunan hücreler tarafından T hücreleri ile birleşebilmeleri için daha küçük peptit parçacıklarına ayrılırlar. Bu antijenik peptitlere Epitop adı verilir.

Doğada bulunan protein yapısındaki antijenlerin her bir peptitine farklı bir T hücre klonu olduğu kabul edilir. T hücrelerinin bir antijeni tanıyıp reaksiyon oluşturabilmesi için bu antijenin bazı hücreler tarafından işlenmesi ve yüzey molekülleri aracılığıyla kendilerine sunulması gerekmektedir. T hücrelerine antijen sunumunu sağlayan hücre yüzeyindeki moleküllere Büyük Doku Uygunluk Kompleksi (BDUK) denir. Büyük doku uygunluk kompleksi (BDUK) molekülünün bağlanması T hücre aktivasyonuna yol açar ve bir seri biyokimyasal reaksiyonu tetikler.

Bağışıklık sistemi hareketinde bir aşı ya da ilaç geliştirmek için T hücre epitoplarının önceden tahmin edilmesi büyük önem taşır. Sürekli mutasyona uğrayarak çeşitlilik gösteren sayısı binleri aşan antijenik peptitlerin tanımlanması laboratuvar ortamında zaman ve maliyet açısından uygun değildir. Bu nedenle bilgisayar ortamında makine öğrenmesi algoritmaları ile çözüm aramak daha uygundur. Bu tez çalışmasında amacımız epitopları kestirmek için yeni makine öğrenmesi öznitelik kodlama teknikleri geliştirmektir.

Yardımcı T hücreleri / BDUK molekülleri bağlanma özgünlüklerinin tanımlanmasında Bağışıklık Epitop Veritabanı (IEDB)'den insan lökosit antijeni (İLA-A, İLA-B) peptit verileri kullanılmıştır. Veri seti dokuz amino asit uzunluğunda peptitlerden oluşmaktadır. Yardımcı T hücreleri / BDUK özgünlüğünü tespit etmek için iki tane öznitelik kodlama yöntemi geliştirildi. Birinci yöntemde Blosum 50 yer değiştirme matrisi ile amino asitlerin fizikokimyasal özellikleri kullanılmıştır. İkinci yöntemde ise amino asitlerin ağırlık ve konum bilgileri ile Blosum 50 yer değiştirme matrisi kullanılmıştır.

Sınıflandırma testleri Weka yazılımı ortamında 10-Kat Çapraz Doğrulama Test Tekniğine göre gerçekleştirilmiştir. Tez kapsamında yapılan deneysel çalışmalarda sınıf doğruluğu, duyarlık, özgünlük ve MKK (Matthews Korelasyon Katsayısı) performans metrikleri elde edilmiştir.

DEVELOPING T HELPER CELLS / MAJOR HISTOCOMPATIBILITY COMPLEX MOLECULES FEATURE ENCODING METHOTS IN DETECTION OF BINDING SITES.

SUMMARY

T cells, has an important role in the formation of immune response. The human body constantly faces with from the outside and ever changing a large number of microorganisms. Antigen proteins having the harmful microorganisms cause the activation of the immune system. Long antigen proteins are divided into smaller peptide fragments by antigen presenting cells to combine with T cells. This antigenic peptides are defined Epitope.

Different T cell clones to each peptide is considered to be an existing in nature of antigens on protein structure. Another important feature of the T cells must be processed to recognize and create reaction the antigen formed by some cells and offered them via surface molecules. The major histocompatibility complex is called enabling T cells to molecules on the cell surface antigen presentation. Connecting the major histocompatibility complex molecule leads to T cell activation and triggers a series of biochemical reactions.

The prediction of T cell epitopes identification is important to develop a vaccine or drug on the immune system. Identification antigenic peptides in excess of one thousand varied undergoes sustained mutations is not appropriate in terms of time and cost in the laboratory. Therefore, it is more appropriate to seek solutions with computerized machine learning algorithms. Develop a new machine learning techniques to predict attributes encoding epitopes is the purpose of fhis thesis.

IEDB database of human leukocyte antigen (ILA-A, ILA-B) peptide data was used for the identification T helper cells / BDUK molecule binding specificity. The data set consists of peptides nine amino acids in length. Two attribute encoding methods was developed to detect T helper cells / BDUK molecule originality. In the first method the physicochemical properties of the amino acid substitution matrix with Blosum 50 was used. In the second method the weight and the position information of amino acids and Blosum 50 substitution matrix was used.

Classification tests were carried out according to the 10-fold cross-validation test technique with Weka software environment. Experimental studies in the thesis were obtained class accuracy, sensitivity, specificity and Matthews Correlation Coefficient (MKK) of performance metrics.

1. GİRİŞ

Bağışıklık sistemi enfeksiyon yapan mikroorganizmalara karşı gelişen fizyolojik bir tepkidir. Bağışıklık sistemi hem yabancı etkenlere karşı organizmayı korur hem de organizmanın yabancı algılanmaması için gerekli toleransı sağlar. Canlıların bağışıklık sistemlerini uyaran ve canlı için yabancı olan tüm moleküllere "antijen" veya "immunojen", vücudumuzun bunlara karşı yok etme amaçlı ürettiği maddelere ise 'antikor' denir. Her antikor antijene özel üretilir. Bazı küçük antijenler immünojenik oldukları halde kendileri antikorlara bağlanamazlar, bağışık yanıt oluşturabilmek için kendilerini bağlayacak bir taşıyıcıya ihtiyaç duyarlar. Bu ihtiyacı Büyük Doku Uygunluk Kompleksi (BDUK) molekülleri karşılar. BDUK'un temel görevi peptit bağlanması ve bunların T hücrelerine sunulmasıdır. BDUK molekülünün bağlanması T hücre aktivasyonuna yol açar. T hücre aktivasyonuyla B hücreleri uyarılır ve antikor salınımı başlar [1].

T/MHC moleküllerinin bağlanma özgünlüklerini tespit etmek için IEDB veri tabanından insan lökosit antijeni (İLA-A, İLA-B) peptitleri kullanıldı. Bu çalışma da T hücre aktivasyonu tahmini için bilgisayar ortamında iki yöntem geliştirilmiştir. Birinci yöntemde Blosum 50 yer değiştirme matrisi ile amino asitlerin fiziko kimyasal özellikleri, ikinci yöntemde ise amino asitlerin ağırlık ve konum bilgileri kullanılmıştır. Blosum 50 yer değiştirme matrisindeki sayılar, amino asitlerin birbirleri yerine geçme eğilimlerine dair bilgi verir [2].

Yer değiştirme matrislerinde amino grubunun birbirlerine benzerlik oranı, satır ve sütunun kesişimi ile elde edilir. Yer değiştirme matrisi ile yapılan kodlamada ilk olarak peptit birimdik olarak kodlanır. Sonra her bir amino grup asidinin yer değiştirme matrisi içindeki satır ve sütundaki değeri ile birimdik vektörü çarpılır [3]. Geliştirilen kodlama yöntemleri makine öğrenmesi sınıflandırıcı algoritmaları ile test edilmiş ve elde edilen sonuçlar literatürdeki diğer öznitelik kodlama yöntemleri ile kıyaslanmıştır.

1.1 Tezin Amacı

Bağışıklık sisteminin harekete geçmesi için yardımcı T hücreleri özgün epitoplarına bağlanır ve bir seri biyokimyasal reaksiyon gerçekleşir. Bağışıklık sistemi hareketinde bir aşı ya da ilaç geliştirmek için T hücre epitoplarının önceden tahmin edilmesi büyük önem taşır. Sürekli mutasyona uğrayarak çeşitlilik gösteren sayısı binleri aşan antijenik peptitlerin tanımlanması laboratuvar ortamında zaman ve maliyet açısından uygun değildir. Bu nedenle bilgisayar ortamında makine öğrenmesi algoritmaları ile çözüm aramak daha uygundur.

Bu tez çalışmasında amacımız yardımcı T hücre epitoplarını kestirmek için yeni makine öğrenmesi öznitelik kodlama teknikleri geliştirmektir.

1.2 Literatür Araştırması

Zvelebil ve arkadaşları 1987 yılında Tvd yöntemine dayalı yeni bir kodlama yöntemi önermişlerdir. Yöntemde yüksek boyutlu olmayan amino asitlerin fizikokimyasal özelliklerini kullanarak %66'lık başarıya ulaşmışlardır [4][5]. İlk adım tüm dizinlerin hizalanmasıdır. İki proteini (veya nükleik asit) hizalamak için standart bir yöntem olan Needleman & Wunsch (1970) dinamik programlama yaklaşımını kullanmışlardır. Amino asit çiftleri arasında benzerlik (kimyasal özellikler veya gözlenen yerdeğiştirme) içeren bir matris seçilir ve algoritma kurulur.

C. Lundegaard ve arkadaşları 2008 yılında MHC I moleküllerinin bağlanma durumlarını tahmin etmek için NetMHC adında bir yöntem geliştirmişler ve bunun için yapay sinir ağlarını (YSA) kullanmışlardır. NetMHC hem Bağışıklık Epitop Veritabanı ve Analiz Kaynağı (IEDB)'den afinite veri kullanılarak, hem de SYYAEITHI'dan elüsyon veri kullanılarak çok çeşitli sayıda nicel peptitler üzerinde çalıştırılmıştır. Bu yöntem MHC peptit bağlanmasının yüksek doğruluk oranıyla tahmin edilmesine olanak sağlamaktadır. Bu yöntem % 75–80 oranında doğrulanmıştır[6].

Aidan NacNamara ve arkadaşları 2009 yılında yaptıkları çalışmada en popüler tahmin yazılımlarından ikisini, NetCTL ve NetMHC'yi kullanarak, tahminleri bir dizi MHC molekülüyle karşılaştırınca yeniden ölçeklendirmenin gerçek biyolojik çeşitliliği tahmin edilmiş yakınlarından uzaklaştırdığı hipotezini test etmişlerdir. Bunun sonucunda bu uzaklaştırmanın tahmin yazılımının performansını sıralı epitoplar açısından niteliksel olarak, bağlama yakınlık tahminlerinin doğruluğu açısından da niceliksel olarak geliştirdiğini gözlemlemişlerdir. Araştırmacılar yapay sinir ağlarını (YSA) kullanmışlar ve bu yöntem sonucunda 0.94'lük bir AUC değerine ulaşmışlardır [7].

Kirsten Roomp ve arkadaşları 2010 yılında yaptıkları çalışmada, bağlayıcıları ve bağlayıcı olmayanları seçmek için kullanılan IC50 kriterlerinde değişiklik gösteren 3 veri seti test etmişlerdir. Sadece güçlü bağlayıcıları (10 nM'den daha az IC50) ve net bağlayıcı olmayanları (10,000 nM'den daha büyük IC50) içeren datasetlerde tahminler uygulandığında, en iyi performans alınmıştır. Bunun yanısıra, tahminlerin sağlamlılığı yeterince büyük (200'den büyük), dengeli bağlayıcı ve bağlayıcı olmayan setlerle temsil edilen aleller için elde edilmiştir. Özetlemek gerekirse, Roomp ve arkadaşları dört farklı yöntem üzerinde BDUK sınıf I moküllerinin bağlanma durumlarını incelemişlerdir. Bu yöntemler: DynaPredPOS, Net-MHC, SVMHC ve YKW % 98'lik başarıya ulaşmışlardır [8].

Ju He ve arkadaşları 2012 yılında yaptıkları çalışmada bir öğrenme yöntemi olan Devamlı Çekirdek Ayrımı'nı (CKD) çeşitli uzunluklardaki BDUK sınıf II bağlayıcılarının tahmini için kullanmışlardır. Kompozisyon değişim ve ayrışım özellikleri peptit dizlimini kodlamak için kullanılmıştır. Metropolis Monte Carlo Simülasyon Yaklaşımı özellik seçiminde kullanılmıştır. Bu çalışma sonucunda, özellik seçiminin modelin performansını geliştirmede önemli bir rol oynadığı görülmüştür. Değerlendirme veri kümesi Dataset-2 için özelliklerin sayısı 147'den 44'e düşürülüp, AUC 0.7349'dan 0.8499'a artırılırken, değerlendirme veri kümesi Dataset-1 için, özelliklerin sayısı 147'den 24'e düşürülmüş ve AUC 0.8088'den 0.9034'e artırılmıştır. 10 kat çapraz doğrulama kullanılarak özellik seçimi ve bant genişliği optimizasyonundan uygun bir sürekli çekirdek ayrımı modeli elde edilmiştir. Bu modelin AUC değerleri değerlendirme veri kümesi, BM-Set1'de değerlendirilen; 0.831 ve 0.980 arasında ve değerlendirme veri kümesi BDUK sınıf

II aleller için BM-Set2'de değerlendirilen; 0.806 ve 0.949 arasındadır. Bu sonuçlar yapılan çalışmadaki modelin, aynı datasetlerin eğitilmesi ve test edilmesine dayanan önceki modellerden daha iyi performans sergilediğini göstermiştir [9].

M. Gök ve A.T. Özcerit 2012 yılında yaptıkları çalışmada, IEDB veri tabanından insan lökosit antijeni (İLA-A, İLA-B) peptitleri üzerinde OEDICHO adlı bir yöntem geliştirmişlerdir. Bu yöntemde, ortonormal kodlama (BKY) ve Amino Asit Dizini Veritabanı (Aaindex)'ten elde edilen amino asitlerin seçilen en iyi 10 fizikokimyasal özelliğinin ikili gösterimi birleştirilmiştir. Araştırmacılar kendi yöntemlerini güncel özellik kodlama teknikleriyle karşılaştırmışlardır. Gözleme dayalı sonuçlar bu çalışmadaki amino asit kodlama şeması bağımsız sınıflayıcılar üzerinde daha iyi sınıflandırma performansını sağlamaktadır. M. Gök ve A. T. Özcerit bu çalışma sonucunda %98'lik başarıya ulaşmışlardır [4].

Mariyana Atanasova ve arkadaşları 2012 yılında yaptıkları çalışmada BDUK sınıf II bağlanma tahmini için yapıya dayalı ilk sunucu olan EpiDOCK'u tanıtmışlardır. EpiDOCK 23 en sık insan, BDUK sınıf II proteinlerine bağlanmayı tahmin eder. Doğru bağlayıcıların %90'ını ve doğru bağlayıcı olmayanların %76'sını, %83 doğruluk oranıyla tanımlamıştır. BDUK sınıf II bağlanma tahmini NetMHCII için en son teknoloji sunucu ile kıyaslandığında, EpiDOCK AUC değerleri açısından daha iyi tahminler vermiştir. 0.667 ile kıyaslandığında 0.892'lik sonuca ulaşmışlardır [11].

Carla Oseroff ve arkadaşları 2012 yılında yaptıkları epitop özgüllüğü ve bunun IgE üretimi ile olan ilişkisini inceleyen çalışmada, çeşitli donör/ alerjen kombinasyonları için Ab ile T hücresi tepkilerini karşılaştırmışlardır. Bunun sonucunda, bağlı olmayan T hücresi-B hücresi yardımı IgE tepkilerinin gelişmesini destekleyebilir diye varsayılarak, ortaya çıkarılabilir olan T hücresi tepkilerinin yokluğunda IgE tepkileri belirlenmiştir. Özet olarak, Bla g alerjenlerine karşı T hücresi tepkilerinin özellikleri IgE tepkileriyle alakasız görünmektedir. Bu tepkileri gözlemlemek, hamamböceği alerjileri ve tedavi yöntemlerini anlamada önemli bilgiler sağlayabilir. Sonuçlar, Bla g alerjenlerine karşı T hücre tepkilerinin, aynı alerjenlere karşı IgE tepkilerinden ayrılan önemli özellikleri olduğunu göstermiştir [12].

Cem Meydan ve arkadaşlarının 2013 yılında yaptıkları çalışmada, tahmin yöntemi BDUK sınıf I için 18 farklı alel ve BDUK sınıf II için 27 farklı alelden oluşan testle değerlendirilmiştir. Elde edilen sonuçlar her iki BDUK sınıfı için de en son teknoloji yöntemlerle karşılaştırılabilir. Bu çalışmada, ortalama AUC tahmin değerleri sınıf I için 0.897 ve sınıf II için 0.858 olarak bulunmuştur. Çalışmada önerilen yöntem peptit uzunluğuna dayanmaz ve hem uzun hem de kısa parçalarla çalışabilir. Bu durum, var olan eğitme verilerinin daha iyi kullanımı ve alışılmamış uzunluklardaki peptitlerin tahmini için bir avantajdır [13].

Edita Karosiene ve arkadaşları 2013 yılında yapmış oldukları çalışmada tanımlanmış bir protein dizilimi olan bir sınıf İnsan Lökosit Antijen (İLA) molekülüne peptit bağlanmasını tahmin edebilen ilk pan-spesifik yöntemi sunmuşlardır. Bu yöntem, sözde dizilim açısından BDUK ortamı peptit bağlanmasını tanımlamak amacıyla, İLA-DR, İLA-DP ve İLA-DQ molekülleri için ortak bir strateji kullanmaktadır. Bu strateji diğer türlerden bile yeni moleküllerin kaynaştırılmasına olanak sağlar. Çalışmada kullanılan yöntem, birkaç kriterle değerlendirilmiş ve daha önceden karakterize edilmeyen BDUK II moleküllerinde peptit bağlanmasının tahmin yeteneğinin yanısıra molekül-spesifik yöntemler üzerinde önemli bir gelişmeyi de gözler önüne sermiştir. Sonuçlar yeni yöntemin sözde dizilimin elde edilmesi için orijinal yönteme göre daha gelişmiş tahminlere ışık tuttuğunu göstermiştir. AUC değerleri 0.688'den 0.695'e ve 0.846'dan 0.847'ye yükselmiştir. Sonuçlar yapay sinir ağı eğitimleri için sözde dizilimlerin elde edildiği yeni yaklaşımın performansı geliştirdiğini göstermiştir [14].

Michael E. Birnbaum ve arkadaşları 2014 yılında yaptıkları çalışmada doğrudan seçime dayanan çok büyük data setleri üreten kombinasyonal, biyokimyasal yaklaşımı kullanan T hücre reseptörleri (TCR) özgüllüğü ve karşı reaksiyon altında yatan mekanizmaları tanımlamayı amaçlamışlardır. Araştırmacılar bu çalışmada, birçok fare ve insan TCR'si tarafından tanınan yüzlerce özgün peptit dizilimini keşfetmek için afinite bazlı pMHC maya kütüphaneleri ve derinlemesine dizilim seçimlerini birleştirmişlerdir. Araştırmacılar, derinlemesine dizilim sonuçlarından dataları kullanarak, doğal yollarla oluşan TCR ligandlarını tahmin etmek için bir sayısal algoritma geliştirmişlerdir. Varsayılan çok çeşitli TCR-reaktif peptitler test edilmiş ve %94'ünün T hücre tepkisini ortaya çıkarabildiği görülmüştür. Sonuçlar,

yapısal prensiplerin TCR'ye çok sayıda özgün pMHC'ye pMHC tanıtımında bozulma gerektirmeden bağlanmasına olanak sağladığını göstermiştir [15].

Julia Schwaiger ve arkadaşları 2015 yılında CD4+ T hücresinin TBE hastalarına kıyasla aşılı insanlardaki viryon proteinlere nasıl cevap verdiğini araştıran bir çalışma yapmışlardır. İnterlökin-2 (IL-2) eliza testi (ELISpot) içindeki örtüşümsel peptitlerle elde edilen data, kapsid (C) ve E proteinlerinin 3 boyutlu yapılarının yanı sıra, BDUK sınıf II peptit eğilimlerini baz alan epitop tahminleri ile de ilişkilendirilerek analiz edilmiştir. C proteini için epitop tahminleri E proteininden daha iyi çıkmıştır ve özellikle transmembran bölgeler için hatalıdır. Bu çalışmadaki data, peptit sürecini etkileyen, deneysel olarak belirlenen ve bilgisayarla tahmin edilen CD4+ T hücre epitopu arasında gözlemlenen farklılıklara katkı sağlayan protein yapısal özelliklerinin güçlü etkisi için kanıt sunmaktadır [16].

1.3 Hipotez

T hücrelerinin bağlanma yerlerinin tespitinde amino asitlerin fizikokimyasal özellikleri ve dizilimi oluşturan kalıntıların ağırlık ile konumlarının kullanılabileceğini düşünüyoruz.

1.4 Tez Organizasyonu

Bu tez çalışmasının içeriğinde bulunan bölümleri kısaca ele alacak olursak ;

Birinci bölüm; tezin amacı, literatür tarama, hipotez ve tez organizsyonu başlıklarından oluşmaktadır. Tezin yazılmasındaki genel amaç, bu konunun seçilmesinde ve çalışmalar süresince bize ışık tutan geçmişte ve günümüzde yapılmış olan çalışmaların neler olduğuna dair literatür araştırması ve tezin sonucunda ulaşmayı hedeflediğimiz başarıya hangi yöntemle ulaşılacağı konularından bahsedilmiştir.

İkinci bölümde; bağışıklık sistemi, epitoplar, amino asitler ve fizikokimyasal özellikleri, yardımcı T hücrelerinin epitoplarına bağlanma durumları ayrıntılı olarak ele alınmıştır.

Üçüncü bölümde; öznitelik çıkarımının nasıl yapıldığı, öznitelik çıkarımı için kulanılan yöntemler, sınıflandırma modeli, sınıflandırma işlemi için kullanılan öğrenme algoritmaları konularından bahsedilmiştir.

Dördüncü bölümde; geliştirilen öznitelik çıkarım yöntemleri örnekler üzerinde ele alınmış ve deneysel sonuçlar kısımında yöntemin veri setlerine uygulandığında elde edilen sonuçlar tablolar üzerinde gösterilmiştir.

Beşinci bölümde; bundan sonraki yapılacak çalışmalara ışık tutması amacıyla ulaşılan sonuçlar yorumlanmıştır. Gelecekte bu alanda yapılabilecek çalışmalara değinilerek tez çalışması bitirilmiştir.

2. T HÜCRELERİ EPİTOPLARI

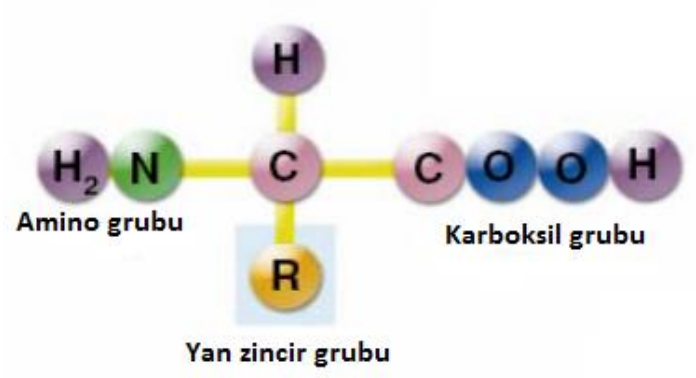
2.1 Amino Asitler ve Fizikokimyasal Özellikleri

Amino asitler; moleküllerinde amino grubu (−NH2) ve karboksil grubu (−COOH) bulunan bileşiklerdir. Doğada 300 civarı amino asit bulunmaktadır. Protein yapısında yer alan 20 amino asit doğada yaygın olarak bulunur [2]. Çizelge 2.1'de 20 standart amino asit görülmektedir.

Çizelge 2.1 :20 standart amino asit [2].

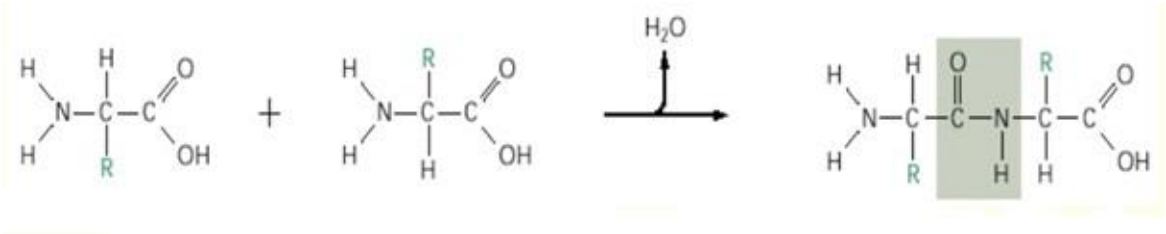
Sıra	Amino Asit	1-harf	3-harf	Sıra	Amino Asit	1-harf	3-harf
1	Alanin	A	Ala	11	Lösin	L	Leu
2	Arginin	R	Arg	12	Lizin	K	Lys
3	Asparajin	N	Asn	13	Metiyonin	M	Met
4	Aspartik asit	D	Asp	14	Fenilalanin	F	Phe
5	Sistein	C	Cys	15	Prolin	P	Pro
6	Glütamin	Q	Gln	16	Serin	S	Ser
7	Glütamik asit	E	Glu	17	Treonin	T	Thr
8	Glisin	G	Gly	18	Triptofan	W	Trp
9	Histidin	H	His	19	Trozin	Y	Tyr
10	İzolösin	I	Ile	20	Valin	V	Val

Şekil 2.1'de görüldüğü üzere her bir standart amino asit merkezi α-karbonu (C_α), amino ($-NH_2$), karboksil (−COOH) grupları ile yan zincir (R) grubu olmak üzere dört kısımdan oluşur. Amino, karboksil ve yan zincir grupları kovalent bağlar ile merkezi α-karbonuna bağlıdırlar [17].



Şekil 2.1: Standart bir amino asitin yapısı [17].

Amino asitler Şekil 2.2'te görüldüğü gibi birbirlerine peptit bağlar ile bağlanırlar. Amino asitlerin -COOH grubu $–NH_2$ grubuna sahip komşu bir $–NH_2$ vericisi ile reaksiyona girerek peptit bağı oluştururlar.



Şekil 2.2 : İki amino asitin peptit bağ oluşturması [17].

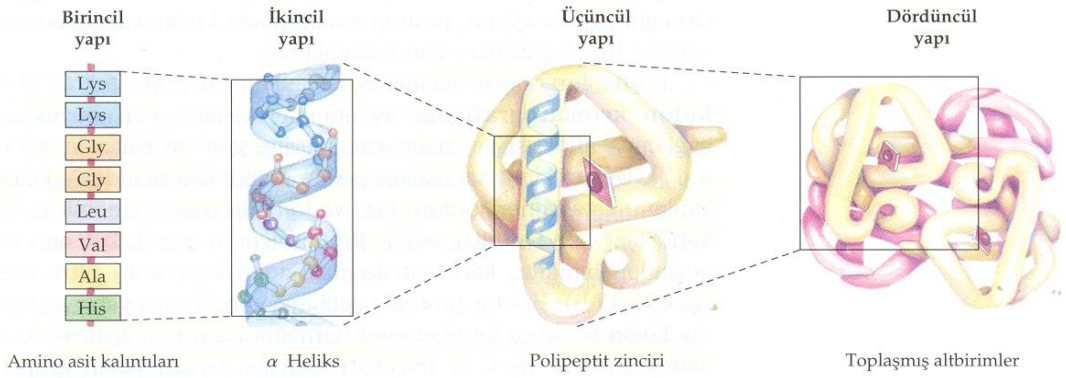
Amino asitlerin standart olarak adlandırdığımız 20 tanesi, DNA tarafından kodlanan ve proteinleri oluşturan amino asitlerdir. Canlılarda DNA ile RNA hangi durumda hangi proteinin gerekli olduğunu enzimler aracılığıyla hücreye bildirip protein sentezini başlatırlar. Hücre içerisinde ribozomların, mesajcı RNA (mRNA) moleküllerini kullanarak amino asitleri uç uca eklemesiyle protein sentezlenir.

Proteinlerin içerisinde çok sayıda ve dizilimde bulunan amino asitler farklı yapıda binlerce çeşit protein oluşturur. Proteinler, amino asitlerin belirli yapıda ve sayıda belirli bir dizilimle düz zincir üzerinde birbirlerine kovalent bağlanarak oluşan polipeptitlerdir. Az sayıda amino asidin bağlanmasıyla oluşan yapıya oligopeptit ya da sadece peptit, çok sayıda amino asidin bağlanmasıyla oluşan yapıya ise polipeptit yada protein adı verilir [10].

Protein yapısının Şekil 2.3'te görüldüğü gibi dört düzeyi tanımlanmıştır:

Birincil yapı (primary structure), proteinleri meydana getiren amino asitlerin hangi sırayla birbirlerine bağlandıklarını gösteren en basit ve düz yapıdır. Diğer üçü ise proteinlerin üç boyutlu katlanma durumlarına dayalıdır. İkincil yapı (secondary structure), hidrojen bağları ile kararlı kılınan, düzenli tekrarlanan, kendi üzerinde kıvrım ve katlanmalar yapan geometrik yapılardır. α sarmalı (α helix) ve β yaprağı (β sheet) en yaygın ikincil yapılardır. Üçüncül yapı (tertiary structure), proteinin üç boyutlu gösterimdir.

Dördüncül yapılar (quarternary structure) ise birden fazla çoklu peptit içeren karmaşık, büyük proteinler için geçerli yapılardır. Proteinin tam işlevsel hali dördüncül yapılardır. Dördüncül yapılar, protein içindeki peptit dizilimlerinin birbirleri ile olan etkileşimlerini tanımlar [10].



Şekil 2.3: Başlıca dört protein yapı düzeyi [10].

Protein veri bankasından protein yapıları ile ilgili bilgilesre ulaşılabilir. Protein veri bankası, proteinler ve nükleik asitler gibi yapıları içerisinde bulunduran büyük bir veri tabanıdır [2].

Amino asitlerin fizikokimyasal özelliklerini her amino asitte bulunan özgün farklılıklar gösteren yan zincir grubu belirler [18]. Fizikokimya, maddenin moleküler ve atomik düzeydeki davranışları ve kimyasal reaksiyonların oluşumunun nasıl olduğunu, moleküllerin doğasını açıklamak için birbirleri ile olan etkileşimleri ve bu etkileşimler sırasında meydana gelen enerji alış verişlerini inceler. Bu etkileşimler moleküllerden oluşan amino asitlerin ait oldukları proteinlerin işlevlerini belirler [19].

Amino asitlerin fizikokimyasal özelliklerinin indeks değerlerini barındıran Aaindex [20] adında bir veri tabanı bulunmaktadır.

Amino asitler, hidrofobiklik, polarlık, moleküler ağırlık gibi birbirinden farklı pek çok fizikokimyasal özelliğe sahiptirler. Bu özellikler, amino asit indeksi adı verilen 20 sayısal değerden oluşan vektörler ile ifade edilebilirler.

Çizelge 2.2'de amino asitlerin fizikokimyasal özelliklerinin niceliksel olarak ifade edildiği indeks değerleri görülmektedir.

Çizelge 2.2: Amino asitlerin fizikokimyasal özelliklerine ait indeks tablosu örneği[2]

	1. fizikokimyasal özellik	…	63. fizikokimyasal özellik	…	544. fizikokimyasal özellik
Amino asit	Alpha-CH kimyasal kaydırma (Andersen et al., 1992	…	Büyüklük (Dawson, 1972)	…	Hidrofobiklik indeksi (Fasman, 1989)
A	4,35	…	2,5	…	-0,21
R	4,38	…	7,5	…	2,11
N	4,75	…	5	…	0,96
D	4,76	…	2,5	…	1,36
C	4,65	…	3	…	-6,04
Q	4,37	…	6	…	1,52
E	4,29	…	5	…	2,3
G	3,97	…	0,5	…	0
H	4,63	…	6	…	-1,23
I	3,95	…	5,5	…	-4,81
L	4,17	…	5,5	…	-4,68
K	4,36	…	7	…	3,88
M	4,52	…	6	…	-3,66
F	4,66	…	6,5	…	-4,65
P	4,44	…	5,5	…	0,75
S	4,5	…	3	…	1,74
T	4,35	…	5	…	0,78
W	4,7	…	7	…	-3,32
Y	4,6	…	7	…	-1,01
V	3,95	…	5	…	-3,5

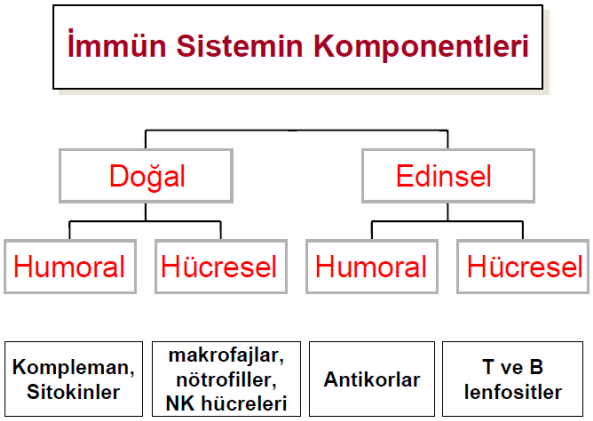
Örneğin 63.fizikokimyasal özellik olan büyüklük, her bir amino asit için 0,5 ile 7 arasında değişen değerlerden oluşur. Bu değerler amino asitlerin büyüklük dereceleridir.

2.2 T Hücreleri

Bağışıklık sistemi vücuda giren zararlı mikro organizmaları antijen-antikor bağlanması ile tespit ve yok eder. Epitoplar (antijen determinantı), antijen molekülü

yüzeyinde yer alan ve antijenlerin kendi özgül antikorları ile birleşebilmelerini sağlayan biyokimyasal gruplardır. Antikorun değişken kısımları "anahtar ve kilit" kuramı uyarınca antijen epitopu aracılığı ile antijene bağlanır. Bunun sonucunda zararlı mikroorganizma hücreye giremez ve bağışıklık sistemi tarafından yok edilir. Bağışıklık sistemi enfeksiyonlardan korunmada anahtar rol oynar. Kemik iliğinde başlayıp daha sonra kan dolaşımı ile doku ve organlarda devam eden bu süreç, bağışıklık sistemini oluşturan hücrelerin gelişimi ve özel hücre türlerine farklılaşmaları ile tamamlanır.

Bağışıklık sistemi Şekil 2.4'te görüldüğü gibi doğal ve kazanılmış olarak ikiye ayrılır. Bu sistemler karşılıklı etkileşim içinde çalışırlar.



Şekil 2.4 : Bağışıklık sisteminin bileşenleri [21].

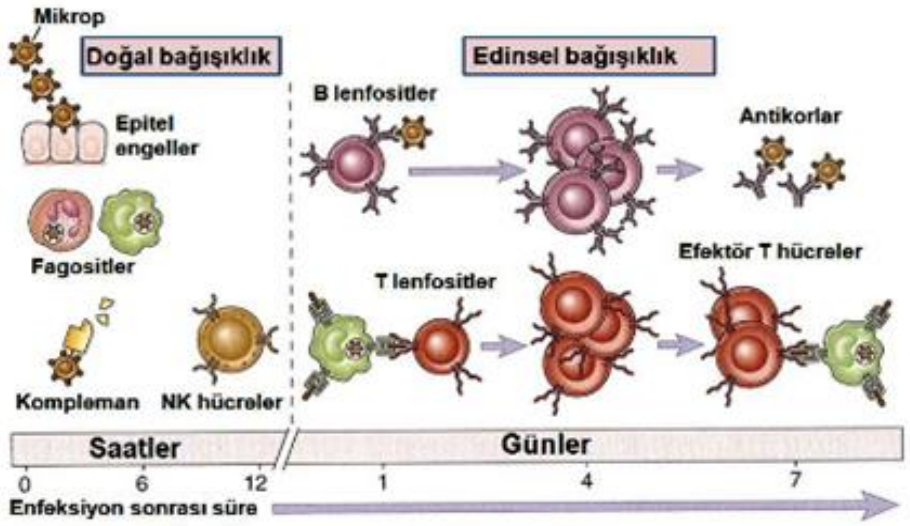
Doğal bağışıklık sistemi, edinilmiş bir bilgi olmadan bir antijenle ilk karşılaşma ile görevi başlayan koruyucu mekanizmalardan oluşur.

Edinsel bağışıklık sistemi, antijene özgü bağışıklık yanıtından sorumludur. Şekil 2.5' te görüldüğü gibi lenfositler edinsel bağışıklığı oluşur.
Lenfositlerin antijenlerle tekrar karşılaştıklarında hatırlamalarını sağlayan hafıza yetenekleri kazanılmış bağışıklık yanıtın temelini oluşturur [21].
Lenfositler, kemik iliğinde gelişen akyuvarlardır ve farklı antijenik yapıları tanıma ve birbirinden ayırabilme yetisine sahip tek hücre topluluğudur. Bu yüzden bağışık yanıt oluşumunda lenfositler kilit rol oynarlar.

Antijene özgü bağışıklık yanıtının oluşmasını sağlarlar. Lenfositler Şekil 2.6'da görüldüğü gibi T, B ve NK hücrelerinden oluşur [22].

B ve T hücrelerinin temel farkı antijen reseptörleridir. T hücre yüzeyinde yüzey immunoglobulin bulunmaz. Bunun yerine antijenleri özgül olarak tanıyan T Hücre Reseptörü (THR) bulunur [23]. Kazanılmış bağışıklık, THR ve antikorlar (B hücreleri) aracılığıyla yürütülen antijen tanıma işlevi ile doğal bağışıklığın eksik kaldığı durumlarda bağışık yanıt oluşturabilir [24].

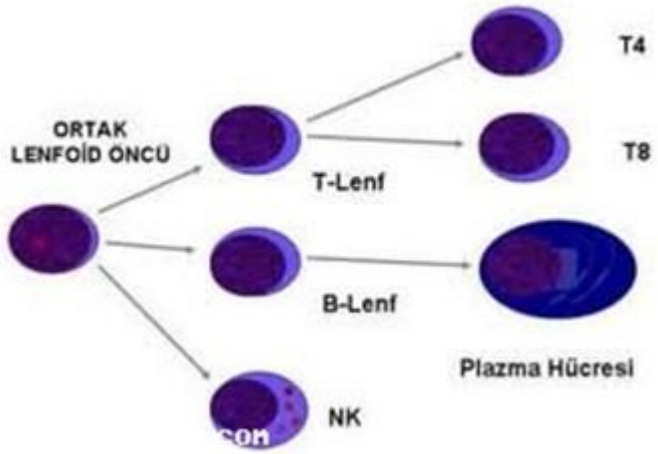
Kemik iliğinde gelişmeye başlayan T lenfositleri olgunlaşmalarını timusta tamamlar. T hücrelerinin, patojenlerin ortadan kaldırılmasında doğrudan rolleri vardır. Ayrıca yardımcı hücre işlevi ile doğrudan hücresel temas ya da sitokinler aracılığı ile B ve T hücre yanıtını etkilerler. Her T hücresi tek bir antijene özgül bir THR'ye sahiptir. T lenfositler antijenik peptidleri THR ile tanır. İmmun sistemde zaman içinde karşılaşma ihtimali olan on binlerce çeşit antijene yanıt verebilecek on binlerce çeşit T lenfositi bulunur [26]. Şekil 2.5'de doğal ve edinsel bağışıklığın temel bileşenleri görümektedir.



Şekil 2.5 : Doğal ve kazanılmış bağışıklık yanıtının bileşenleri [26].

Şekil 2.6'da görüldüğü gibi T hücreleri ikiye ayrılır. CD4 yardımcı T lenfositler (Thelper-Th), BDUK Sınıf I molekülüne bağlanabildiği için lenfositlerin THR ile birlikte ASH ile bağlantısında görev alır [27].

CD8 öldürücü T lenfositler (Tcytotoxic-Tc) virüs, parazit ve hücre içi bakteriler ile tümör hücrelerine, organizmada yabancı kabul edilen doku ve organ hücrelerine doğrudan saldırırlar [25].



Şekil 2.6: Lenfositler [25].

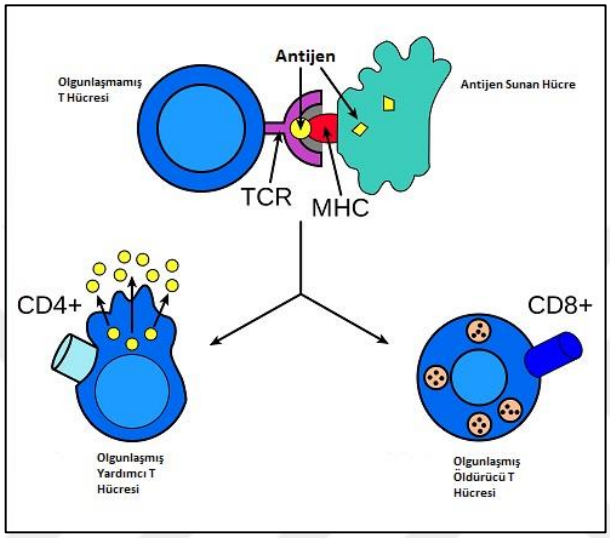
2.2.1 T Hücreleri/ BDUK moleküllerinin bağlanma özgünlüğü

T hücreleri antijene özgünlüğü çok yüksek olan hücrelerdir. Doğada bulunan protein yapısındaki antijenlerin her bir peptidine özgün, farklı bir T hücre klonu olduğu kabul edilir. T hücrelerinin diğer bir önemli özelliği ise bir antijeni tanıyıp reaksiyon oluşturabilmesi için bu antijenin bazı hücreler tarafından işlenmesi ve membranlarında bulunan yüzey molekülleri aracılığıyla kendilerine sunulması gerekmektedir. T hücrelerine antijen sunumunu sağlayan hücre yüzeyindeki moleküllere "Büyük Doku Uygunluk Kompleksi" (BDUK) denir [28].

BDUK molekülleri ya da diğer adıyla insan lökosit antijenleri, T hücrelerine antijen sunup onları aktive eden ve T hücre immün yanıtının yönünü belirleyen hücre yüzey molekülleridir. Bu moleküllerin özgünlüğü yüksek değildir. Yapıca birbirine benzeyen farklı peptitler aynı BDUK molekülüne bağlanıp sunulabilirler.

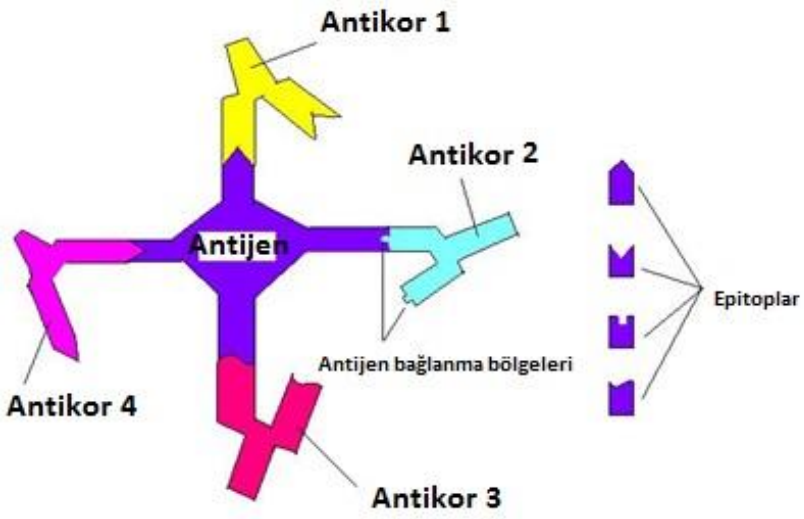
Kazanılmış bağışıklık sisteminin en önemli hücrelerinden biri olan T hücrelerinin protein yapıdaki antijene karşı oluşturacağı özgün yanıtın gerçekleşebilmesi için Şekil 2.7'de görüldüğü gibi bu antijenin antijen sunan hücreler (ASH) tarafından alınması, işlenmesi ve yüzeylerinde bulunan bazı moleküllere bağlanarak T hücrelerine sunulması gerekmektedir. Bu amaçla kullanılan yüzey molekülleri büyük

doku uygunluk kompleksi molekülleridir [28]. Hücre yüzeyinde bulunan BDUK molekülleri yabancı antijenleri bağlayarak bağışıklık sisteminin effektör hücrelerine sunar ve bu şekilde bağışık yanıtın başlamasında anahtar rol oynarlar [29]. BDUK molekülleri, insanda 6.kromozomun kısa kolunda yerleşen yaklaşık 4000 kilobazlık dev bir gen kompleksidir.



Şekil 2.7: T hücrelerinin bağlanması [29].

Vücuda yabancı olup bağışıklık sisteminin harekete geçmesini sağlayan, sayıları çok fazla olan antijen molekülleri protein yapısındalardır. Uzun antijen proteinlerinin T hücreleri ile birleşebilmeleri için küçük peptitlere ayrılırlar. Bu antijenik peptitlere "Epitop" adı verilir. Bağışıklık sisteminin harekete geçmesi için T hücreleri üzerinde bağlanma ve tetikleme olması gereklidir. Bu bağlanma BDUK moleküllerinin yüzeye çıkarmış olduğu antijenik peptit parçaları diğer bir deyişle epitoplar üzerinde olur [28].



Şekil 2.8 : Antijenlerin bağlanma örneği [28].

Epitoplar, Şekil 2.8'de görüldüğü gibi antikorun bağlandığı antijene özel bir peptit parçacığıdır.

2.3 T Hücrelerinin Bağlanma Özgünlüğü Veri Setleri

Bu tez çalışmasında yardımcı T hücreleri / BDUK molekülleri bağlanma özgünlüklerinin tanımlanmasında Bağışıklık Epitop Veritabanı (IEDB)[30]'den insan lökosit antijeni (İLA-A, İLA-B) peptit verileri kulllanılmıştır.

Veri seti dokuz amino asit uzunluğundaki peptitlerden oluşmaktadır. Peptitlerin bağlanma / bağlanmama durumu IC 50 kesim değerine göre belirlenmiştir. Veri setindeki peptitler tam, orta ve güçlü olmak üzere üç seviyeye ayrılmıştır. Orta veri seti (OV) 50-500 nM bağlanma, 500-1000 nM bağlanmama afinitesindeki peptitleri, güçlü veri seti (GV) 10 nM'den küçük bağlanma, 10.000 nM'den büyük bağlanmama afinitesindeki peptitleri içermektedir. Tam veri seti (TV) ise tüm olası peptitleri içermektedir [31].

3. PROTEİNLER İÇİN KULLANILAN ÖZNİTELİK ÇIKARIM YÖNTEMLERİ

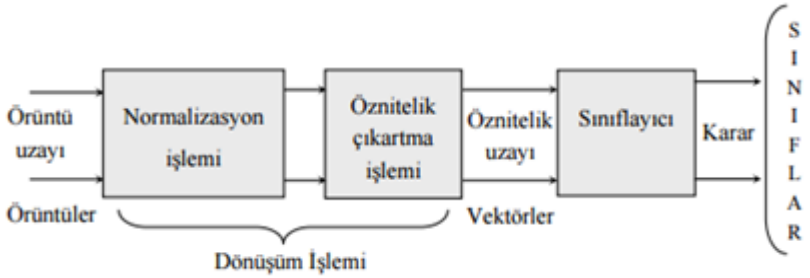
Sınıflandırma, ilk kez karşılaşılan veri kümesi içerisindeki örneklerin daha önceden karşılaşılmış olan verilerden elde edilen bilgilerle karşılaştırılarak sınıflara ayrılması işlemidir. Veri kümesinin bir kısmı eğitim kümesi olarak kullanılarak sınıflandırma kuralları oluşturulur. Bu kurallar yardımıyla yeni bir veri ortaya çıktığında nasıl karar verileceği belirlenir.

Proteinleri oluşturan amino asitlerin, sayısal olarak temsil edilmesi örüntü modellemenin bir ayağıdır. Proteinlerin temsilinde, amino asitlerin protein içindeki yeri, sayısı veya fizikokimyasal özellikleri, öznitelik vektörler kümesi olarak ifade edilebilir [2]. Örüntüye ait karakteristik özellikler ne kadar net temsil edilirse sınıflandırıcı örüntüyü o kadar iyi tanır. Bilgisayarların örüntüleri algılayabilmesi için bu uzaydaki her bir örüntünün, bilgisayarın işleyebileceği bir vektörel form ile temsil edilmesi gerekmektedir.

Dönüşüm işlemini, kategorileri birbirinden ayıran ve kendi kategorisini en iyi temsil eden özniteliklerin bulunması aşamasıdır [32]. Dönüşüm kısmının amacı, örüntünün temsilini sonraki adımlar için daha anlaşılır ve işlenebilir hale getirmektir. Bu amaç doğrultusunda örneğin biyobilişimde protein verileri üzerine araştırma yapılıyorsa örüntü verisi normalize edilerek bir kodlamaya tabi tutulabilir.

Öznitelik, örüntüye ait ölçülebilir veya gözlenebilir bilgi olarak tanımlanabilir. Eğitim sürecinde, öznitelik çıkartılması kısmında temsil edilen örüntü verileri için en uygun öznitelikler tespit edilir ve sınıflandırıcı öznitelik uzayını bu yönde çeşitli bölümlere ayırır. Sınıflandırma kısmında eğitilmiş sınıflandırıcı, giriş örüntülerini öznitelik ölçümlerine göre hangi sınıflara ait olduğuna karar verir. Makine öğrenmesi ile gerçekleştirilen örüntü tanıma modellemelerinde amaç, deneysel gözlemlerin tam olarak temsilini öğrenmek değil, temel fonksiyonu üretebilmek ve eğitim

verilerinden farklı örnekler üzerinde başarılı biçimde genelleme yapabilmesini sağlamaktır [2]. Şekil 3.1'de bir sınıflandırma işleminin genel yapısı görülmektedir.



Şekil 3.1 : Karar verme mekanizması [32].

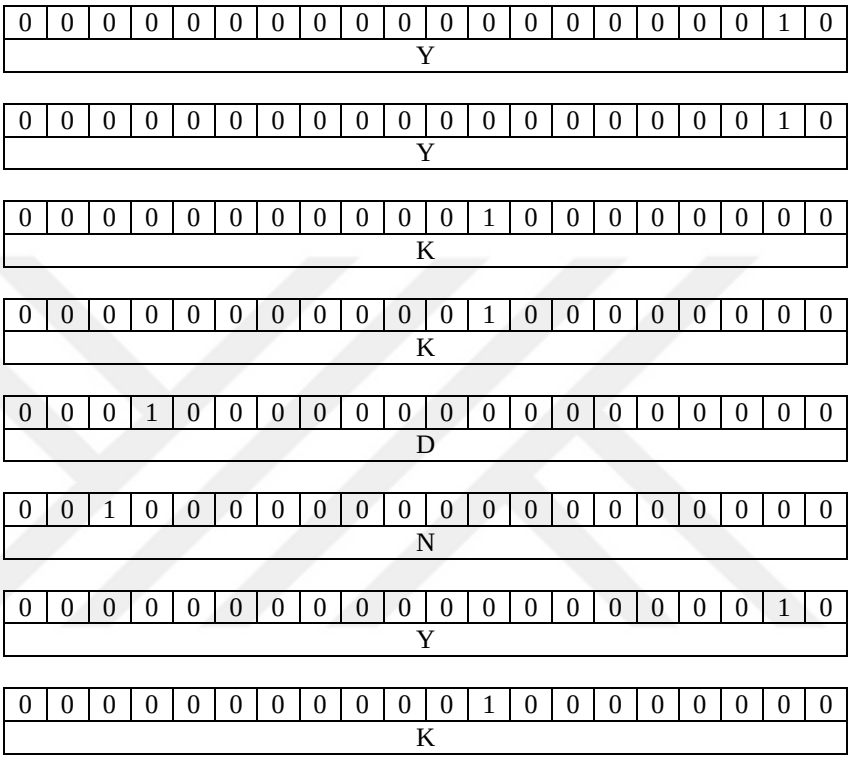
3.1 Birimdik (Orthonormal) Öznitelik Kodlama Yöntemi

Birimdik öznitelik kodlama yönteminde (BKY) peptidi oluşturan her bir amino asit sembolü, birbirine dik vektörlerle ile ifade edilir [2]. Şekil 3.2'de görüldüğü gibi BKY'de her bir kalıntı 20 bit uzunluğunda vektör ile temsil edilir. Bu temsilde, her bir kalıntının sırasına karşılık gelen bit 1 ile geri kalan değerler ise 0 ile temsil edilir.

AMİNO ASİTLER	20 BİT VEKTÖR																			
A	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**

Şekil 3.2 : Amino asitlerin standart BKY ile temsil edilmeleri[2].

YYKKDNYK şeklinde verilen örnek peptiti BKY yöntemi ile açıklayalım. Peptitin ilk amino asiti Y'nin BKY öznitelik vektörü Şekil 3.3'te görüldüğü gibi [00000000000000000010]'dir. Peptit dizilimi 8 amino asit uzunluğunda olduğu için oluşan öznitelik vektörünün büyüklüğü 1*160 olacaktır.



Şekil 3.3: YYKKDNYK peptit diziliminin BKY ile kodlanması.

BKY, amino asitlerin birbirleri ile olan fizikokimyasal benzerlikleri veya farklılıkları hakkında bilgi içermiyor oluşu verinin içeriğinin tam olarak temsil edilememesine neden olur [33][20].

3.2 Yer Değiştirme Matrisleri Tabanlı Öznitelik Kodlama Yöntemleri

Yer değiştirme matrislerindeki sayılar, amino asitlerin birbirleri yerine geçme eğilimlerine dair bilgi verir. PAM, BLOSUM olmak üzere çeşitli yer değiştirme matrisleri vardır[34].

Yer değiştirme matrislerinde P_i ve P_j amino asitlerinin birbirlerine benzerlik oranı, i. satır ve j. sütunun kesişimi ile elde edilir. Yer değiştirme matrisi ile yapılan kodlamada önce peptit birimdik olarak kodlanır. Sonra her bir amino asitin yer değiştirme matrisi içindeki i. satır ve sütundaki değeri ile birimdik vektör çarpılır [2]. Şekil 3.4'te Blosum 50 amino asit yer değiştirme matrisi görülmektedir.

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	5	-2	-1	-2	-1	-1	-1	0	-2	-1	-2	-1	-1	-3	-1	1	0	-3	-2	0
R	-2	7	-1	-2	-4	1	0	-3	0	-4	-3	3	-2	-3	-3	-1	-1	-3	-1	-3
N	-1	-1	7	2	-2	0	0	0	1	-3	-4	0	-2	-4	-2	1	0	-4	-2	-3
D	-2	-2	2	8	-4	0	2	-1	-1	-4	-4	-1	-4	-5	-1	0	-1	-5	-3	-4
C	-1	-4	-2	-4	13	-3	-3	-3	-3	-2	-2	-3	-2	-2	-4	-1	-1	-5	-3	-1
Q	-1	1	0	0	-3	7	2	-2	1	-3	-2	2	0	-4	-1	0	-1	-1	-1	-3
E	-1	0	0	2	-3	2	6	-3	0	-4	-3	1	-2	-3	-1	-1	-1	-3	-2	-3
G	0	-3	0	-1	-3	-2	-3	8	-2	-4	-4	-2	-3	-4	-2	0	-2	-3	-3	-4
H	-2	0	1	-1	-3	1	0	-2	10	-4	-3	0	-1	-1	-2	-1	-2	-3	2	-4
I	-1	-4	-3	-4	-2	-3	-4	-4	-4	5	2	-3	2	0	-3	-3	-1	-3	-1	4
L	-2	-3	-4	-4	-2	-2	-3	-4	-3	2	5	-3	3	1	-4	-3	-1	-2	-1	1
K	-1	3	0	-1	-3	2	1	-2	0	-3	-3	6	-2	-4	-1	0	-1	-3	-2	-3
M	-1	-2	-2	-4	-2	0	-2	-3	-1	2	3	-2	7	0	-3	-2	-1	-1	0	1
F	-3	-3	-4	-5	-2	-4	-3	-4	-1	0	1	-4	0	8	-4	-3	-2	1	4	-1
P	-1	-3	-2	-1	-4	-1	-1	-2	-2	-3	-4	-1	-3	-4	10	-1	-1	-4	-3	-3
S	1	-1	1	0	-1	0	-1	0	-1	-3	-3	0	-2	-3	-1	5	2	-4	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	2	5	-3	-2	0
W	-3	-3	-4	-5	-5	-1	-3	-3	-3	-3	-2	-3	-1	1	-4	-4	-3	15	2	-3
Y	-2	-1	-2	-3	-3	-1	-2	-3	2	-1	-1	-2	0	4	-3	-2	-2	2	8	-1
V	0	-3	-3	-4	-1	-3	-3	-4	-4	4	1	-3	1	-1	-3	-2	0	-3	-1	5

Şekil 3.4: Blosum 50 yer değiştirme matrisi [2].

YYKKDNYK şeklinde verilen örnek peptit üzerinde Şekil 3.4'te görülen Blosum 50 yer değiştirme matrisine göre öznitelik kodlamasını açıklayalım. İlk adımda YYKKDNYK peptitindeki her bir amino asitin kendi i. satır ve sütun değerleri elde edilir. Buna göre Y=8, K=6, D=8, N=7 olur.

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
Y																			

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
Y																			

0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
K																			

0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
K																			

0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D																			

0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N																			

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
Y																			

0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
K																			

Şekil 3.5: YYKKDNYK peptit diziliminin Blosum 50 yerdeğiştirme matrisine göre kodlanması.

İkinci adımda, yukarıda belirlenmiş olan değerler, Şekil 3.5'te görüldüğü gibi BKY öznitelik vektöründe karşılık gelen amino asit ile çarpılarak yeni öznitelik vektörleri elde edilmiş olur. Peptit dizilimi 8 amino asit uzunluğunda olduğu için oluşan öznitelik vektörünün büyüklüğü 1*160 olacaktır.

3.3 Ağırlık ve Konum Tabanlı Öznitelik Kodlama Yöntemi

Ağırlık ve konum tabanlı öznitelik kodlama yöntemi (AKKY), peptit dizilimlerinin ağırlıkları ve konumları üzerine kurulu bir öznitelik çıkarım yöntemidir. Yöntemin uygulanmasında ilk olarak peptit dizilimini oluşturan her bir amino asitin ağırlığı (a_i) (3.1),

$$a_i = \frac{P_T}{N} \tag{3.1}$$

ve konumu(k_i) (3.2),

$$k_i = \frac{R_T}{N(N-1)} \tag{3.2}$$

denklemleri ile hesaplanır. Burada P_T, P kalıntısının peptit içindeki sayısı, R_T kalıntının peptit içindeki indeks numaralarının toplamı, N ise peptitte bulunan kalıntı sayısıdır. Sonraki adımda BKY ile AKKY birleştirilir. Buna göre peptit içinde yer alan her bir amino asit ağırlığı ve konumu ile amino asite karşılık gelen birimdik vektörü çarpılır [22].

YYKKDNYK şeklinde verilen örnek peptit için AKKY'yi açıklayalım. İlk adımda peptit dizilimi BKY'ye göre kodlanır. İkinci adımda peptitit içindeki amino asitlerin ağırlık ve konumları hesaplanır. Y amino asidi için ağırlık,

$$a_Y = \frac{3}{8} = 0{,}3 \tag{3.3}$$

Y amino asiti için konum Şekil 3.6'dan yararlanılarak,

İndis	**1**	**2**	3	4	5	6	**7**	8
Amino asit	Y	Y	K	K	D	N	Y	K

Şekil 3.6: Örnek peptit dizilimi indisi.

$$k_Y = \frac{1+2+7}{7*8} = 0.1 \tag{3.4}$$

olarak hesaplanır. Bu işlem dizilim içindeki diğer amino asitler için de tekrarlanır.

$a_K = 3/8 = 0{,}3$ $\quad k_K = 15/56 = 0{,}2$

$a_D = 1/8 = 0{,}1$ $\quad k_D = 5/56 = 0{,}08$

$a_N = 1/8 = 0{,}1$ $\quad k_N = 6/56 = 0{,}1$ olarak bulunur.

AĞIRLIK																			
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
0	0	0,1	0,1	0	0	0	0	0	0	0	0,3	0	0	0	0	0	0	0,3	0

KONUM																			
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
0	0	0,1	0,08	0	0	0	0	0	0	0	0,2	0	0	0	0	0	0	0,1	0

Şekil 3.7 : YYKKDNYK peptit diziliminin AKKY ile kodlanması.

Ağırlık ve konum bilgileri, 20 standart amino asit için 1*20'şer uzunluklukta vektörlerden oluşur. Dolayısıyla AKKY 1*40 büyüklüğündeki vektörlerden meydana gelir.

3.4 n-grams Öznitelik Kodlama Yöntemi

n-grams öznitelik kodlama yöntemi, kalıntı çiftlerinin sıklığını bulmaya dayalı bir öznitelik çıkarımı yöntemidir. n-grams yönteminde P peptit diziliminde ardışık n tane kalıntı sayısı aranır. Bu sayılar n ardışık kalıntı çiftinin, üçlüsünün, dörtlüsünün vb. meydana gelme sıklığını verir [35]. Bu durumda öznitelik sayısı 20^n olur. Bu tez çalışmasında n=2 ye göre öznitelik kodlaması yapılmıştır.

YYKKDNYK şeklinde verilen örnek peptit üzerinde n = 2'ye göre n-grams öznitelik kodlamasını açıklayalım:

Bu durumda 20^2 uzunlukta bir öznitelik vektörü ile mevcut tüm ikili kalıntı ihtimalleri temsil edilebilir. Peptit dizilimi meydana getiren 7 adet ikili, ardışık kalıntı çiftinin sayısı elde edilir. Buna göre;

YY = 1,
YK = 2,
KK = 1,
KD = 1,
DN=1,
NY = 1 olur.

Elde edilen değerler Şekil 3.8'de görüldüğü gibi 1x400 büyüklüğündeki öznitelik vektöründe karşılık gelen yere yerleştirilir.

1	2	3	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	400
AA	AC	AD	..	..	DN	..	..	KD	KK	..	NY	..	..	YK	..	..	..	YY
0	0	0	0	0	1	0	0	1	1	0	1	0	0	2	0	0	0	1

Şekil 3.8 : YYKKDNYK peptit diziliminin n-grams yöntemine göre kodlanması.

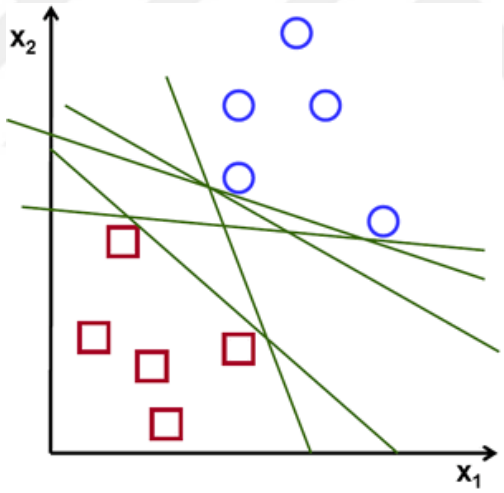
Proteini oluşturan amino asit sayısı artsa bile öznitelik sayısı daima 20^n dir. n-grams kodlamanın protein dizilimindeki amino asit sayısından bağımsız olması önemli bir üstünlüğüdür.

3.5 Sınıflandırma Algoritmaları

Yardımcı T hücreleri / BDUK molekülleri bağlanma konumları tespitinde DVM ve RO makine öğrenmesi sınıflandırma algoritmaları kullanılmıştır.

3.5.1 Destek vektör makineleri

Destek Vektör Makineleri (DVM), örüntü tanıma ve sınıflandırma problemlerinin çözümü için 1979 yılında Vapnik tarafından geliştirilmiştir [36]. DVM'nin çalışma prensibi iki sınıfı birbirinden en uygun şekilde ayırabilen hiper düzlemin tanımlanması esasına dayanmaktadır.

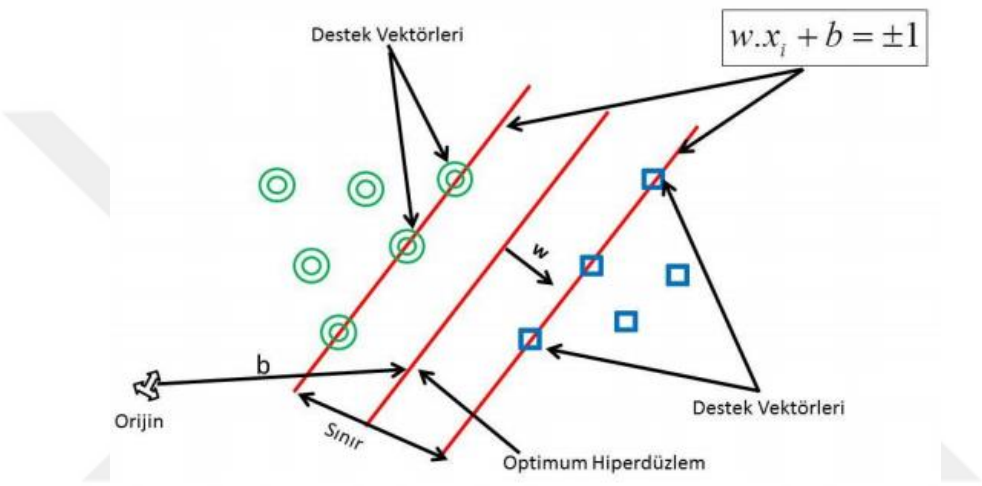


Şekil 3.9: Ayırıcı hiper düzlemler [36].

Şekil 3.9'de görüldüğü gibi problem uzayını, tüm örnekleri başarılı sınıflandıracak şekilde ayıran birçok doğru bulunabilir. Bu doğrulardan hangisinin diğerlerinden daha uygun olduğunu belirlemek için bir kritere ihtiyaç duyulmaktadır. Belirlenen doğru bütün noktalara yakın olmamalıdır. Çünkü bu doğrunun kullanılarak yapıldığı sınıflandırma gürültüye çok duyarlı olacak ve yeterince genelleyici olmayacaktır. Bu yaklaşıma göre en iyi ayırıcı doğru her noktadan mümkün olduğunca uzaktan geçen olmalıdır. DVM'nin çalışma mekanizması, eğitim verilerinden kendine en yakın

örnek için en uzak mesafaye sahip ayırıcı hiper düzlemi bulmaktır. Şekil 3.10'da optimum hiper düzlem gösterilmiştir [37].

Sınıflandırma için bir düzlemde bulunan iki grup arasına bir sınır çizilerek iki grubu ayırmak mümkündür. DVM algoritması ait olduğu sınıfın sınırını belirler [38]. Destek vektörlerinin üzerinde bulunduğu ve noktalarla gösterilmiş düzlemlere sınır düzlemleri denir. Sınır düzlemlerinin tam ortasından geçen ve her iki düzleme de eşit uzaklıkta bulunan düzlem ise hiper düzlem olarak ifade edilir [39].



Şekil 3.10 : Doğrusal olarak ayrılabilen iki sınıflı DVM ve optimum hiperdüzlem[39]

Örneğin iki sınıflı, r sayıda örnekten oluşan ve birbirinden ayrılabilir (x1,y1),……,(xr ,yr) şeklinde tanımlanan bir eğitim verisi olmak üzere; burada x ϵ R^N olup N boyutlu bir uzayı, y ϵ {+1,-1} olup sınıf etiketlerini temsil etmektedir ve bu iki sınıf çeşitli N-1 boyutlu hiperdüzlemler tarafından ayrılabilmektedir [41].
Bir hiperdüzlem denklem (3.5)'deki gibi tanımlanır.

$$w.xi+b=0 \qquad (3.5)$$

Denklemde, x_i hiperdüzlem üzerindeki noktayı, w hiperdüzlemin normalini ve b ise hiperdüzlemin orijinden olan uzaklığı olarak ifade edilen biası (eğilim değeri) temsil etmektedir.

İki sınıflı doğrusal olarak ayrılabilen veriler için, ayrım yapan hiperdüzlem;

$w.x_i+b \geq +1$, her y = +1 için (3.6)

$w.x_i+b \leq -1$, her y= -1 için (3.7)

olarak tanımlanabilir.

(3.6) ve (3.7)'deki iki eşitsizlik tek bir eşitsizlik haline getirilirse;

$y_i\ (w.x_i+b) -1 \geq 0$, $y_i \in \{1,-1\}$ (3.8)

denklem (3.8)'deki gibi olur.

Optimum hiperdüzleme paralel ve $w.x_i+b = \pm 1$ eşitliği ile tanımlanan ayrım yapabilen iki hiper düzlem üzerinde bulunan ve sınır genişliğini belirleyen noktalar destek vektörleri olarak adlandırılır [42] [43]. Bu iki hiper düzlem arasındaki sınır genişliği (marjin) dir. (3.9)

$$\frac{2}{||w||} \tag{3.9}$$

İki sınıfı maksimum sınır genişliği ile ayıran optimum hiperdüzlem, $||w||^2$ ifadesinin (3.8)'deki sınırlamalara bağlı olarak minimum hale getirilmesi ile bulunabilir [40] [44].

$$\min\left[\frac{1}{2}||w||^2\right] \tag{3.10}$$

Bu problem Lagrange denklemleri kullanılarak çözülebilecek bir optimizasyon problemidir ve çözüm için karar fonksiyonu;

$$f(x) = sign\left(\sum_{i=1}^{r} \lambda_i.y_i(x.x_i) + b\right) \tag{3.11}$$

şeklinde olur (3.11).

Burada λ_i, i=1, 2, …, r Lagrange çarpanlarını[45] temsil etmektedir.

3.5.2 Rastgele orman

RO, 2001 yılında Leo Breiman ve Adele Cutler tarafından geliştirilmiş olan içerisinde oylama yöntemi barındıran bir sınıflama algoritmasıdır. Birden çok karar ağacının biraya gelmesiyle oluşur ve her bir ağacın verdiği oy değerlendirilerek kazanan sınıf belirlenir. Karar ağaçları, birbirinden bağımsızdır ve veri setinden bootstrap tekniği ile çekilen örneklerden oluşturulur [47][48] .

Diğer bütün karar ağaçları sınıflandırıcılarında olduğu gibi RO yönteminde de dallanma kriterlerinin belirlenmesi ve uygun bir budama yönteminin seçilmesi gerekmektedir. Dallanma kriterlerinin belirlenmesinde Gini indeksi yöntemi kullanılmaktadır. Gini indeksi işin temelinde sınıf özniteliklerinin zayıflık derecesini ölçmektedir [48].

RO yönteminde kullanıcıların belirtmesi gereken parametreler vardır. Bu parametre değerleri, her bir düğümde kullanılacak örneklerin sayısı ve oluşturulacak ağaç sayısıdır. Yani bir sınıflandırma işlemi sırasında karar ormanı, kullanıcı tarafından belirlenen K adet ağaçtan oluşturulur. Yeni bir sınıflandırma işleminde ise veri seti bu K adet karar ağacından geçirilir ve RO sınıflandırıcısı ile K adet ağaçtan elde edilen K adet oy arasından en fazla oya sahip olan sınıf seçilir [49][50].

RO yönteminde, üretilen modeli test etmek için ayrı bir veri seti yoksa ya da orijinal veri setinden test veri seti ayrılmamış ise, sınıf dağılımı dikkate alınmak suretiyle orijinal veri setinin 2/3'ü öğrenme verisi (inBag), 1/3'ü ise test verisi (Out-Of-Bag(OOB)) olarak ayrılır [51]. Karar ormanını oluşturacak K tane karar ağacı için, gene K adet bootstrap tekniği kullanılarak örneklem oluşturulur ve her bir örneklem için inBag ve OOB verisi ayrılır. Tüm ağaçlar ayrılan OOB verisi ile test edilerek hata oranı hesaplanır. Bu hata oranlarının ortalaması alınarak karar ormanının OOB hatası hesaplanır. Ortaya çıkan bu OOB hata oranına göre tüm ağaçlara bir ağırlık verilir. Hata oranı ve ağaca verilen ağırlık değeri ters orantılıdır. Hata oranı en yüksek olan karar ağacı en düşük ağırlığı, hata oranı en düşük olan karar ağacı ise en yüksek ağırlığı alır. Belirlenen ağırlıklara göre tüm ağaçlar sınıflandırılmaları için bir oylama işleminden geçirilir. En yüksek oyu alan ağacın sınıf bilgisi tahmini olarak belirlenmiş olur.

4. ÖNERİLEN ÖZNİTELİK KODLAMA YÖNTEMLERİ

Yardımcı T hücreleri / BDUK molekülleri bağlanma konumlarının tahmini için BloFTKY ve BloAKKY adları verilen iki yöntem geliştirilmiştir.

4.1 BloFTKY

Yardımcı T hücrelerinin Mhc moleküllerine bağlanma durumlarını belirlemek için kalıntıların biyokimyasal etkileşimlerini anlamak gerekir. Bu temel düşünceden yola çıkarak ikinci bölümde bahsi geçen Aaindex tablosunda bulunan toplam 544 özellik kullanılarak bir öznitelik kodlama yöntemi geliştirilmiştir.

Blosum 50 yer değiştirme matrisi ve amino asitlerin fizikokimyasal özellikleri birlikte kullanıldığından BloFTKY olarak isimlendirilmiştir. Yer değiştirme matrisleri amino asitlerin doğada birbiri yerine geçme sıklığını ifade eder, fizikokimyasal benzerlikleri veya farklılıkları hakkında bilgiler içermez. Bu durum verinin içeriğinin tam olarak temsil edilememesine neden olur. BloFTKY'de bu eksikliğin üstesinden gelmek için seçilen en iyi fizikokimyasal özellik, kalıntıyı temsil eden öznitelik vektöründe yerine yerleştirildi. Böylece Blosum 50 matrisi ile amino asitlerin fizikokimyasal indeks değerlerini birleştirerek öznitelik vektörlerinin sınıflandırıcı algoritmalar tarafından daha iyi tanınması sağlandı.

BloFTKY iki aşamadan oluşmaktadır. Birinci aşamada Aaindex tablosunda bulunan 544 fizikokimyasal özellikten veri seti için en iyi olan fizikokimyasal özellik belirlenmiştir. İkinci aşamada en iyi fizikokimyasal özellik kullanılarak BloFTKY matrisi elde edilmiştir.

Birinci aşama 5 adımdan oluşur. İlk adımda Aaindex tablosundaki her bir fizikokimyasal özellik değeri z-skor [52] yöntemine göre normalize edilmiştir. İkinci adımda 9 kalıntıdan oluşan her bir peptit dizilimi birimdik (BKY) olarak kodlanarak 1*180 boyutunda öznitelik vektörleri elde edilmiştir. Üçüncü adımda BKY, Blosum 50 yer değiştirme matrisinin i.satır ve sütunu ile çarpılarak yeni 1x180 büyüklüğünde öznitelik vektörleri elde edilmiştir. Dördündü adımda Blosum 50 matrisine göre

kodlanmış öznitelik vektörleri ile 544 tane fizikokimyasal özelliğin her biri sırasıyla çarpılarak her bir set için 544 tane öznitelik vektörü oluşturulur. Bu işlem tüm veri setleri için uygulanmıştır. Dördüncü adımda elde edilen öznitelik vektörleri doğrusal DVM ile sınıflandırılarak her bir fizikokimyasal özellik için sınıf doğruluğu değerleri elde edilmiştir. Son adımda elde edilen sınıf doğruluğu değerleri büyükten küçüğe doğru sıralanarak en iyi fizikokimyasal özellik belirlenmiştir.

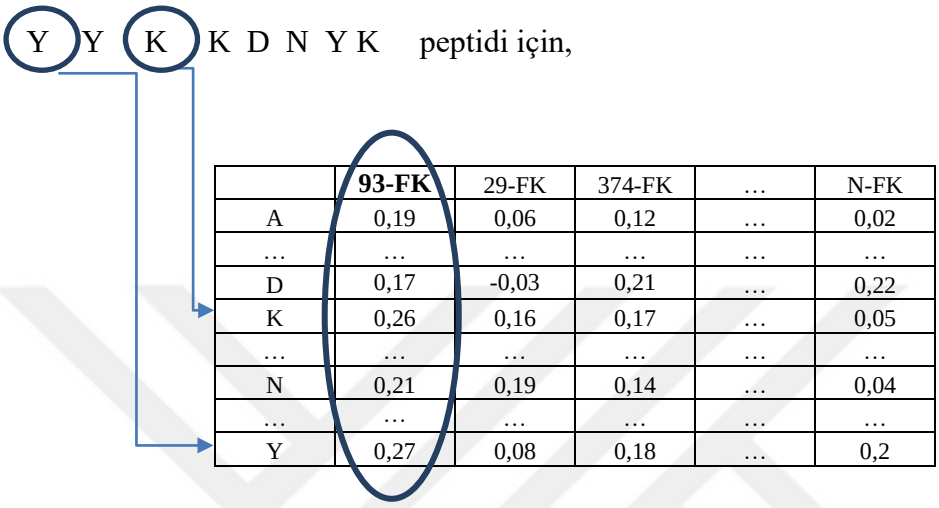
İkinci aşamada ise, elde edilen en iyi fizikokimyasal özelliğin indeks değerindeki özellik peptit diziliminde ait olduğu kalıntının yerine yazılmış ve BloFTKY matrisi elde edilmiştir.

BloFTKY'yi, YYKKDNYK şeklinde verilen örnek peptit üzerinde açıklayacak olursak; ilk aşamada Şekil 4.1'de görüldüğü gibi tüm 544 amino asit indeks değerleri normalize edilmiş ve sırasıyla Blosum 50 ile çarpılmış olan BKY öznitelik vektörüne yerleştirilir. Şekil 4.1'de 100.fizikokimyasal özellik görülmektedir.

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8*0.15	0
Y																		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8*0.15	0
Y																		
0	0	0	0	0	0	0	0	0	0	0	6*0.21	0	0	0	0	0	0	0
K																		
0	0	0	0	0	0	0	0	0	0	0	6*0.21	0	0	0	0	0	0	0
K																		
0	0	0	8*0.19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D																		
0	0	7*0.26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N																		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8*0.15	0
Y																		
0	0	0	0	0	0	0	0	0	0	0	6*0.21	0	0	0	0	0	0	0
K																		

Şekil 4.1 : 100-Fk isimli özelliğin Blosum 50 matrisinin i.satır ve sütunu ile çarpılması.

Her bir fizikokimyasal özelliğe göre kodlanan öznitelik vektörleri doğrusal DVM ile sınıflandırılarak sınıf doğrulukları elde edilir. Bu sınıf doğruluğu değerleri büyükten küçüğe sıralanarak en yüksek fizikokimyasal özellik belirlenir. İkinci aşamada Şekil 4.2'de görülen en iyi fizikokimyasal özelliğe göre BloFTKY öznitelik vektörleri elde edilir.



	93-FK	29-FK	374-FK	…	N-FK
A	0,19	0,06	0,12	…	0,02
…	…	…	…	…	…
D	0,17	-0,03	0,21	…	0,22
K	0,26	0,16	0,17	…	0,05
…	…	…	…	…	…
N	0,21	0,19	0,14	…	0,04
…	…	…	…	…	…
Y	0,27	0,08	0,18	…	0,2

Şekil 4.2 : BloFTKY için en iyi fizikokimyasal özelliğin büyükten küçüğe sıralanarak belirlenmesi.

Şekil 4.2'de görüldüğü gibi en yüksek sonuçları veren fizikokimyasal özellik belirlenir. Son aşama olarak peptit dizilimlerindeki amino asit değerleriyle çarpılarak BloFTKY vektörü elde edilir. Şekil 4.3'te BloFTKY öznitelik vektörü görülmektedir.

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8*0.27	0
Y																		

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8*0.27	0
Y																		

0	0	0	0	0	0	0	0	0	0	0	6*0.26	0	0	0	0	0	0	0
K																		

0	0	0	0	0	0	0	0	0	0	0	6*0.26	0	0	0	0	0	0	0
K																		

0	0	0	8*0.17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D																		

0	0	7*0.21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N																		

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8*0.27	0
Y																		

0	0	0	0	0	0	0	0	0	0	0	6*0.26	0	0	0	0	0	0	0	0
K																		

Şekil 4.3 : En iyi fizikokimyasal özellik ile BloFTKY öznitelik vektörü kodlanması.

Çalışmada kullanılmış olan veri setlerinin her biri 9'ar peptitden oluştuğu için BloFTKY öznitelik vektörünün büyüklüğü 1*180 olur.

4.2 BloAKKY

Ağırlık ve konum tabanlı öznitelik çıkarım yöntemine dayanarak oluşturulmuştur. Peptit dizilimlerinin ağırlıkları ve konumları bilgisi üzerine Blosum 50 matrisinin eklendiği bir öznitelik çıkarım yöntemidir. BloAKKY yönteminin kalıntıdaki amino asit sayısından bağımsız bir şekilde kodlanması büyük bir üstünlüğüdür. Peptit dizilimi içerisindeki amino asit sayısı kaç olursa olsun oluşturulan öznitelik vektörleri 1*40 boyutunda olacaktır.

BloAKKY dört adımdan oluşur. İlk adımda peptit dizilimi amino asite karşılık gelen birimdik vektörü çarpılır. İkinci adımda peptit dizilimini oluşturan her bir amino asitin ağırlığı (a_i) ve konumu (k_i) hesaplanır. Ağırlık ve konum bilgileri, 20 standart amino asit için 1*20'şer uzunluklukta vektörlerden oluşur. Dolayısıyla ağırlık ve konum bilgileri 1*40 uzunluğunda vektörlerden meydana gelir. Üçüncü adımda birimdik olarak kodlanmış olan amino asitler ikinci adımda hesaplanan ağırlık ve konum bilgileri ile çarpılarak üçüncü bölümde bahsi geçen AKKY öznitelik vektörleri elde edilir. Son olarak dördüncü adımda BloAKKY, Blosum 50 matrisinin i. satır ve sütun değerlerinin AKKY vektörlerinde karşılık gelen sırada çarpılmasıyla oluşur.

BloAKKY'yi YYKKDNYK şeklinde verilen bir örnek peptit üzerinde açıklayalım. Şekil 4.4'te görüldüğü gibi ilk adımda her bir peptit dizilimi birimdik olarak kodlanır.

AĞIRLIK																			KONUM																			
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Şekil 4.4: YYKKDNYK peptidinin birimdik olarak kodlanması.

İkinci adımda ağırlık ve konum bilgileri hesaplanır.

$a_Y = 3/8 = 0,3$ $k_Y = 10/56 = 0,1$

$a_K = 3/8 = 0,3$ $k_K = 15/56 = 0,2$

$a_D = 1/8 = 0,1$ $k_D = 5/56 = 0,08$

$a_N = 1/8 = 0,1$ $k_N = 6/56 = 0,1$ olarak bulunur.

Üçüncü adımda hesaplanan ağırlık ve konum bilgileri birimdik olarak kodlanmış vektörlerle çarpılarak Şekil 4.5'te görülen AKKY öznitelik vektörleri elde edilir.

AĞIRLIK																				KONUM																	
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
0	0	0,1	0,1	0	0	0	0	0	0	0	0,3	0	0	0	0	0	0	0,3	0	0	0	0,1	0,08	0	0	0	0	0	0	0	0,2	0	0	0	0	0	0	0,1	0

Şekil 4.5 : YYKKDNYK peptit diziliminin AKKY ile kodlanması.

Son olarak dördüncü adımda, Blosum 50 matrisinin i. satır ve sütunundaki değerler AKKY vektörleri ile çarpılarak Şekil 4.6'da görülen BloAKKY öznitelik vektörleri elde edilir.

$a_Y = 0,37 * 8 = 3$ $k_Y = 0,17 * 8 = 1,3$

$a_K = 0,37 * 6 = 2,2$ $k_K = 0,26 * 6 = 1,5$

$a_D = 0,12 * 8 = 0,9$ $k_D = 0,08 * 8 = 0.6$

$a_N = 0,12 * 7 = 0,8$ $k_N = 0,1 * 7 = 0,7$ olarak hesaplanır.

AĞIRLIK																				KONUM																	
A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
0	0	0,8	0,9	0	0	0	0	0	0	0	2,2	0	0	0	0	0	0	3	0	0	0	0,7	0,6	0	0	0	0	0	0	0	1,5	0	0	0	0	0	0	1,3	0

Şekil 4.6 : YYKKDNYK peptit diziliminin BloAKKY ile kodlanması.

Bu işlem bütün AKKY vektörleri için yapılarak BloAKKY matrisi elde edilir.

4.3 Deneysel Sonuçlar

Yardımcı T hücreleri / BDUK molekülleri özgünlüklerinin belirlenmesinde geliştirilen BloFTKY ile BloAKKY yöntemleri, sınıflandırma algoritmaları ile GV, OV ve TV veri setlerine uygulanmıştır.

Tez kapsamında yapılan deneysel çalışmalarda, BloFTKY ve BloAKKY için sınıf doğruluğu, duyarlık, özgünlük ve (Matthews Korelasyon Katsayısı) MKK performans metrikleri elde edilmiştir.

Performans metrikleri Şekil 4.7'da görülen karmaşıklık matrisinden elde edilir:

		Öngörülen Sınıf	
		Sınıf =1	Sınıf=0
Doğru Sınıf	Sınıf=1	DA	YE
	Sınıf=0	YA	DE

Şekil 4.7 : Karmaşıklık matrisi.

DA; doğru tahmin edilen birinci sınıfa ait örneklerin, DE; doğru tahmin edilen ikinci sınıfa ait örneklerin, YA; yanlış tahmin edilen birinci sınıfa ait örneklerin, YE; yanlış tahmin edilen birinci sınıfa ait örneklerin sayısıdır.

Doğruluk, hasta ve sağlam kişilerin yüzde kaçının önerilen yöntemle tanınabildiğini gösterir [53] (4.1).

$$Doğruluk = \frac{\mathrm{DA+DE}}{\mathrm{DA+YA+DE+YE}} \tag{4.1}$$

Duyarlık, hastalığın gerçekten var olduğu bilinen kişilerden yüzde kaçının önerilen yeni yöntemle tanınabildiğini gösterir (4.2).

$$Duyarlık = \frac{\mathrm{DA}}{\mathrm{DA+YE}} \tag{4.2}$$

Özgünlük değeri, hastalığı taşımayanların yüzde kaçının önerilen yöntemle tanınabildiğini gösterir (4.3).

$$Özgünlük = \frac{\mathrm{DE}}{\mathrm{DE + YA}} \tag{4.3}$$

MKK, iki sınıflı problemlerde modelin gücünü belirtir. En önemli özelliği, sınıflardaki veri sayısı karışık olduğunda diğer kriterlere göre daha doğru sonuç vermesidir. -1 ile 1 arasında değerler alır. 1 en iyi tahmini, 0 şansa bağlı bir tahmin yapıldığını, -1 ise ters tahmin yapıldığını belirtir (4.4).

$$MKK = \frac{\text{DA x DE} - \text{YAxYE}}{\sqrt{(\text{DA}+\text{YA})(\text{DA}+\text{YE})(\text{DE}+\text{YA})(\text{DE}+\text{YE})}} \quad (4.4)$$

BloFTKY 'nin ilk aşamasında en iyi fizikokimyasal özellik DVM algoritaması ile GV veriseti üzerinde belirlenmiştir. Her iki yöntemde de sınıflandırma testleri Weka yazılımı [54] ortamında 10-kat Çapraz Doğrulama Test Tekniği (ÇDT)'ye göre gerçekleştirilmiştir. ÇDT ile veriler 10 ayrı kümeye ayrılır ve 1'i test, 9'u eğitim için kullanılır. Her çevrimde sınıf doğruluğu elde edilir. 10 çevrim sonunda ortalama sınıf doğruluğu hesaplanır.

GV, OV ve TV üzerinde yapılan deneysel çalışmalar sonucunda Çizelge 4.1'deki 93 numaralı fizikokimyasal özelliğin bizim problemimiz için en iyi sonucu verdiğini gördük. Fizikokimyasal özelliklerin indeks değerleri ortalama 0, varyans 1 olacak şekilde z-skor yöntemine göre normalize edilerek BloFTKY kodlanmıştır.

Çizelge 4.1 : BloFtky için elde edilen en iyi fizikokimyasal özellik değerleri.

Amino asit	93-Fk
A	0,07
R	0,05
N	0,07
D	0,12
C	0,07
Q	0,07
E	0,12
G	0,09
H	0,07
I	0,07
L	0,07
K	0,05
M	0,07
F	0,07
P	0,94
S	0,07
T	0,07

W	0,07
Y	0,07
V	0,072

BloFTKY'nin fizikokimyasal özellikleri içeren bir kodlama yöntemi olması amino asitlerin biyokimyasal özelliklerinin modellemeye yansıtılması açısından önemlidir. Bununla beraber BloFTKY'nin birkaç kısıtı bulunmaktadır. Öncelikle BloFTKY'de ilk aşamada en iyi özellik seçilirken her bir amino asit bağımsız olarak kodlamaya dahil edilmektedir. Yani amino asitler arasındaki bağımlılık görmezden gelinmektedir. Ayrıca BloFTKY amino asitlerin dizilim içindeki ağırlık ve konumları hakkında bilgi içermez. Dolayısıyla bu durum örneklerinin sınıflandırıcı tarafından açık olarak tanımlanmayışına yol açar [20].

Deneyler k-Eyk, BayesNet, NaiveBayes, C4.5, RO ve Doğrusal DVM üzerinde denenmiş ancak en yüksek sonucu doğrusal DVM ile RO sınıflandırıcı algoritmaları vermiştir. Aşağıdaki tablolarda DVM ile RO sınıflandırıcı algoritmaları ile test edilmiş verilerden elde edilen performans metrikleri görülmektedir.

Çizelge 4.2 : GV verileri üzerinde öznitelik kodlama yöntemlerinin Doğrusal DVM algoritması başarım sonuçları.

	BKY				n-grams				BloFTKY				BloAKKY			
ALEL	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	Mkk	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK
A0101	94,97	0,71	0,979	0,72	88,05	0,265	0,954	0,266	93,39	0,618	0,972	0,63	91,19	0,382	0,975	0,455
A0201	94,77	0,96	0,936	0,89	81,55	0,818	0,813	0,630	95,24	0,962	0,942	0,90	87,83	0,889	0,867	0,756
A0202	93,90	0,94	0,929	0,87	76,66	0,779	0,753	0,532	94,97	0,969	0,903	0,90	86,17	0,883	0,839	0,723
A0203	92,04	0,93	0,902	0,84	75,01	0,751	0,749	0,499	94,12	0,956	0,925	0,88	86,36	0,872	0,855	0,726
A0206	92,84	0,912	0,938	0,85	82,11	0,759	0,857	0,615	95,57	0,958	0,954	0,90	88,58	0,843	0,911	0,754
A0301	88,79	0,805	0,919	0,72	71,64	0,504	0,795	0,294	90,33	0,854	0,922	0,76	82,85	0,683	0,883	0,586
A1101	91,14	0,895	0,926	0,82	69,61	0,684	0,706	0,389	90,33	0,908	0,911	0,82	80,68	0,789	0,822	0,611
A2402	80,05	0,478	0,882	0,37	77,41	0,391	0,871	0,273	81,52	0,435	0,912	0,38	85,04	0,522	0,934	0,501
A2601	96,31	0,533	0,988	0,6	92,61	0	0,98	-0,03	95,56	0,4	0,988	0,50	94,09	0	0,996	-0,02
A3101	92,44	0,842	0,951	0,8	81,64	0,588	0,891	0,492	93,73	0,877	0,957	0,83	88,76	0,711	0,946	0,687
A3301	96,79	0,583	0,99	0,65	92,53	0,222	0,966	0,208	97,25	0,583	0,995	0,70	95,42	0,361	0,989	0,463
A6801	92,82	0,91	0,94	0,85	71,53	0,645	0,762	0,406	92,30	0,923	0,923	0,84	81,02	0,748	0,851	0,602
A6802	93,65	0,8	0,966	0,77	85,04	0,526	0,92	0,467	92,71	0,726	0,97	0,74	90,09	0,568	0,973	0,629
B0702	90,78	0,844	0,925	0,74	78,64	0,511	0,863	0,374	90,29	0,756	0,944	0,71	81,06	0,422	0,919	0,389
ORTALAMA	92,23	**0,79**	0,941	0,74	80,3	0,53	0,848	0,39	**92,67**	0,78	**0,944**	**0,75**	87,09	0,63	0,911	0,56

Çizelge 4.2'de görüldüğü gibi GV verileri üzerinde gösterilen yardımcı T hücreleri / BDUK molekülleri bağlanma yerleri tahmininde doğrusal DVM sınıflandırıcısı ile %92,67 ile en iyi doğruluk, 0,944 ile en iyi özgünlük ve 0,75 ile en iyi MKK sonucunu BloFTKY vermiştir. %80,3 sınıf doğruluğu, 0,53 duyarlık ve 0,39 MKK sonucu açısından en düşük başarımı ise n-grams yöntemi vermiştir.

Çizelge 4.3 : OV verileri üzerinde öznitelik kodlama yöntemlerinin Doğrusal DVM algoritması başarım sonuçları.

	BKY				n-grams				BloFTKY				BloAKKY			
ALEL	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK
A0201	79,22	0,91	0,23	0,18	73,11	0,857	0,156	0,013	82,16	0,987	0,067	0,13	*Na*	*Na*	*Na*	*NaN*
A0202	66,48	0,80	0,218	0,02	65,68	0,776	0,264	0,040	73,73	0,923	0,067	0,07	*Na*	*Na*	*Na*	*NaN*
A0203	64,08	0,73	0,444	0,18	58,13	0,701	0,333	0,034	64,60	0,789	0,349	0,14	64,59	0,92	0,079	-0,001
A0206	75,44	0,86	0,351	0,23	68,34	0,788	0,311	0,097	74,26	0,909	0,149	0,08	*Na*	*Na*	*Na*	*NaN*
A0301	67,8	0,79	0,311	0,1	68,25	0,791	0,34	0,131	71,43	0,872	0,217	0,11	*Na*	*Na*	*Na*	*NaN*
A1101	74,83	0,87	0,231	0,11	69,46	0,818	0,187	0,005	79,14	0,963	0,088	0,09	*Na*	*Na*	*Na*	*NaN*
A3101	63,51	0,76	0,282	0,04	60,89	0,737	0,262	-0,04	67,98	0,853	0,214	0,08	*Na*	*Na*	*Na*	*NaN*
A3301	54,72	0,68	0,306	-0,01	60,19	0,705	0,417	0,123	54,22	0,682	0,292	-0,02	60,69	0,783	0,292	0,0833
A6801	63,14	0,76	0,26	0,02	64,76	0,769	0,302	0,072	67,75	0,857	0,167	0,03	*Na*	*Na*	*Na*	*NaN*
A6802	57,42	0,71	0,325	0,03	59,71	0,67	0,463	0,13	61,71	0,775	0,325	0,1	60,57	0,872	0,114	-0,021
B1501	73,76	0,85	0,152	0,003	76,23	0,876	0,182	0,062	78,21	0,911	0,121	0,04	*Na*	*Na*	*Na*	*NaN*
ORTALAMA	67,31	0,796	0,282	0,082	65,9	0,77	**0,292**	-0,35	**70,47**	**0,866**	0,186	**0,77**	61,95	0,858	0,161	0,0204

Çizelge 4.3'te görüldüğü gibi OV verileri üzerinde gerçekleştirilen yardımcı T hücreleri / BDUK molekülleri bağlanma yerleri tahmininde %70,47 ile en iyi doğruluk, 0,866 ile en iyi duyarlık sonucunu ve 0,77 ile en iyi MKK sonucunu BloFTKY vermiştir. BloAKKY pek çok alelde duyarlık değerini 1, özgünlüğü 0 olarak yanlış hesaplamaktadır. Bunun nedeni tüm peptitleri bağlanabilir kabul etmesinden kaynaklanmaktadır. Nitekim bu durum karmaşıklık matrisinden de görülmektedir. Bu nedenle bu yöntem ilgili alellere uygulanamaz. Diğer bir ifade ile sadece A0203, A3301, A6802 alelerine uygulanabilir. 0,77 duyarlık ve -0,35 MKK sonucu açısından en düşük başarımı n-grams yöntemi vermiştir.

Çizelge 4.4 : TV verileri üzerinde öznitelik kodlama yöntemlerinin Doğrusal DVM algoritması başarım sonuçları.

	BKY				n-grams				BloFTKY				BloAKKY			
ALEL	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK
A0101	91,48	0,59	0,954	0,56	87,49	0,35	0,94	0,313	91,82	0,55	0,964	0,56	89,24	0,104	0,99	0,209
A0201	88,22	0,89	0,872	0,76	76,27	0,745	0,777	0,521	88,08	0,89	0,869	0,76	77,25	0,753	0,788	0,540
A0202	81,05	0,83	0,788	0,62	66,26	0,672	0,653	0,325	79,86	0,82	0,769	0,61	71,19	0,726	0,697	0,423
A0203	80,66	0,82	0,784	0,61	61,25	0,646	0,577	0,223	80,88	0,84	0,768	0,62	70,14	0,723	0,679	0,401
A0206	84,59	0,84	0,849	0,69	69,25	0,624	0,748	0,375	83,25	0,82	0,838	0,67	74,34	0,682	0,793	0,478
A0301	85,85	0,74	0,905	0,65	72,99	0,449	0,841	0,304	85,41	0,71	0,908	0,64	77,07	0,454	0,896	0,390
A1101	86,23	0,84	0,874	0,71	72,27	0,634	0,781	0,418	86,33	0,85	0,87	0,71	76,45	0,682	0,819	0,505
A2402	72,37	0,55	0,8	0,35	65,16	0,426	0,75	0,175	73,27	0,54	0,817	0,36	74,47	0,505	0,849	0,371
A2601	92,03	0,36	0,964	0,36	89,20	0,145	0,95	0,106	93,19	0,26	0,984	0,35	*Na*	*Na*	*Na*	*NaN*
A3101	86,58	0,68	0,927	0,64	76,03	0,427	0,875	0,328	86,03	0,67	0,923	0,62	80,41	0,455	0,924	0,437
A3301	86,81	0,54	0,935	0,51	81,95	0,345	0,916	0,293	87,05	0,48	0,949	0,51	83,54	0,069	0,992	0,17
A6801	76,12	0,76	0,761	0,52	62,52	0,606	0,644	0,249	76,63	0,76	0,768	0,53	68,28	0,659	0,705	0,364
A6802	79,84	0,59	0,888	0,51	72,23	0,46	0,84	0,318	80,34	0,58	0,9	0,52	75,47	0,437	0,897	0,381
B0702	88,64	0,69	0,928	0,61	75,66	0,303	0,854	0,157	89,17	0,68	0,935	0,63	86,21	0,487	0,942	0,481
B0801	94,78	0,26	0,971	0,22	95,77	0,087	0,987	0,105	96,62	0,13	0,994	0,22	*Na*	*Na*	*Na*	*NaN*
B1501	86,83	0,57	0,932	0,53	75,24	0,247	0,862	0,115	87,33	0,53	0,947	0,53	83,69	0,148	0,987	0,273
B2705	94,38	0,64	0,971	0,62	91,28	0,432	0,955	0,398	94,89	0,59	0,98	0,63	93,28	0,321	0,987	0,439
B3501	76,85	0,61	0,839	0,46	68,62	0,505	0,772	0,277	78,97	0,63	0,865	0,51	69,68	0,396	0,839	0,257
B4001	91,95	0,44	0,96	0,42	88,97	0,245	0,944	0,197	92,87	0,39	0,974	0,43	*Na*	*Na*	*Na*	*NaN*
B4402	70,75	0,61	0,757	0,37	69,81	0,539	0,787	0,332	66,98	0,52	0,75	0,28	66,98	0,355	0,846	0,230
B4403	64,31	0,39	0,768	0,17	62,44	0,408	0,732	0,143	63,38	0,31	0,796	0,12	64,78	0,31	0,817	0,143
B5101	74,50	0,59	0,811	0,41	73,95	0,565	0,815	0,381	74,79	0,51	0,847	0,38	73,11	0,463	0,847	0,330
B5301	78,59	0,71	0,833	0,53	71,54	0,559	0,803	0,369	80,56	0,69	0,868	0,57	73,24	0,559	0,829	0,402
B5801	94,23	0,60	0,972	0,6	88,46	0,218	0,943	0,171	94,54	0,48	0,985	0,58	*Na*	*Na*	*Na*	*NaN*
ORTALAMA	83,65	0,63	0,876	0,51	75,61	0,44	0,822	0,28	**83,84**	**0,66**	**0,886**	**0,52**	76,44	0,46	0,856	0,361

Çizelge 4.4'te görüldüğü gibi TV verileri üzerinde yapılan yardımcı T hücreleri / BDUK molekülleri bağlanma yerleri tahmininde doğrusal DVM sınıflandırıcısında % 83,84 ile en iyi doğruluk, 0,66 ile en iyi duyarlık ve 0,52 ile en iyi MKK sonucunu BloFTKY vermiştir. %75,61 sınıf doğruluğu , 0,44 duyarlık, 0,82 özgünlük ve 0,28 MKK değeri ile en düşük başarımı n-grams yöntemi vermiştir.

Çizelge 4.5 : GV verileri üzerinde öznitelik kodlama yöntemlerinin RO algoritması başarım sonuçları.

ALEL	BKY				n-grams				BloFTKY				BloAKKY			
	Doğ. (%)	Duy.	Özg.	MKK	Doğ. (%)	Duy.	Özg.	Mkk	Doğ. (%)	Duy.	Özg.	MKK	Doğ. (%)	Duy.	Özg.	MKK
A0101	91,19	0,176	1	0,401	*Na*	*Na*	*Na*	*NaN*	92,13	0,265	1	0,493	90,56	0,235	0,986	0,359
A0201	92,3	0,964	0,879	0,847	82,51	0,867	0,779	0,651	92,11	0,958	0,881	0,844	83,84	0,86	0,815	0,676
A0202	93,35	0,966	0,899	0,868	81,32	0,855	0,768	0,627	92,45	0,959	0,888	0,850	83,66	0,859	0,813	0,673
A0203	93,37	0,952	0,914	0,868	77,84	0,813	0,741	0,556	93,18	0,96	0,902	0,865	79,16	0,802	0,802	0,583
A0206	92,67	0,88	0,954	0,842	81,77	0,708	0,881	0,602	93,01	0,889	0,954	0,849	85,51	0,75	0,916	0,684
A0301	88,35	0,667	0,964	0,692	74,94	0,293	0,919	0,272	90,32	0,724	0,97	0,746	82,19	0,577	0,913	0,525
A1101	90,14	0,886	0,914	0,801	70,62	0,654	0,751	0,407	89,33	0,877	0,907	0,785	82,89	0,794	0,859	0,655
A2402	83,57	0,217	0,993	0,388	81,81	0,159	0,985	0,284	83,87	0,232	0,993	0,403	84,16	0,319	0,974	0,111
A2601	*Na*	*Na*	*Na*	*NaN*	*Na*	*Na*	*Na*	*NaN*	*Na*	*Na*	*Na*	*NaN*	*Na*	*Na*	*Na*	*NaN*
A3101	89,84	0,684	0,968	0,714	80,12	0,289	0,968	0,379	90,06	0,675	0,974	0,827	88,33	0,684	0,948	0,672
A3301	*Na*	*Na*	*Na*	*NaN*	94,35	0	0,998	-0,09	*Na*	*Na*	*Na*	*NaN*	94,96	0,139	0,997	0,301
A6801	89,23	0,819	0,94	0,773	68,46	0,445	0,843	0,317	87,43	0,8	0,923	0,736	81,28	0,69	0,894	0,604
A6802	87,47	0,337	0,991	0,500	85,61	0,242	0,989	0,396	88,97	0,4	0,995	0,574	87,47	0,411	0,975	0,506
B0702	88,83	0,511	0,994	0,650	78,64	0,133	0,969	0,188	89,81	0,578	0,988	0,682	88,34	0,533	0,981	0,630
ORTALAMA	90,02	0,671	**0,95**	0,695	79,83	0,454	0,882	0,382	**90,22**	**0,69**	0,947	**0,72**	85,56	0,588	0,913	0,536

Çizelge 4.5'te görüldüğü gibi GV verileri üzerinde yapılan yardımcı T hücreleri / BDUK molekülleri bağlanma yerleri tahmininde RO sınıflandırıcısında %90,22 ile en iyi doğruluk, 0,69 ile en iyi duyarlık ve 0,72 ile en iyi MKK sonucunu BloFTKY vermiştir. Özgünlük sonucunda en yüksek başarımı BKY vermiştir. 0,38 duyarlık, 0,89 özgünlük ve 0,33 MKK sonucu ile en düşük başarımı n-grams yöntemi vermiştir. Sınıf doğruluğu, duyarlık, özgünlük ve MKK sonucu açısından en düşük başarımı ise n-grams yöntemi vermiştir.

Çizelge 4.6 : OV verileri üzerinde öznitelik kodlama yöntemlerinin RO algoritması başarım sonuçları.

	BKY				n-grams				BloFTKY				BloAKKY			
ALEL	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK	Doğ.(%)	Duy.	Özg.	MKK
A0201	81,49	0,99	0,007	-0,03	80,82	0,98	0,007	-0,028	81,35	0,99	0	-0.038	80,95	0,98	0	-0,04
A0202	75,87	0,99	0	-0,04	73,72	0,95	0,034	-0,029	76,4	0,99	0	-0.028	74,53	0,96	0,03	-0,08
A0203	70,8	0,93	0,23	0,247	64,59	0,87	0,17	0,064	69,76	0,93	0,21	0.212	66,92	0,91	0,16	0,11
A0206	76,62	0,98	0	-0,07	76,92	0,97	0,02	0,012	76,33	0,97	0	-0.071	77,81	0,96	0,10	0,14
A0301	75,73	0,98	0,047	0,081	73,92	0,94	0,07	0,046	75,28	0,97	0,04	0.059	72,56	0,94	0,03	-0,03
A1101	*Na*	*Na*	*Na*	*NaN*	79,13	0,97	0,02	0,002	80,21	0,99	0	-0.022	79,78	0,99	0	-0,03
A3101	70,6	0,96	0,01	-0,07	69,81	0,94	0,02	-0,052	71,12	0,97	0	-0.083	71,91	0,94	0,11	0,09
A3301	61,69	0,86	0,167	0,04	65,67	0,93	0,16	0,152	60,19	0,86	0,12	-0.009	65,17	0,86	0,27	0,16
A6801	72,62	0,97	0,01	-0,04	72,89	0,95	0,09	0,086	73,17	0,98	0,01	-0.016	72,35	0,96	0,05	0,02
A6802	60,28	0,86	0,114	-0,03	60,85	0,85	0,16	0,017	60,57	0,87	0,11	-0.026	58,28	0,82	0,13	-0,05
B1501	*Na*	*Na*	*Na*	*NaN*	83,16	0,99	0	-0,031	83,16	0,99	0	-0.031	*Na*	*Na*	*Na*	*NaN*
ORTALAMA	71,74	0,94	0,06	0,009	72,86	0,944	0,07	0,0217	**73,42**	**0,97**	0,04	-0,004	72,02	0,93	**0,08**	**0,029**

Çizelge 4.6 ‘te görüldüğü gibi OV verileri üzerinde gösterilen yardımcı T hücreleri / BDUK molekülleri bağlanma yerleri tahmininde RO sınıflandırıcısında BloFTKY yöntemi %73,42 doğruluk, 0,97 duyarlık; BloAKKY yöntemi 0,08 özgünlük ve 0,02 MKK sonuçlarıyla ile en yüksek başarımı yakalamışlardır. %71,74 sınıf doğruluğu, 0,06 özgünlük sonuçları ile BKY en düşük başarımı vermiştir.

Çizelge 4.7 : TV verileri üzerinde öznitelik kodlama yöntemlerinin RO algoritması başarım sonuçları.

	BKY				n-grams				BloFTKY				BloAKKY			
ALEL	Doğ,	Duy,	Özg,	MKK	Doğ,	Duy,	Özg,	MKK	Doğ,	Duy,	Özg,	MKK	Do,	Duy,	Özg,	MKK
A0101	91,21	0,276	0,991	0,434	88,57	0,018	0,993	0,041	91,07	0,288	0,988	0,428	88,64	0,153	0,977	0,216
A0201	85,2	0,862	0,844	0,703	71,29	0,671	0,747	0,418	84,53	0,855	0,838	0,690	75,78	0,716	0,792	0,509
A0202	78,16	0,820	0,742	0,564	65,98	0,682	0,637	0,319	79,36	0,826	0,76	0,588	70,07	0,729	0,671	0,401
A0203	78,19	0,813	0,749	0,563	63,16	0,664	0,597	0,261	79,25	0,824	0,759	0,585	67,11	0,691	0,65	0,341
A0206	81,97	0,777	0,854	0,634	69,04	0,618	0,749	0,371	80,84	0,771	0,839	0,612	71,8	0,635	0,785	0,426
A0301	82,52	0,557	0,931	0,5423	73,39	0,228	0,933	0,229	82,17	0,553	0,928	0,533	79,24	0,505	0,905	0,454
A1101	82,94	0,774	0,866	0,642	68,37	0,51	0,799	0,323	83,22	0,777	0,869	0,649	77,77	0,695	0,832	0,533
A2402	71,17	0,342	0,873	0,250	69,81	0,218	0,907	0,171	70,87	0,342	0,869	0,244	70,42	0,312	0,875	0,223
A2601	*Na*	*Na*	*Na*	*NaN*	92,66	0	0,999	-0,09	*Na*	*Na*	*Na*	*NaN*	*Na*	*Na*	*Na*	*NaN*
A3101	84,07	0,527	0,949	0,549	74,87	0,145	0,957	0,175	84,67	0,537	0,953	0,567	80,7	0,48	0,92	0,452
A3301	84,71	0,182	0,983	0,298	81,7	0,02	0,98	-0,01	84,87	0,177	0,986	0,306	83,45	0,217	0,961	0,262
A6801	74,37	0,718	0,768	0,486	58,51	0,547	0,621	0,168	76,21	0,74	0,782	0,523	66,86	0,645	0,69	0,336
A6802	77,3	0,374	0,952	0,423	71,45	0,335	0,885	0,262	77,51	0,374	0,955	0,430	73,572	0,353	0,907	0,318
B0702	88,5	0,563	0,954	0,573	82,27	0,118	0,974	0,173	88,2	0,555	0,952	0,562	86,94	0,45	0,959	0,493
B0801	*Na*	*Na*	*Na*	*NaN*	*Na*	*Na*	*Na*	*NaN*	*Na*	*Na*	*Na*	*NaN*	*Na*	*Na*	*Na*	*NaN*
B1501	85,16	0,275	0,977	0,384	81,23	0,033	0,982	0,041	85,16	0,269	0,978	0,383	83,01	0,148	0,978	0,236
B2705	94,18	0,309	0,998	0,515	93,88	0,272	0,998	0,481	93,98	0,284	0,998	0,492	93,78	0,259	0,998	0,468
B3501	77,32	0,495	0,905	0,447	71,68	0,337	0,896	0,284	77,08	0,491	0,903	0,442	73,32	0,421	0,881	0,342
B4001	92,2	0,128	0,989	0,224	91,54	0,021	0,991	0,033	92,12	0,106	0,99	0,198	91,37	0,032	0,988	0,048
B4402	71,69	0,526	0,824	0,365	68,39	0,316	0,89	0,254	75	0,553	0,86	0,438	72,64	0,539	0,831	0,387
B4403	68,07	0,352	0,845	0,224	65,72	0,183	0,894	0,108	67,13	0,352	0,831	0,205	71,83	0,437	0,859	0,327
B5101	75,91	0,435	0,9	0,383	74,22	0,324	0,924	0,317	75,91	0,481	0,88	0,394	78,15	0,472	0,916	0,444
B5301	76,33	0,591	0,86	0,471	71,54	0,354	0,917	0,338	76,61	0,567	0,877	0,473	74,92	0,559	0,855	0,436
B5801	92,17	0,077	0,996	0,195	91,55	0	0,996	-0,02	92,5	0,128	0,996	0,282	92,58	0,154	0,993	0,297
ORTALAMA	81,51	**0,489**	0,897	**0,448**	75,68	0,287	0,881	0,202	**81,72**	0,359	**0,923**	0,367	78,36	0,436	0,873	0,361

Çizelge 4.7’de görüldüğü gibi TV verileri üzerinde yapılan yardımcı T hücreleri / BDUK molekülleri bağlanma yerleri tahmininde RO sınıflandırıcısı %81,72 ile en iyi doğruluk ve 0,923 ile en iyi özgünlük sonucunu BloFTKY, 0,489 ile en iyi duyarlık ve 0,448 ile en iyi MKK sonucunu BKY vermiştir. %75,68 sınıf doğruluğu, 0,287 duyarlık ve 0,202 MKK sonucu açısından en düşük başarımı n-grams yöntemi vermiştir.

5. SONUÇLAR VE ÖNERİLER

İnsan vücudu her an dışarıdan gelen ve sürekli değişiklik gösteren çok sayıda mikroorganizma ile karşı karşıya kalır. Vücuda yabancı ve zararlı olan en küçük mikroorganizmalar dahi bağışıklık sistemi tarafından fark edilir. Fark edilen zararlı mikroorganizmalar bir seri kimyasal reaksiyon sonucunda yok edilir. Bağışıklık sisteminin harekete geçmesine zararlı mikroorganizmalara ait olan antijen proteinleri neden olur. Uzun antijen proteinleri ASH tarafından T hücreleri ile birleşebilmeleri için daha küçük peptit parçacıklarına ayrılırlar. Böylece T hücreleri bu antijen peptitlerine bağlanabilirler. Bu antijenik peptitlere Epitop adı verilir. Bağışıklık sisteminin tetiklenmesi için önce yardımcı T hücrelerinin epitopları tanıması gereklidir. Yardımcı T hücreleri bu tanıma işlemini BDUK molekülleri ile gerçekleştirirler. BDUK molekülleri T hücreleri ile hücre yüzeyinde bağlanırlar. Bu bağlanma esnasında epitoplar yardımcı T hücresine tanıtılır.

Bu tez çalışmasında yardımcı T hücresi / BDUK molekülleri bağlanma noktolarının tespiti için BloFTKY ve BloAKKY adı verilen iki protein öznitelik kodlama yöntemi geliştirilmiştir. BloFTKY'de önce proteini temsil eden amino asitlerin en iyi fizikokimyasal özelliği belirlenmiştir. Belirlenen özellik Blosum 50 yer değiştirme matrisi ile çarpılarak yöntem geliştirilmiştir. BloAKKY'de ise protein dizilimi içindeki amino asitlerin ağırlık ve konumları hesaplanmıştır. Sonrasında hesaplanan bu değerler Blosum 50 yer değiştirme matrisinde karşılık gelen amino asit değerleri ile çarpılmıştır. Böylece proteinler sayısal olarak ifade edilmişlerdir.

Her iki öznitelik çıkarım yöntemi yardımcı T hücreleri /BDUK molekülleri bağlanma durumlarının tespiti için k-eyk, BayesNet, NaiveBayes, doğrusal ve polinom DVM, C4.5 ve RO sınıflandırıcı algoritmaları ile test edilmiştir.

Elde edilen deneysel sonuçlara göre BloFTKY ve BloAKKY öznitelik kodlama yöntemleri doğrusal DVM ve RO algoritmaları ile en yüksek başarımı vermişlerdir. Fizikokimyasal özelliklerin yöntemde kullanılması BloFTKY yönteminin her üç veri

setinde başarılı sonuçlara ulaşılmasında önemli rol oynamıştır. OV veri seti üzerinde yapılan testlerde BloFTKY yöntemi BKY yöntemine göre %3'lük bir artışla başarılı olmuştur. BloFTKY yönteminin hem doğrusal DVM sınıflandırıcısında hemde RO sınıflandırıcısında BloAKKY yönteminden daha başarılı olduğu görülmüştür.

Gelecekte yapılacak çalışmalarda, BloAKKY öznitelik kodlama yöntemine amino asitlerin fizikokimyasal etkileşimlerini tanımlayan bilgilerin eklenmesi planlanmaktadır. Böylece öznitelik kodlama yönteminin proteinleri sınıflandıırcıya daha iyi tanıtabileceği ve böylece sınıflandırıcı başarımının artacağı öngörülmektedir. BloFTKY yönteminde Blosum 50 yer değiştirme matrisinin yanı sıra PAM ve Blosum matrislerinin negatif değerlerden arındırılmış hali olan VOGG yer değiştirme matrislerinin kullanılması planlanmaktadır. Üçüncü olarak geliştirilen öznitelik kodlama yöntemlerinin uygulanmasını sağlayan çevrimiçi bir web uygulamasının yapılması planlanmaktadır.

KAYNAKLAR

[1] **Abbas, A.K., Lichtman, A.H., Pillai, S.** (2007). Molecular and Cellular Immunology (6.bs.). Philadelphia: Saunders Elsevier.

[2] **Gök, M.,** Hıv-1 Proteaz Enziminin Kesme Konumlarının Tespitinde Yeni Öznitelik Vektörleri,Sakarya Üniversitesi . Sakarya,Türkiye.

[3] **Chaplin, D.D.,** (2010). Overview of the Immune Response. J Allergy ClinImmunol. 125, 3-23.

[4] **Gök,M., Özcerit, A.T**.(2011): OETMAP: a new feature encoding scheme for MHC class I binding prediction, Mol, Cell, Biochem, (2011). 1–6.

[5] **M.,J., Zvelebil, G.,J., Barton, WR., Taylor, M.,J., Sternberg.,** Prediction of protein secondary structure and active sites using the alignment of homologous sequences, J, Mol, Biol, 195 (4).(1987). 957–961.

[6] **Lundegaard,C., Lamberth,K., Harndahl,M., Buus,S., Lund,O.,Nielse,M., (**2008). NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. Nucleic Acids Research, 2008, Vol. 36, Web Server issue W509–W512 doi:10.1093/nar/gkn202.

[7] **MacNamara A., Kadolsky U., Bangham CRM., Asquith B.,** (2009). T-Cell Epitope Prediction:Rescaling Can Mask Biological Variation between MHC Molecules. PLoS Comput Biol 5(3): e1000327. doi:10.1371/journal.pcbi.1000327.

[8] **Roomp,K., Antes,I., Lengauer,T.,** Predicting MHC class I epitopes in large datasets. BMC Bioinformatics ,2010 11:90.

[9] **He, J.,Yang,G., Rao, H., Li, Z., Ding,X. Chen, Y.** (2012). Prediction of human major histocompatibility complex class II binding peptides by continuous kernel discrimination method. Artificial Intelligence in Medicine 55 (2012) 107–115.

[10] **Lee, B., & Richards, F. M.** (1971). The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology*, *55*(3),379-IN4.

[11] **Atanasova,M., Atanas Patronov, A., Ivan Dimitrov, I., Flower, D. R. and Doytchinova, I.** (2012). EpiDOCK: a molecular docking-based tool for MHC class II binding prediction. Protein Engineering, Design & Selection vol. 26 no. 10 pp. 631–634, 2013 Published online May 9, 2013 doi:10.1093/protein/gzt018.

[12] **Oseroff, C., Sidney, J., Tripple, V., Grey, H., Wood, R., Broide, D. H., Greenbaum, J., Kolla, R., Peters, B., Pomés, A. and Sette,A,.** (2012). Analysis of T Cell Responses to the Major Allergens from German Cockroach: Epitope Specificity and Relationship to IgE Production. J *Immunol* 2012; 189:679-688; Prepublished online 15 June 2012; doi: 10.4049/jimmunol.1200694.

[13] **Meydan, C., Otu, H. H., Sezerman, O.U**, Prediction of peptides binding to MHC class I and II alleles by temporal motif mining. BMC Bioinformatics ,2013,14 (Suppl 2):S13.

[14] **Karosiene, E., Michael Rasmussen, M., Blicher,T., Lund, O., Buus, S., Nielsen, M.** (2013). NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics (2013) 65:711–724 .DOI 10.1007/s00251-013-0720-y.

[15] **Birnbaum, M. E., Mendoza, J. L., Sethi, D. K., Dong, S., Glanville, J., Dobbins, J., Özkan, E., Davis, M.M., Wucherpfennig, K. W., and Garcia K. C**.(2010). Deconstructing the Peptide-MHC Specificity of T Cell Recognition. Cell 157, 1073–1087.

[16] **Schwaiger,J., Aberle, J. H., Stiasny,K., Knapp,B., Schreiner,W., Fae,I., Fischer,G., Scheinost,O., Chmelik,V., Heinz, F. X.** (2015). Specificities of Human CD4_ T Cell Responses to an Inactivated Flavivirus Vaccine and Infection: Correlation with Structure and Epitope Prediction. Journal of Virology p. 7828–7842.

[17] **WATSON, J.D., CRICK, F.H.C.,** Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature, 171,pp. 737-738, 1953.

[18] **BARNES, M., GRAY, I.,** Bioinformatics for Genetics. John Wiley & Sons Inc, 2003.

[19] **Url-1** <http://www.kimyasalgelismeler.com/kimya-kutuphanesi/kimya-dallari/fizikokimya.html> ,alındığı tarih: 25.04.2016.

[20] **KAWASHIMA, S., KANEHISA, M.,** AAindex: amino acid index database, Nucleic Acids Res. 20 (1): 374, 2000.

[21] **Dalva, K., Beksaç, M.,** (1999). Mononüklear Hücrelerin Maturasyonu. Ş.(Ed)(s.153-159). Temel Ve Klinik Mikrobiyoloji. Ankara. Güneş Kitabevi.

[22] **Chaplin, DD.,** (2010). Overview of the Immune Response. J Allergy ClinImmunol. 125, 3-23.

[23] **Peter, J., Delves, D., Ivan, M.,** (2000). The Immune System. N Engl J Med, 343,37-49.

[24] **Kuby, H.,** (1997). Overview of The Immune System: Cells and Organs of The Immune System. Immunology. (3. bs)(s.1-83). USA: W.H. Freeman And Company.

[25] **Punt, J.A., Singer, A.T.,** (1997). Cell Development. Rich, RR., Fleisher, TA, Schwartz,B.D., Shearer, W.T., Strober, W.(Ed). Clinical Immunology Principles and Practice (1.bs) (s.157-176). St Louis: Mosby.

[26] **Bonilla, FA.,Oettgen HC**. (2010). Adaptive immunity. J Allergy Clin Immunol. 125,33-40.

[27] **Kılıçturgay, K.,** (2003). Kan Hücrelerinin Gelişimi. (s.15-51) .İmmünoloji. Bursa: Nobel&Güneş Kitabevi.

[28] **Abbas KA, Andrew HL, Pillai S.,** Cellular and Molecular Immunology. 6'ncı_baskı. Philadelphia, Saunders Elsevier, 2007;189-214.

[29] **Choo SY.,** The HLA system: Genetics, immunology, Clinical testing and Clinical Implications.Yonsei Med. Journal. 2007, 48:11-23.

[30] **B. Peters, J. Sidney, P. Bourne, H.H. Bui, S. Buus, G. Doh, W. Fleri, M. Kronenberg,R. Kubo, O. Lund, D. Nemazee, J.V. Ponomarenko, M. Sathiamurthy, S. Schoenberger, S. Stewart, P. Surko, S. Way, S. Wilson, A. Sette,** The immune epitope database and analysis resource:from vision to blueprint, PLoS Biol. 3 (3) (2005) e91.

[31] **Gök, M.** (2013). A Comparison of Several Feature Encoding Techniques for MHC Class I Binding Prediction, International Journal of Bioscience, Biochemistry and Bioinformatics, Vol. 3, No. 2, March 2013.

[32] **Ölmez, T., Dokur Ölmez, Z.** (2009).Uzman sistemlerde örüntü tanıma: Yapay sinir ağları, Genetik algoritmalar,bulanık mantık,makine öğrenmesi, İ.T.Ü Elektrik – Elektronik Fakültesi.

[33] **RÖGNVALDSSON, T., L., YOU.** (2004) Why neural networks should not be used for HIV-1 protease cleavage site prediction. Bioinformatics 20(11): 1702- 1709.

[34] **NANNI L., LUMINI A**., (2006). A reliable method for HIV-1 protease cleavage site prediction methods. Neuro Computing 69: pp. 838-841.

[35] **WU, C., WHITSON, G.,** (1992). Protein Classification Artificial Neural System. Protein Science 1(5): pp. 667-677.

[36] **V. Vapnik**, (1995) .The nature of statistical learning theory, Springer, New York.

[37] **CORINNA, C. and VLADIMIR, V.N.,** (1995) Support-Vector Networks, Machine Learning, vol.20, no.3, s.273-297.

[38] **Burges, C. J. C.,** 1998, A tutorial on support vector machines for pattern recognition, data mining and knowledge discovery, Kluwer Academic Publishers, 2 (2), 121-167.

[39] **A.Sevgi.** (2013). Kaba Küme ve Destek Vektör Makineleri Kullanılarak Nitelik İndirgeme ve Sınıflandırma Problemlerinin Çözümü İçin Bütünleşik Bir Yaklaşım. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü.Doktora Tezi.

[40] **Huang, C., Davis L. S.,** vd. (2002). "An assessment of support vector machines for land cover classification." International Journal of Remote Sensing 23(4): 725-749.

[41] **Mathur, A., Foody, G.M.** (2008). "Multiclass and binary SVM classification: Implications for training and classification users." IEEE Geoscience and Remote Sensing Letters (5):241–245.

[42] **Kavzoglu, T. ve Colkesen, I.** (2009). "A kernel functions analysis for support vector machines for land cover classification." International Journal of Applied Earth Observation and Geoinformation 11(5): 352-359.

[43] **Mathur, A. ve Foody G. M.** (2008). "Crop classification by support vector machine with intelligently selected training data for an operational application." International Journal of Remote Sensing 29(8): 2227-2240.

[44] **Song, X., Duan Z., at al**. (2011). "Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image." International Journal of Remote Sensing 33(10): 3301-3320.

[45] **Gunn, S. R.,**1998, Support vector machines for classification and regression, Technical Report, Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science.

[46] **Hastie, T., Tibshirani, R., & Friedman, J.** (2009). The elements of statistical learning: data mining, inference and prediction: Springer-Verlag.

[47] **Breiman, L.** (2001). Random Forest [Eloktronik Sürüm]. Machine Learning, 45, 5-32.

[48] **Mather, P.M.,** Computer Processing Of Remote-Sensed Images. John Wiley And Sons Ltd., 1987.

[49] **Çölkesen, İ.,** 2009, Uzaktan Algılamada İleri Sınıflandırma Tekniklerinin Karşılaştırılması ve Analizi, Yüksek Lisans Tezi, Gebze, 2009.

[50] **Pal, M.,** Random Forest Classifier For Remote Sensing Classification. International Journal Of Remote Sensing, 26, 217-222, 2005.

[51] **Akman, M.,** Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama, Yüksek Lisans Tezi, Ankara.Üniversitesi, Ankara, 2010.

[52] **JAIN A., NANDAKUMAR K.,**(2005) Score normalization in multimodal biometric systems, Pattern Recognition, 2005.

[53] **Url-2** <http://kanser.org/saglik/upload/17.kanser_Kongresi/Tani_Testlerinin_Degerlendirilmesi %23Rian_Disci.pdf>, alındığı tarih : 04.05.2016.

[54] **Frank E., Hall MA., Holmes G., Kirkby R., Pfahringer B, Witten, Trigg, L.,** Weka-a machine learning workbench for data mining, In: Maimon O, Rokach L (eds), The Data Mining and Knowledge Discovery Handbook, Springer,2005, 1305-14.

ÖZGEÇMİŞ

Ad Soyad : İlknur ÇINAR EFE

Doğum Yeri ve Tarihi : İstanbul-02/02/1987

Adres : Esentepe Mah. Karasu Sok. No:23/7 Kartal /İstanbul

E-Posta : ilknurcinar34@gmail.com

Lisans : Marmara Üniversitesi / Elektronik-Bilgisayar Eğitimi Bölümü / (2011)

Yüksek Lisans : Yalova Üniversitesi (2016)

TEZDEN TÜRETİLEN YA YINLAR/SUNUMLAR

- **Çınar İ.**, Gök M., A New Feature Encoding Technique For MHC Class I Molecules / T Cells Binding Specificity, International Symposium on Health Informatics and Bioinformatics (HIBIT) October 16-17/ 2015 Muğla, Turkey.