

Trombofili hastalığı ile genetik bozukluklar arasındaki ilişkinin adaptif ağ tabanlı bulanık mantık çıkarım sistemi (ANFIS) ile tespit edilmesi

Bülent Haznedar^{1*}, Adem Kalınlı²

29.06.2015 Geliş/Received, 29.08.2015 Kabul/Accepted

ÖZ

Kanın damar içinde ya da bir organda pıhtılaşması olan Tromboz'a yol açan hastalıklarda önemli sağlık sorunları ortaya çıkmakta ve hatta birçok vakada insanlar hayatını kaybedebilmektedir. Tromboz gelişimi multifaktoriyel olup, çok sayıda edinsel ve kalıtsal faktörün değişik mekanizmalarla tromboz oluşumuna neden olduğu bilinmekte ise de tromboza yatkınlık olarak bilinen Trombofili tanısının konmasında önemli zorluklar yaşanmaktadır. Bu kapsamda, geleneksel sınıflandırma yöntemlerinin klinik, laboratuvar ve genetik tetkiklere ait verilerin değerlendirilmesindeki başarımları ise çoğunlukla sınırlı kalabilmektedir. Bu çalışmada, Trombofili hastalığı ile genetik bozukluklar arasındaki ilişkinin tespit edilebilmesi için Adaptif Ağ Tabanlı Bulanık Mantık Çıkarım Sistemi (ANFIS) kullanılmış ve elde edilen sonuçlar literatürde yaygın olarak kullanılan bazı sınıflandırma algoritmalarına ait sonuçlarla karşılaştırılmıştır. Elde edilen sonuçlardan, ANFIS ile elde edilen sonuçların daha başarılı olduğu görülmüştür.

Anahtar Kelimeler: ANFIS, bulanık mantık, Trombofili

Detection of the relationship between thrombophilia disease with genetic disorders by adaptive neuro-fuzzy inference system (ANFIS)

ABSTRACT

Thrombosis is a condition involving the clotting of bloods in the veins or in an organ, whereas thrombophilia is a term used to describe a predisposition for thrombosis. In diseases causing major health problems that are related to thrombosis, people may even lose their lives in many cases. Thrombosis is multifactorial; it is known to cause a number of acquired and hereditary factors which lead to thrombosis formation through various mechanisms. Therefore, many difficulties are experienced in the diagnosis of thrombophilia. In this context, traditional statistical methods are often inadequate for the evaluation of clinical and laboratory data. In this study, in order to determine the relationship between genetic disorders and thrombophilic disease, the Adaptive Neuro-Fuzzy Inference System (ANFIS) is used and our results are compared to the results of some of the commonly-used classification algorithms. Simulation results showed that the results from using ANFIS were more successful than those obtained from considered classification algorithms.

Keywords: ANFIS, neuro-fuzzy, fuzzy logic, Thrombophilia

* Sorumlu Yazar / Corresponding Author

1 Hasan Kalyoncu Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Gaziantep – bulent.haznedar@hku.edu.tr

2 Erciyes Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Kayseri - kalinlia@erciyes.edu.tr

1. GİRİŞ (INTRODUCTION)

İnsanda damar hasarından sonra kanamanın durabilmesi için gerçekleşen normal pıhtılaşma sürecine Hemostaz denir. Kanın bir atardamar veya toplardamar içinde ya da bir organda pıhtılaşmasına ise Tromboz, pıhtıya ise trombus denir. Trombofili (Thrombo-philia: trombozu sevme) ise tromboza eğilim yaratan tabloları tanımlamakta kullanılan bir terimdir. Tromboz gelişimi multifaktöriyel olup, çok sayıda edinsel ve kalıtsal faktörün değişik mekanizmalarla tromboz oluşumuna neden olduğu bilinmektedir [1-19]. Tromboz genellikle bir damarın iç yüzündeki bir bozukluk sonucu meydana gelir. Bu bozukluk o tabakada oluşan bir yozlaşma (aterom, travma) nedeniyle olabileceği gibi, kaza sonucu, ameliyatlar veya bazı zehirler, ilaçlar, mikrobik hastalıklar, bazı kan bozuklukları ve uzun süre yatakta kalmayı gerektiren ya da hastayı zayıf düşüren hastalıklar da Tromboza yol açabilmektedir. Bu hastalıklarda kanın pıhtılaşmasına bağlı damar tıkanıkları kalp, akciğer ve beyin gibi organlarda pıhtı oluşması, gebeliklerin düşük veya ölü doğum ile sonuçlanması gibi önemli problemler yaşanmaktadır.

İleri yaş, sigara içme, hipertansiyon, diyabet, östrojen kullanımı, atriyal fibrilasyon, varisler, gebelik, genel cerrahi girişimler ve travmalar gibi edinsel pek çok faktörün Tromboza neden olabileceği bilinmektedir [8-12, 20]. Kalıtsal pıhtılaşma bozuklukları ise genel olarak basit tanımlamayla kanın pıhtılaşmaya eğilim gösterdiği bazı hastalıkları kapsamaktadır ve genetik bozukluklara bağlı olarak irsi yani kuşaktan kuşağa geçebilen ve doğuştan edinilmiş hastalıklardır. Kalıtsal Trombofilide, Antitrombin eksikliği, Protein C eksikliği, Protein S eksikliği, Faktör V Mutasyonu ve bazı gen mutasyonlarının etkili olduğu gösterilmiştir [2-6]. Bunlar dışında her geçen gün kalıtsal trombofiliye neden olduğu iddia edilen bozukluklar tanımlanmaktadır. Bununla beraber günümüzde hala kalıtsal trombofili düşünülen vakaların %40-60'ında tüm incelemelere rağmen hastalığın nedenini ortaya koymak mümkün olamamaktadır [7].

Trombofilinin multifaktöriyel nedenlere bağlı olmasından dolayı tanı konmasında zorluklar yaşanmaktadır. Örneğin klinik olarak Derin Ven Trombozu (DVT) tanısı konulan hastaların çok az kısmında gerçek hastalık bulunduğu ve bazı yüksek riskli olgularda ise lokal bulgu ve semptomlar olmadan DVT geliştiği görülmektedir. Yine bazı hastalarda ani gelişen ve ölümcül seyirli Pulmoner Emboli (Akciğer embolisi) görülebilmektedir [3].

Klinik tanısal yaklaşımda, hastaların öyküsüne ilave olarak fiziksel muayene ve rutin laboratuvar testleri yapılmaktadır. Seçilen hasta alt guruplarına ise kalıtsal

trombofili sebeplerine yönelik genetik testler uygulanmaktadır. Ancak kalıtsal trombofili tanısı için yapılacak testler oldukça zahmetlidir ve pahalıdır [21]. Ayrıca, uygun testler kullanılmazsa yanıltıcı sonuçlar elde edilebileceğinden bu hastalara uygulanacak testlerin doğru seçilmesi de önemlidir.

Çalışmalar, çoklu parametrelere bağlı kompleks hastalıkların tanısının tahmin edilmesinde, genel popülasyonda ilgili risklerin genliğinin düşük olması ve genotiplerin tamamlanmamış özellikleri (incomplete penetrance of genotypes) nedeniyle genetik tetkiklerin sınırlı bilgi sağlandığını göstermektedir. Diğer taraftan, bu tür çalışmalar genellikle aynı anda bir genin etkisini incelemektedir. Ancak, bu tür kompleks hastalıkların tanısında genetik ve çevresel faktörlerin birlikte değerlendirilmesinin hastalığın etiyolojisi için daha anlamlı ve önemli sonuçlar sağlayacağına inanılmaktadır [22, 23]. Bununla birlikte, çoklu parametrelerin trombofili hastalığının tanısındaki etkilerinin araştırılmasına yönelik çalışmalar ise henüz sınırlı düzeydedir [24, 25]. Ayrıca henüz kişinin venöz tromboz riskinin doğru tahmin edilmesine yönelik risk modelleri de mevcut değildir ve bu konu oldukça zordur [26]. Venöz trombozun matematiksel olarak tahmin edilmesine yönelik yalnızca birkaç çalışma bulunmaktadır. Çalışmalar göstermektedir ki, kişisel emboli riskinin etkili bir şekilde tahmin edilebilmesi ancak klinik, çevresel ve genetik değişkenleri de dikkate alan sofistike modeller ile mümkün olabilecektir [27]. Ancak bu konularda yapılan çalışmalar ise henüz sınırlı düzeydedir.

Çoklu hastalık durumları, hastalara ait verilerin çokluğu ve bu veriler arasında çoğunlukla doğrusal olmayan ilişkiler bulunması gibi nedenlerden dolayı, klinik ve laboratuvar verilerinin değerlendirilmesinde geleneksel sınıflandırma algoritmalarının başarımları sınırlı kalabilmektedir. Hastalıkların sınıflandırılması, görüntülerden bilgi çıkarımı, hastalık seyrinin kestirimi, tedavi planlama ve hasta verileri arasındaki bilinmeyen ilişkilerin tespit edilmesi gibi tıbbi uygulamaların zorluğu araştırmacıları daha modern ve güçlü yöntemler kullanmaya sevk etmiştir. Bu amaçla, zor problemlerin çözümünde araştırmacıların ilgisi giderek artan bir oranda yapay zeka yöntemleri üzerine yoğunlaşmaya başlamıştır.

Penco ve arkadaşları Yapay Sinir Ağları (YSA) kullanarak venöz trombozun önemli risk faktörlerinin belirlenmesine yönelik bir çalışma yapmışlardır. Yaptıkları çalışmada YSA'nın diğer analitik metotlara göre daha başarılı sonuçlar verdiğini göstermişlerdir [28]. Shanthi ve ark. beyinde trombolik tıkanma olasılığını tahmin etmek için çok katmanlı ileri beslemeli YSA modeli kullanmışlardır. Hastanın yaş, cinsiyet,

hipertansiyon, diyabet, sigara ve alkol kullanımı, kalp damar tıkanıklığı, görüntü kaybı gibi 25 parametrenin dikkate alındığı bu çalışmada YSA'nın beyinde trombolik tıkanıklık ihtimalinin belirlenmesinde başarılı bir şekilde kullanılabileceği gösterilmiştir [29]. Dahabiah ve ark. Klinisyenlerin daha önce değerlendirdiği ultrason görüntülerinden venöz trombolinin karakterize edilmesi üzerine bir çalışma yapmışlardır [30]. Markus ve ark. Doppler ultrason sinyallerinin sinyal ilmeme teknikleri ile değerlendirilerek embolik sinyallerin tespiti üzerine bir çalışma yapmışlardır. Dikkate alınan veriler üzerinde başarılı sonuçlar elde edilmesine rağmen, dikkate alınan yöntem çevrim için değerlendirme yapabilecek bir otomasyon sistemi olmaması nedeniyle geliştirilmeye muhtaçtır [31]. Kemeny ve ark. Yaptıkları çalışmada doppler ultrason sinyallerinin kullanarak emboli tespiti için YSA kullanmışlardır [32]. Yaptıkları çalışmada sonuçların çok başarılı olmadığını ancak bu başarısızlıkta uygun sinyal-gürültü oranının uygun seçilmemesinin etkili olduğunu ifade etmişlerdir. Bununda bazı eşikleme algoritmaları ile aşılabileceğini ve emboli tespitinde otomasyon sistemlerinin geliştirilmesine büyük ihtiyaç duyulduğunu ifade etmişlerdir.

Bu çalışmada, Trombofili hastalığı tanısı konmuş bu hastaların klinik verileri ile genetik bozukluk olup olmama durumu arasındaki ilişkinin belirlenmesi amacıyla Adaptif Ağ Tabanlı Bulanık Mantık Çıkarım Sistemi (ANFIS)'nin kullanılması araştırılmıştır. Elde edilen sonuçlar sınıflandırma algoritmalarına ait sonuçlarla karşılaştırılarak yorumlanmıştır.

Makalenin ikinci Bölümünde, Adaptif Ağ Tabanlı Bulanık Mantık Çıkarım Sistemi (ANFIS) hakkında bilgi verilmiş, üçüncü bölümde gerçekleştirilen uygulama ve elde edilen sonuçlar izah edilmiş ve dördüncü bölümde sonuçlar tartışılmıştır.

2. ADAPTİF AĞ TABANLI BULANIK MANTIK ÇIKARIM SİSTEMİ (ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM, ANFIS)

Adaptif ağ tabanlı bulanık mantık çıkarım sistemi (ANFIS), Sugeno tipi bulanık sisteminin, sinirsel öğrenme kabiliyetine sahip bir ağ yapısı olarak temsilinden ibarettir [33]. Yapısında hem yapay sinir ağları hem de bulanık mantık kullanılır [34]. Sinir ağı ile bulanık sistemin kombinasyonu bulanık-sinir ağı olarak adlandırılır. ANFIS, öğrenme ve adaptasyon işlemlerini kolaylaştırmak için, adaptif sistemlerden yararlanan bulanık Sugeno modelidir. Böyle bir yaklaşım bulanık mantığı daha sistematik ve tecrübeye daha az bağlı hale getirmektedir [35]. ANFIS'in temel amacı, eşdeğer bulanık mantık sisteminin parametrelerini, giriş-çıkış

veri kümelerini kullanarak bir öğrenme algoritması vasıtasıyla optimize etmektir. Parametre optimizasyonu, gerçek çıkış ile hedef çıkış arasındaki hata değeri minimum olacak şekilde yapılmaktadır [36].

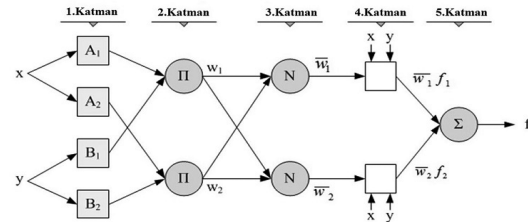
ANFIS, bulanık çıkarım sistemindeki eğer-ise kuralları ve giriş çıkış bilgi çiftlerinden oluşur. Ancak sistem eğitiminde ve denetiminde YSA öğrenme algoritmaları kullanılır [37,38]. x ve y giriş, z ise çıkış olarak alınırsa temel kural yapısı şu şekilde yazılabilir:

$$\text{Eğer } x \in A_i \text{ ve } y \in B_i \text{ ise } z_i = p_i x + q_i y + r_i$$

Burada A_i ve B_i sırasıyla öncül kısımdaki x ve y değişken uzayını bulanık alt uzaylara ayıran kümelerin etiketidir. p_i , q_i ve r_i eğitime işlemi boyunca belirlenen tasarım parametreleridir. z_i ise o kurala ait çıkış değeri olup giriş değişkenlerinin bir fonksiyonudur. Herhangi bir x , y girdi çifti için sonuç çıkış değeri ise tüm kuralların çıkış değeri olan z_i 'lerin ağırlıklı ortalamasıdır [35]. Z çıkış değerinin hesaplanması Eşitlik (1)'de verilmiştir.

$$Z = \sum_{m=1}^k w_m Z_m / \sum_{m=1}^k w_m \quad (1)$$

Aşağıda verilen iki bulanık kurala bağlı olarak, birinci derece bulanık Sugeno modeli için olası ANFIS mimarisi Şekil 1'de verilmiştir. ANFIS modelinin katmanlarını genel olarak özetlersek, 1. katmanda giriş verilerine üyelik fonksiyonları uygulanarak bulanıklaştırma işlemi yapılır. 2. katmanda bulanık mantık çıkarım sistemine göre kurallar oluşturulur. 3. katmanda kural katmanından gelen her bir düğüme, ağırlıklı ortalama ile normalizasyon işlemi uygulanır. 4. katmanda ise bulanık sonuçlar sayısal değerlere dönüştürülür ve son olarak 5. katmanda tüm düğümlerin çıkış değerleri toplanarak sistemin tek çıkış değeri üretilir.



Şekil 1. Adaptif ağ tabanlı bulanık mantık çıkarım sistemi (Adaptive Neuro-Fuzzy Inference System) [39].

Şekilden görüleceği gibi ANFIS, 5 katmanlı ileri beslemeli bir yapay sinir ağı mimarisine sahiptir. Bu

mimari için öğrenme algoritmasının temel görevi, ANFIS çıktısı ile öğrenme verilerini benzetmek için tüm uyarlanabilir parametreleri ayarlamaktır. Eğitim veri seti, sinir ağına tanıtılır ve herhangi bir eğitim algoritması yardımıyla ağ eğitilir. Model çıktısı ile öğrenme verileri arasındaki hata fonksiyonun minimum olduğu şartların belirlenmesi hedeflenir [40].

ANFIS yapısındaki her katmana ait düğüm işlevleri ve katmanların işleyişi aşağıda açıklanmıştır.

1.Katman: Bulanıklaştırma katmanı olarak adlandırılır. Giriş değerlerini bulanık kümeler ayırmada ANFIS modeli kullanılmıştır. Bu katmandaki hücreler, adaptif (uyarlanabilir) hücrelerdir ve hücre sayısı girdi değişkeni sayısına eşittir. Burada, her bir düğümün çıkışı, giriş değerlerine ve kullanılan üyelik fonksiyonuna bağlı olan üyelik derecelerinden oluşmaktadır. Bu tabakadaki hücrelerin çıktıları (O_{1i}) için Eşitlik (2) ve Eşitlik (3) verilmektedir [41].

$$O_{1i} = \mu A_i(x) \quad i = 1,2 \quad (2)$$

$$O_{1i} = \mu B_{i-2}(x) \quad i = 3,4 \quad (3)$$

Burada, A_i ve B_i herhangi bir bulanık küme parametresi, μA_i ve μB_i bu küme parametreleri için üyelik dereceleridir. Çan eğrisi şeklinde üyelik fonksiyonu kullanıldığında, μA_i için aşağıdaki eşitlik verilebilir.

$$\mu A_i = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i} \right)^2 \right]^{b_i}} \quad i = 1,2 \quad (4)$$

Burada, a_i , b_i ve c_i sırasıyla, çan eğrisi şekilli üyelik fonksiyonun sigması, eğimi ve merkezidir.

2.Katman: Kural katmanıdır. Bu katmandaki her bir düğüm, Sugeno bulanık mantık çıkarım sistemine göre oluşturulan kuralları ve sayısını ifade etmektedir. Bu katmandaki hücreler sabittir ve hücre sayısı kural sayısına eşittir. Hücre girdileri, kuralların öncül kısmındaki değişkenlerinin üyelik fonksiyonu değerleridir ve hücre çıktıları (O_{2i}) kuralların ağırlık derecelerini (w_i) vermektedir. Her bir kural düğümünün çıkışı μ_i 1. katmandan gelen üyelik derecelerinin çarpımı olmaktadır.

$$O_{2i} = w_i = \mu A_i(x) \cdot \mu B_i(y) \quad i = 1,2 \quad (5)$$

3.Katman: Normalizasyon katmanıdır. Bu katmandaki her bir düğüm, kural katmanından gelen tüm düğümleri giriş değeri olarak kabul etmekte ve her bir kuralın normalleştirilmiş ateşleme seviyesini hesaplamaktadır. Bu katmandaki hücreler de sabit hücrelerdir, girdileri önceki tabakadan aldıkları ağırlık dereceleridir. Bu tabakada ağırlık dereceleri normalize edilir.

$$O_{3i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1,2 \quad (6)$$

4.Katman: Arındırma katmanıdır. Arındırma katmanındaki her bir düğümde verilen bir kuralın ağırlıklandırılmış sonuç değerleri hesaplanmaktadır. Bu katmandaki hücreler adaptif hücrelerdir ve bu hücrelerin çıktısı Eşitlik (7)'deki şekilde hesaplanır.

$$O_{4i} = \bar{w}_i \cdot f_i = \bar{w}_i \cdot (p_i x + q_i y + r_i) \quad i = 1,2 \quad (7)$$

5.Katman: Toplam katmanıdır. Bu katmanda sadece bir düğüm vardır ve \sum ile etiketlenmiştir. Burada, 4. katmandaki her bir düğümün çıkış değeri toplanarak sonuçta ANFIS sisteminin gerçek değeri elde edilir. Sistemin çıkış değeri ise,

$$O_{5i} = f = \sum \bar{w}_i \cdot f_i = \frac{\sum w_i \cdot f_i}{\sum w_i} \quad i = 1,2 \quad (8)$$

formülüne göre hesaplanmaktadır [42-47].

ANFIS parametrelerinin optimizasyonunda, geriye yayma, en küçük kareler kestirimi, Kalman filtresi yada birden fazla matematiksel optimizasyon yönteminin birleşmesinden oluşan hibrid öğrenme algoritmaları gibi değişik yöntemler kullanılabilir [48, 49].

3. DENEYSEL ÇALIŞMALAR VE ANALİZ (EXPERIMENTAL STUDIES AND ANALYSIS)

3.1. Makine Öğrenmesi Algoritmaları ile Başarım Analizi (Performance Analysis with Machine Learning Algorithms)

Çalışma kapsamında Erciyes Üniversitesi Tıp Fakültesi Genetik Anabilim dalından temin edilen 180 hastaya ait veriler dikkate alınmıştır. Veri setinde uzman hekimlerin öneri ve yönlendirmeleri ile genetik bozukluk olup olmaması hususunda etkili olabileceği öngörülen 47 farklı parametreye ait bilgilere yer verilmiştir. Ancak, veriler üzerine yapılan incelemelerde hastaların tamamının dikkate alınan 47 parametreye ait verilerinin

bulunmadığı görülmüştür. Analizlerin genellenebilir sonuçlar üretebilmesi amacıyla mümkün olduğunca fazla sayıda hasta için mevcut olan ve istatistiksel olarak anlamlılık düzeyi yüksek olan veriler üzerine gerekli çalışmaların yapılması kararına varılmıştır.

Veriler üzerinde yapılan incelemede hangi parametre verilerinin kaç hastada mevcut olduğu tespit edilmiştir. Çalışmamızda uygulayacağımız mühendislik modeline uygun sayıda eğitim ve test veri seti belirleyebilmemiz için yeterli miktarda hasta sayısı olması gerekmektedir. Bu kapsamda, elimizde bulunan verileri dikkate aldığımızda 70 adet hastanın çalışmamız için yeterli olduğu düşünülerek 72 veya daha az sayıda hastada bulunan parametreler veri setinden çıkartılmıştır. Sonuç olarak 180 hastaya ait 39 adet parametrenin dikkate alınmasına karar verilmiştir.

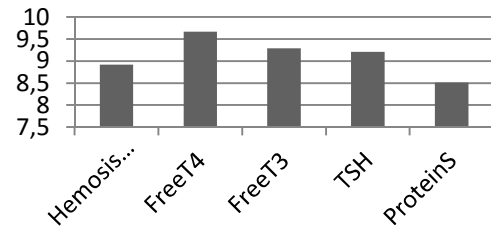
ERÜ Genetik Anabilim dalı öğretim üyeleri ile yapılan çalışmalarda, *Demir* ve *Demir Bağlama* parametrelerinin birbirini temsil edebilir parametreler oldukları öngörülmüştür [50,51]. İstatistiksel yöntemlerden olan *Roc* analizi verilerimize uygulanmış, elde edilen sonuçlar doğrultusunda *Demir Bağlamanın* genetik bozukluk varlığının tespitinde *Demir* parametresinden daha etkili olduğu görüldükçe, *Demir* parametresi veri setinden kaldırılmıştır. *Roc* analizinde her değişken için ayrı ayrı *Roc* eğrileri çizilmiş ve eğri altında kalan alanları hesaplanmıştır. Eğri altında kalan alanlar *Demir* ve *Demir Bağlama* parametreleri için sırasıyla 0.569 ve 0.585 olarak elde edilmiştir. Dolayısıyla genetik bozukluğun ayırt ediciliğinin belirlenmesinde *Demir Bağlamanın* daha yüksek bir kestirim gücü olduğu belirlenmiştir. Aynı şekilde ERÜ Genetik Anabilim dalı öğretim üyelerinden alınan bilgiler doğrultusunda birbirini temsil etme durumunun *Glukoz-Alkalen Fosfataz*, *LDL-APC Resistansı* ve *Kolesterol-Magnezyum* parametreleri arasında da olduğu anlaşılmıştır [52,53]. Yine bu parametreler için de yapılan *Roc* analizi neticesinde *Glukoz*, *APC Resistansı* ve *Magnezyum* parametrelerinin genetik bozukluk varlığının tespitinde daha etkili oldukları anlaşılmıştır. Böylelikle *Alkalen Fosfataz*, *LDL* ve *Kolesterol* parametreleri de veri setinden çıkarılmıştır. Sonuç olarak veri setinde, parametre sayısı 35 olan toplam 180 hastaya ait veriler bırakılmıştır. Tablo 1'de dikkate alınan 35 adet parametre ve bu parametre değerine sahip hasta sayıları verilmiştir.

Yapılan bu çalışmalardan sonra kalan hasta verileri içerisinde hala 35 parametrenin tüm değerleri dolu olmayan hastaların bulunduğu görülmüştür. Oluşturulacak model için kullanılacak örnek veri seti boyutunu daha yüksek tutabilmek amacıyla, hastaların boş olan parametre değerlerinin *Eksik Veri Atama* (*Missing Value Imputation*) yönteminin *Beklenti-*

Maksimizasyonu (*Expectation-Maximization*) algoritması ile rastgele doldurulması yoluna gidilmiştir [54]. Ancak, bu yöntemle üretilen parametre değerlerinin büyük oranda klinik olarak anlamlı olmadığı görülmüş ve üretilen bu değerler çalışmada kullanılmamıştır.

Genellenebilir sonuçlar elde edebilmek için eğitime ve test gurubunda yer alacak hasta sayılarının belirli bir miktarda olması gerektiği bilinmektedir. Bu nedenle toplamda en az 100 hastaya ait bir veri seti ile çalışmanın uygun olacağı kanaatine varılmıştır. Bu kapsamda yapılan incelemeler neticesinde yine ERÜ öğretim üyeleri ile yapılan çalışmalar neticesinde, klinik olarak anlamlı olabilecek 8 adet parametre değeri bulunan (*ACA IGG,ACA IGM, APC Resistansı, FreeT3, FreeT4, Hemosistein, Protein S* ve *TSH*) 103 adet hasta olduğu görülmüştür. Bu hastaların 60'ında genetik bozukluk mevcut olup, 43 tanesinde ise genetik bozukluk bulunmamaktadır.

103 hastaya ait veriler üzerinde istatistiksel çalışmaların *Öznitelik Seçimi* (*Feature Selection*) yöntemi uygulanarak 8 adet parametrenin anlamlılık düzeyleri belirlenmiştir. Bu yöntemde Ki-kare (χ^2) dağılımı testi kullanılmıştır. Bu test ile sınıf değişkeni ve her öznitelik arasındaki bağımsızlık araştırılmıştır. Elde edilen test istatistikleri sıralanarak öznitelik seçimi yapılmıştır [55]. Öznitelik seçimi kriteri olarak $p < 0.10$ düzeyi kullanılmıştır. Yapılan bu çalışma ile sadece *FreeT3*, *FreeT4*, *Hemosistein*, *Protein S* ve *TSH* parametrelerinin model için anlamlı oldukları görülmüştür. Anlamlı olarak bulunan bu 5 parametrenin *Öznitelik Seçimi* yöntemiyle bulunan Ki-kare değerleri Şekil 2'de verilmiştir. Yapılan istatistiksel ön çalışmalar sonucunda, başarımlı analizleri için dikkate alınan veri setinde 103 adet örnek ve 5 adet öznitelik yer almıştır.



Şekil 2. Öznitelik Seçimi yöntemiyle bulunan Ki-kare değerleri (Chi-Square values found by Feature Selection method)

Tablo 1. Parametreler ve Veri Değerine Sahip Hasta sayıları (Parameters and The Number Of Patients with Data Values)

Parametre	Hasta Sayısı
Kreatinin	182
PROTEIN S	182
ALT	179
AST	177

BUN	173
Glukoz	171
APC Resistansı	167
TSH	164
FREET4	157
Hemosistein(HPLC İle)	154
Sodyum	153
Total_Bilirubin	148
Direk_Bilirubin	146
LDH	144
FREET3	143
ACA IGG	139
ACA IGM	138
Kalsiyum	134
Albümin	133
Total Protein	130
Potasyum C	129
Ürik_asit	114
Potasyum	112
GGT	107
Klor	103
Fosfor	99
HBS AG (Doğrulama Dahil)	89
Magnezyum	86
Vitamin B-12	85
Trigliserit	82
HDL Kolesterol	81
Sedimantasyon	80
CPK	78
Demir Bağlama	78
ANTI-HCV1	72

Lojistik Regresyon, Sınıflandırma Ağaçları, Rastgele Orman ve Destek Vektör Makinaları yöntemleri kullanılarak, Trombofili tanısı konmuş 103 hastaya ait verilerden hareketle, belirlenen 5 parametre ile hastalarda genetik bozukluk bulunup bulunmadığının tespit edilmesine yönelik analizler yapılmıştır. Hasta sayısının sınırlı olması nedeniyle, modelin genelleştirebilme yeteneğini doğru bir şekilde değerlendirebilmek için *çapraz doğrulama (cross validation)* yöntemlerinden *K parçalı* yöntemi kullanılmıştır. *K parçalı çapraz geçerlilik* yönteminde amaç bir deneyi bağımsız koşullarda yineleyerek sonuçlarının geçerliliğini sınamaktır. Örneğin, istatistiksel sınıflandırma problemlerinde, bir veri kümesini aşağı yukarı eşit k tane kümeye bölüp, her bir k için, sınıflandırıcıyı oluşturmak için, $k-1$ kümeyi kullanıp ve arta kalan k 'inci küme üzerinde test işlemini yapmaktır. Bu çalışmada, 5-kat çapraz doğrulama yöntemi kullanılmış ve elde edilen sonuçlar Tablo 2'de verilmiştir. Sonuçlardan ilgili

yöntemlerin bu problem üzerindeki performanslarının genel olarak düşük olduğu görülmüştür. Özellikle literatürde en etkili sınıflandırma algoritması olarak bilinen *Destek Vektör Makinaları* yönteminin de anlamlı sonuçlar vermediği görülmüştür. Çalışmada kullanılan sınıflandırma algoritmaları için *Statistica 7 (Stat Soft Inc., Tulsa, USA)* ve *R 3.0.1 (www.r-project.org)* yazılımının *caret* kütüphanesi kullanılmıştır [56]. Destek Vektör Makinaları yönteminde, C-SVM (C-Style Soft Margin Support Vektor Machine) algoritması ve radyal tabanlı çekirdek (kernel) fonksiyonu kullanılmıştır. Cost parametresi 0.5 olarak optimize edilmiştir [57].

Tablo 2' de verilen Duyarlılık kriteri genetik bozukluk var olarak bulunan hasta sayısının gerçekte genetik bozukluk bulunan hasta sayısına oranını, Seçicilik kriteri ise genetik bozukluk yok olarak bulunan hasta sayısının gerçekte genetik bozukluk bulunmayan hasta sayısına oranını vermektedir. Bu iki kriter sınıflandırma modelinde iki önemli başarımlı ölçütüdür. Tablo 2'de verilen sonuçlar incelendiğinde *Destek Vektör Makinaları* için başarımlı oranı yüzdesi diğerlerine göre yüksek olsa da, seçicilik değerinin 0 olması bu yöntem için elde edilen sonuçların da uygulanabilir olmadığını ortaya koymaktadır.

Tablo 2. Farklı yöntemlere ait sınıflandırma sonuçları (Classification results of different methods)

Yöntem	Başarımlı (%)	Duyarlılık	Seçicilik
Lojistik Regresyon	56.3	0.6500	0.4419
Sınıflandırma Ağaçları	56.3	0.6500	0.4419
Rastgele Orman	56.3	0.7833	0.2558
Destek Vektör Makinaları	57.2	0.9833	0.0000

3.2. ANFIS ile Başarımlı Analizi (Performance Analysis with ANFIS)

Sınıflandırma algoritmaları ile elde edilen sonuçları karşılaştırmak üzere Yapay zeka tekniklerinden ANFIS (Adaptive Neuro Fuzzy Inference System) kullanılarak ilave çalışmalar yapılmıştır. Bu çalışmada ANFIS için nonlinear üyelik fonksiyonları ve verilerin hazırlanmasında çapraz geçerlilik (cross validation) yöntemi kullanılmıştır. Toplam 103 hastaya ait veri aşağı yukarı olmak üzere 5 parçaya bölünmüştür. Her bir parça kadar tekrar edilerek 4 parça veri eğitim için ve kalan 1 parçada test için kullanılarak elde edilen sonuçlar kaydedilmiştir. Sınıflandırma modelini oluşturmak ve oluşturulan modelin test işlemin yapmak için Matlab ortamında gerekli kodlama çalışmaları yapılmıştır. Kodlama içerisinde ANFIS modelinin tüm giriş değerlerine 0 ve 1 aralığında veri normalizasyonu

uygulanmıştır. ANFIS modelinin en uygun üyelik fonksiyonu, üyelik fonksiyonu sayısı, iterasyon sayısı ve modelin çıkış fonksiyonunu belirleme çalışmaları yapılmıştır. Modelin bir çok kez farklı parametreler ile koşturulmasından sonra ANFIS modelinin üyelik fonksiyonu *trimf*, üyelik fonksiyonu sayısı sırasıyla 5 giriş için [2 3 3 3 3], iterasyon sayısı 100 ve çıkış fonksiyonu *constant* olarak seçilmiştir. Oluşturulan ANFIS modelinde 103 hastadaki 5 giriş parametresi ve bir çıkış parametresi için başarımlar Tablo 3’de verilmiştir.

Tablo 3. ANFIS modeli ile başarımlar sonuçları (Benchmark results with Anfis model)

Yöntem	Başarımlar (%)	Duyarlılık	Seçicilik
ANFIS	63.1	0.7333	0.4883

4. SONUÇLAR VE TARTIŞMA (CONCLUSIONS AND DISCUSSION)

Bu çalışmada, Trombofili hastalığı tanısı konmuş hastaların klinik verileri ile genetik bozukluk olup olmama durumu arasındaki ilişkinin belirlenmesi amacıyla Adaptif Ağ Tabanlı Bulanık Mantık Çıkarım Sistemi (ANFIS)’nin kullanılması araştırılmıştır. Bu amaçla, Erciyes Üniversitesi Tıp Fakültesi Hastanesi Genetik Anabilim Dalından temin edilen 180 hastaya ait veriler kullanılmıştır. Hastalığın tespiti için önem arz eden parametrelere ait verilerin tüm hastalar için mevcut olmaması nedeniyle, bu hastalardan 103’ünün verileri bu çalışma için kullanılabilmiştir. Hasta verileri üzerinde, *Lojistik Regresyon*, *Sınıflandırma Ağaçları*, *Rastgele Orman* ve *Destek Vektör Makinaları* yöntemleri kullanılarak yapılan analizlerde, başarımların en fazla %56.3 olduğu ve ayrıca seçicilik değerlerinin düşük olduğu görülmüştür.

ANFIS kullanılarak elde edilen sonuçlar incelendiğinde, modelinin başarımlar oranının, sınıflandırma algoritmalarına göre daha yüksek olduğu görülmektedir. ANFIS Modeli ile başarımlar oranının % 63.1 olarak gerçekleştiği, duyarlılık ve seçicilik değerlerinin de diğer yöntemlere göre daha başarılı olduğu görülmüştür.

Sınıflandırma algoritmalarının doğrusal olmayan karmaşık problemler üzerindeki başarımları genel olarak sınırlı olabilmektedir. ANFIS, bulanık sistemlerin ve sinirsel ağların bilinen en iyi özelliklerini kendi içyapısında birleştirmektedir. Bulanık sistemlerin, giriş bilgisini ağ üzerine dağıtarak araştırma uzayını azaltması, Sinirsel ağların ise geri yayılım davranışlarını kullanarak kontrol parametrelerinin doğrusal olmayan problemler için en uygun değerlerinin bulunması ANFIS modeline önemli üstünlükler kazandırmaktadır.

ANFIS’in eğitilmesi, sınıflandırma algoritmalarına göre daha fazla işlem sayısı ve etkin hafıza kullanımı gerektirmektedir. Ancak, ANFIS bir kez eğitildiğinde, girişlerine uygulanan veriler için çok hızlı bir şekilde sonuç verebilmektedir. Dolayısıyla, yüksek başarımlar düzeyi ve uygulamada girişlere karşılık çok kısa sürelerde cevap verebilmesi ANFIS yönteminin önemli bir üstünlüğü olarak ortaya çıkmaktadır.

Genel olarak, belirlenmesi gereken parametre sayısının fazla olması ve ağı eğitilmesinde kullanılan geri yayılım algoritmasının türe dayalı olması nedeniyle bölgesel optimuma takılma riski içermesi gibi hususlar ANFIS’in başarımlarını olumsuz etkileyebilen temel unsurlar arasındadır. Bu kapsamda, ANFIS ağına geri yayılım algoritması yerine yapay zeka optimizasyon algoritmaları kullanılarak eğitilmesi ve ilgili problem üzerindeki başarımlarının incelenmesi ise gelecek çalışmalar olarak planlanmaktadır.

KAYNAKLAR (REFERENCES)

- [1] R. Küçükaya, ve M. Aydın, “Trombofili Genetiği”, Türk Hematoloji Derneği - Moleküler Hematoloji, Klinik Hematoloji için Pratik Genetik Yaklaşım Kursu, Antalya, 39-43, Kasım 2006.
- [2] V. De Stefano, G. Finazzi ve M. Mannucci, “Inherited thrombophilia: Pathogenesis, clinical syndromes and management”, Blood, vol. 87, no. 9, pp. 3531-3544, 1996.
- [3] D.A. Lane, P.M. Manucci ve K.A. Bauer, “Inherited thrombophilia: Part-1”, Thromb Haemost, vol. 76, 1996.
- [4] B. Dahlback, M. Carlsson ve P.J. Svensson, “Familial thrombophilia due to a previously unrecognized mechanism characterized by poor anticoagulant response to activated protein C: Prediction of a cofactor to activated protein C.”, Proc Natl Acad Sci USA, vol. 90, pp. 1004-1008, 1993.
- [5] R.M. Bertina et al., “Mutation in blood coagulation factor V associated with resistance to activated protein C.”, Nature, vol. 369, no. 64, 1994.
- [6] C.R. Falcon et al., “High prevalence of hyperhomocyst(e)inemia in patients with juvenile venous thrombosis”, Arterioscler Thromb, vol. 14, 1080, 1994.
- [7] S.R. Poort, F.R. Rosendaal, P.H. Reitsma, R.M. Bertina, “A common genetic variation in the 3’-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis”, Blood, vol. 88, 3698, 1996.

- [8] M. Makris, F.R. Rosendaal ve F.E. Preston, "Familial thrombophilia: Genetic risk factors and management", *J Int Med*, vol. 242, no. 9, 1997.
- [9] M. Kolodziej ve P.C. Comp, "Hypercoagulable states due to natural anticoagulant deficiencies", *Curr Op Hematol*, vol. 301, 1993.
- [10] G.M. Rodgers, "Thrombosis and antithrombotic therapy. In: Wintrobe's Clinical Hematology, 10th Edition", Williams and Wilkins Comp, pp. 1781-1818, 1998.
- [11] K.A. Bauer, "The Hypercoagulable state. In: Williams Hematology 5th edition", McGraw Hill inc, pp. 1531-1549, 1995.
- [12] D.A. Lane, P.M. Manucci ve K.A. Bauer, "Inherited thrombophilia: Part-2.", *Thromb Haemost*, vol. 76, 824, 1996.
- [13] R.L. Bick ve H. Kaplan, "Syndromes of thrombosis and hypercoagulability", *Med Clin North Am.*, vol. 82, 409, 1998.
- [14] M.B. Hultin, "Antithrombin III assays. In: Williams Hematology 5th edition", McGraw Hill inc., 101, 1995.
- [15] P.C. Comp, "Protein C and protein S. In: Williams Hematology 5th edition", McGraw Hill inc., 99, 1995.
- [16] G.N. Welch ve J. Loscalzo, "Homocysteine and atherothrombosis", *N Engl J Med*, vol. 338, 1042, 1998.
- [17] A. D'Angelo ve J. Selhub, "Homocysteine and thrombotic disease", *Blood*, vol. 90, no. 1, 1997.
- [18] T. Koster et al., "Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep venous thrombosis", *Lancet*, vol. 345, 152, 1995.
- [19] P.W. Kamphuisen, J.J. Houwing-Duistermaat ve J.C. van Hauwelingen, "Familial clustering of factor VIII and von Willebrand factor levels", *Thromb Haemost*, vol. 79, 561, 1998.
- [20] F.R. Rosendaal, "Venous thrombosis: a multicausal disease", *Lancet*, vol. 353, 1167, 1999.
- [21] O. Wu et al., "Screening for thrombophilia in high-risk situations: a meta analysis and cost effectiveness analysis", *Br J Haematol*, vol. 131, pp. 80-90, 2006.
- [22] P.H. Reitsma ve F.R. Rosendaal, "Past and future of genetic research in thrombosis", *J Thromb Haemost*, pp. 264-269, 2007.
- [23] Q. Yang et al., "Improving the Prediction of Complex Diseases by Testing for Multiple Disease-Susceptibility Genes", *Am. J. Hum. Genet.*, vol. 72, pp. 636-649, 2003.
- [24] K. Santhi et al., "Detection of Venous Thromboembolism by Proteomic Serum Biomarkers", *PLoS ONE*, vol. 6, 2007.
- [25] T. Baglin, "Using the laboratory to predict recurrent venous thrombosis", *International Journal of Laboratory Hematology*, vol. 33, pp. 333-342, 2011.
- [26] H.G. de Haan et al., "Multiple SNP testing improves risk prediction of first venous thrombosis", *Blood*, vol. 120, no. 3, pp. 656-663, 2012.
- [27] J.C. Souto ve J.M. Soria, "Predicting individual risk of venous thrombosis", *Blood*, vol. 120, no. 3, pp. 500-501, 2012.
- [28] S. Penco et al., "Assessment of the role of genetic polymorphism in venous thrombosis through artificial neural networks", *Ann Hum Genet*, vol. 69, pp. 693-706, 2005.
- [29] D. Shanthi, G. Sahoo ve N. Saravanan, "Designing an Artificial Neural Network Model for the Prediction of Thrombo-embolic Stroke", *International Journals of Biometric and Bioinformatics (IJBB)*, vol. 3, no. 1, 2009.
- [30] A. Dahabiah et al., "Comparative Neural Network Based Venous Thrombosis Echogenicity and Echostructure Characterization Using Ultrasound Images", *Information and Communication Technologies, ICTTA '06*, vol. 1, pp. 992 - 997, 2006.
- [31] H. Markus, M. Cullinane ve G. Reid, "Improved Automated Detection of Embolic Signals Using a Novel Frequency Filtering Approach, Stroke", *Journal of the American Heart Association*, pp. 1610-1615, 1999.
- [32] V. Keme'ny et al., "Automatic Embolus Detection by a Neural Network", *Stroke, Journal of the American Heart Association*, pp. 807-810, 1998.
- [33] H. Özçalık ve A. Uygur, "Dinamik Sistemlerin Uyumlu Sinirsel - Bulanık Ağ Yapısına Dayalı Etkin Modellenmesi", *KSÜ Fen ve Mühendislik Dergisi, Kahramanmaraş*, vol. 6, no. 1, pp. 36-46, 2003.
- [34] E. Avcı ve Z.H. Akpolat, "Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi İle Doğru Akım Motorlarının Hız Denetimi", *ELECO'2002 Elektrik-Elektronik-Bilgisayar Mühendisliği Sempozyumu, Bursa*, pp. 193-196, Aralık 2002.
- [35] E. Özgan et al., "Asfalt Betonunda Marshall Stabilesinin Uyarlamalı Sinirsel Bulanık Mantık Yaklaşımı İle Tahmini", *Uluslar arası İleri Teknolojiler Sempozyumu, Karabük*, 2009.
- [36] K. Güney ve N. Sarıkaya, "Dairesel Mikroserit Antenlerin Yama Yarıçapının Çeşitli Algoritmalarla Optimize Edilen Bulanık Mantık Sistemine Dayalı Uyarlanır Ağlar İle Hesaplanması", *ELECO'2008 Elektrik-Elektronik-Bilgisayar Mühendisliği Sempozyumu ve Fuarı, Bursa*, 26-30 Kasım 2008.
- [37] J.S.R. Jang, "ANFIS: Adaptive network-based fuzzy inference systems", *IEEE Trans. Syst., Man. and Cybern.*, vol. 23, pp. 665-685, 1993.

- [38] G.F. Franklin, J.D. Powell ve M.L. Workman, "Digital Control of Dynamic Systems", Addison Wesley, pp. 130-132, 1990.
- [39] S. Uzundurukan, "Zeminlerin Şişme Özelliklerine Etkiyen Temel Parametrelerin Belirlenmesi ve Modellenmesi", Doktora Tezi, İnşaat Mühendisliği Anabilim Dalı, Isparta, 2006.
- [40] Z. Himer et al., "Neuro- fuzzy modeling and genetic algorithms optimization for flue gas oxygen control", 2nd IFAC Workshop on Advanced Fuzzy/Neural Control, University of Oulu, 2004.
- [41] J.S.R. Jang ve C.T. Sun, "Neuro-Fuzzy Modeling and Control", proceedings of the IEEE, vol. 83, no. 3, 1995.
- [42] Ö. Başkan, "İzole Sinyalize Kavşaklardaki Ortalama Taşıt Gecikmelerinin Yapay Bulanık Sinir Ağları ile Modellenmesi", Yüksek lisans tez çalışması, Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü, Denizli, 2004.
- [43] F. Hocaoglu ve M. Kurban, "Adaptif Ağ Tabanlı Bulanık Mantık Çıkarım Sistemi İle Eskişehir Bölgesi İçin Güneşlenme Süreleri Tahmini", Elektrik-Elektronik-Bilgisayar Mühendisliği 11. Ulusal Kongresi ve Fuarı, İstanbul, 22-25 Eylül 2005.
- [44] M. Fırat, M. Yurdusev ve M. Mermer, "Uyarlamalı Sinirsel Bulanık Yaklaşımı İle Aylık Su Tüketiminin Tahmini", Gazi Üniversitesi, Mühendislik-Mimarlık Fakültesi Dergisi, vol. 23, no. 2, pp. 449-457, 2008.
- [45] E. Günay, M. Alçı ve S. Parmaksızoğlu, "Çoklu Kaotik Çeker Üreten DK-HSA Devresinde Parçalı – Doğrusal Fonksiyon Karakteristiğinin Bulanık Sinir Ağı Kontrolü İle Elde Edilmesi", URSI-Türkiye 2006 Bilimsel Kongresi ve Ulusal Genel Kurul Toplantısı, Hacettepe Üniversitesi, Ankara, 6-8 Eylül 2006.
- [46] G. Civelekoğlu, "Arıtma Proseslerinin Yapay Zeka ve Çoklu İstatiksel Yöntemler İle Modellenmesi", Doktora Tezi, Çevre Mühendisliği Anabilim Dalı, Isparta, 2006.
- [47] E. Özgan, "NAPO₃ Miktarının Zemin Tanelerine Etkisi ve Tane Büyüklüklerinin ANFIS Yöntemiyle Tahmini", Uluslar arası İleri Teknolojiler Sempozyumu, Karabük, 2009.
- [48] O. Akyılmaz, T. Ayan ve M.T. Özlüdemir, "Geoid surface approximation by using Adaptive Network based Fuzzy Inference Systems", AVN, no. 8, 2003.
- [49] B. Haznedar, "Yapay Zeka Teknikleri Kullanılarak Erkek Kısır Hastalarda Genetik Bozukluk Varlığının Tespiti", Yüksek Lisans Tezi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, 2010.
- [50] Y. Kurtulmuş et al., "Demir Düzeyi ve İskemi Modifiye Albumin", Smyrna Tıp Dergisi, pp. 30-35, 2013.
- [51] Z. Uysal, V. Çulha ve Ş. Cin, "Demir Eksikliği Tanısında Serum Transferrin Reseptörü", Ankara Üniversitesi Tıp Fakültesi Dergisi, vol. 51, no. 3, 1998.
- [52] I. Onyesom et al., "Relationship between placental alkaline phosphatase activity and cord blood glucose, albumin and neonatal birth weight at term.", Invest Clin, vol. 50, no. 4, 2009.
- [53] G.M. Rao ve L.O. Morghom, "Correlation between serum alkaline phosphatase activity and blood glucose levels.", Enzyme, vol. 35, no. 1, 1986.
- [54] B.M. Webb-Robertson et al., "Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics", Journal of Proteome research, pp. 1993-2001, 2015.
- [55] B.Y. Cheng, J.G. Carbonell ve J. Klein-Seetharaman, "Protein Classification Based on Text Document Classification Techniques", PROTEINS: Structure, Function, and Bioinformatics, pp. 955-970, 2005.
- [56] M. Kuhn, "Building Predictive Models in R Using the caret Package", Journal of Statistical Software, vol. 28, no. 5, 2008.
- [57] X. Wu ve R. Srihari, "New v-Support Vektor Machines and their Sequential Minimal Optimization", Proceedings of the 20. International Conference on Machine Learning (ICML-2003), Washington DC, 2003.