Gazi Üniv. Müh. Mim. Fak. Der.
Cilt 29, No 3, 477-486, 2014

Journal of the Faculty of Engineering and Architecture of Gazi University
Vol 29, No 3, 477-486, 2014

SINIFLANDIRMA KURALLARININ ÇIKARIMI İÇİN ETKİN VE HASSAS YENİ BİR YAKLAŞIM

Murat KÖKLÜ[1], Humar KAHRAMANLI[2], Novruz ALLAHVERDİ[3]
Bilgisayar Mühendisliği Bölümü, Teknoloji Fakültesi, Selçuk Üniversitesi, Konya
mkoklu@selcuk.edu.tr, hkahramanli@selcuk.edu.tr, noval@selcuk.edu.tr

(Geliş/Received: 04.03.2013; Kabul/Accepted: 17.07.2014)

ÖZET

Bu çalışmada çok sınıflı verilerden kural çıkarımı için yeni bir yöntem önerilmiştir. Önerilen metot 4 farklı veri kümesi üzerinde uygulanmıştır. Ayrık ve gerçel nitelikler farklı şekilde kodlanmıştır. Ayrık nitelikler ikili olarak, gerçek nitelikler ise, iki gerçel değer kullanılarak kodlanmıştır. Bu değerler kuralları oluşturan niteliklerin değer aralıklarının orta noktası ve genişlemesidir. Kural çıkarım işlemi için sınıflandırma başarısı uygunluk fonksiyonu olarak kullanılmıştır. Uygunluk fonksiyonunun optimizasyonu amacıyla Yapay Bağışıklık Sistemi (YBS) algoritması olan CLONALG kullanılmıştır. Önerilen yöntemi uygulamak için Süsen çiçeği (Iris), Şarap (Wine), Cam Kimliklendirme (GlassIdentification) ve Deniz Kabuğu (Abalone) veri kümeleri kullanılmıştır. Veriler Irvine California Üniversitesi (UCI) makine öğrenmesi veri tabanından elde edilmiştir. Önerilen metot kullanılarak Süsen çiçeği için %100, Şarap için %99,44, Cam kimliklendirme için %77,10 ve Deniz Kabuğu için %62,59 doğruluk elde edilmiştir. Diğer yöntemlerle kıyaslandığında önerilen yöntem kullanılarak daha başarılı sonuçlar elde edildiği ve hesaplama karmaşıklığı açısından da avantajlı olduğu görülmüştür.

Anahtar Kelimeler: Kural çıkarımı, Gerçel değer kodlaması, Çok sınıflı problemler.

A NEW ACCURATE AND EFFICIENT APPROACH TO EXTRACT CLASSIFICATION RULES

ABSTRACT

A new method for extracting rules from multi-class datasets was proposed in this study. The proposed method was applied to 4 different data set. Discrete and real attributes were decoded in different ways. Discrete attributes were coded as binary whereas real attributes were coded by using two real values These values indicate the midpoint and the expansion of intervals of the attributes that form the rules. Classification success was used as fitness function of rule extraction. CLONALG which is Artificial Immune Systems (AIS) algorithm was used to optimize the fitness function. To apply the proposed method Iris, Wine, Glass and Abalone datasets were used. The datasets were obtained from machine learning repository of University of California at Irvine (UCI). The proposed method achieved prediction accuracy ratios of 100%, 99,44%, 77,10%, and 62,59% for Iris, Wine, Glass and Abalone datasets, respectively. When it is compared with the previous studies it has been seen that the proposed method achieved more successful results and has advantage in terms of complexity.

Keywords: Rules extraction, Real value coding, Multi-class problems.

1. GİRİŞ (INTRODUCTION)

Modern bilgisayarlar yüksek miktarda veri depolamaktadır. Son yıllarda veri madenciliğine olan ilgi giderek artmaktadır. Veri madenciliği yöntemleri veriyi bilgiye dönüştürmek ve bu bilgiden yararlanarak yeni bilgilere ulaşmak için kullanılır. Veri sınıflandırma, veri madenciliğinin önemli bir alanıdır. Bugüne kadar araştırmacılar sınıflandırma için değişik teori ve yöntemler üzerine farklı çalışmalar yapmış ve bir çok önemli sonuca ulaşmışlardır. Sınıflandırma yöntemleri genellikle şeffaf olmayan yöntemler olduğundan birçok yeni araştırmacı bu metotları kural çıkarımı olarak isimlendirilen şeffaf modellere çevirmeye çalışmışlardır. Önerilen yöntemler genellikle iki

sınıflı problemler için kullanılmaktadır. Bu çalışmada verilerden kural çıkaran bir yöntem açıklanmıştır. Önerilen yöntem kullanılarak iki sınıflı problemlerin yanında, çok sınıflı problemlerden de sınıflandırma kuralları çıkarılabilir. Optimizasyon için Yapay bağışıklık sistemleri (YBS) kullanılmıştır. Uygunluk fonksiyonu olarak sınıflandırma başarısı kullanılmış ve optimize edilmiştir. Literatürde genellikle ikili kodlama kullanılmış, gerçel değer alan nitelikler belirli mantıkla kesikli hale getirilmiştir [1].

Bu çalışmada ise öncekilerden farklı olarak aralıklar optimize edilerek oluşturulmuştur. Yalnız birçok veri kümesi ayrık değerler alan niteliklere sahip olmaktadır. Örneğin Deniz kabuğu veri kümesinin cinsiyet özelliği erkek, kadın ve çocuk olmak üzere 3 farklı değer alabilmektedir. Bu tip özelliklerin ikili olarak kodlanmasının daha uygun olacağı düşünülmüştür. Böylece bu çalışmada kullanılan yöntemin ana fikri gerçek ve ayrık değerlerin farklı kodlanması olmuştur.Bu amaçla ayrık değerler ikili sistemde kodlanmıştır. Gerçek değerler ise iki sayı kullanılarak kodlanmıştır. Bu sayılar kuralları oluşturan aralıkların genişlemesi ve orta noktasını ifade etmektedir. Yöntem dört veri kümesine uygulanmıştır. Elde edilen sonuçlar diğer çalışmalarla karşılaştırılmış ve sonuçların daha başarılı olduğu gözlenmiştir.

Yapılan literatür araştırmasında az sayıda veriden doğrudan kural çıkarma çalışmalarının yapıldığı görülmektedir. Kural çıkarımı için genellikle ilk aşama olarak Yapay Sinir Ağları (YSA) ve Destek Vektör Makinaları (DVM) kullanılmıştır. Bu konuda ilk çalışma Gallant tarafından yapılmıştır [2]. Towell ve Shavlik kural oluşturmak için seçilen bir nöronu aktive etmek amacıyla ağırlıkları incelemeye dayanan SUBSET isimli algoritma geliştirmişlerdir [3]. Lu ve arkadaşları gizli katman nöron aktivasyon değerlerini kümeleyerek YSA tabanlı bir kural çıkarım metodu geliştirmişlerdir [4]. Keedwell ve arkadaşları YSA giriş uzayında kural aramak için genetik algoritmaları kullanan bir sistem geliştirmişlerdir [5, 6]. Aliev ve arkadaşları etkin ve hızlı batarya şarjının kontrol kurallarının çıkarımı için esnek hesaplama tabanlı bir veri madenciliği tekniği kullanmışlardır. Elde edilen kurallar NiCd batarya şarjı için kullanılmıştır [7]. Ang ve arkadaşları kural çıkarımı için evrimsel algoritmaların (EA)global araştırma kapasitesini tamamlamak için yerel arama yoğunluk şeması kullanan evrimsel memetik algoritma (EMA) önermişler. Onlar yerel arama için iki şema kullanmışlar: EMA-μGA olarak isimlendirilen yöntem Mikro genetik algoritma (GA) tekniği kullanırken, yapay bağışıklık siteminden esinlenerek hücre üretimi için klonal seçme kullanan diğer yöntem EMA-AIS olarak adlandırılmıştır [8]. Papageorgiou birçok farklı bilgi çıkarım yöntemi kullanarak verilerden mevcut bilgiyi bulanık kurallar olarak çıkarmış ve bunu bulanık bilişsel haritalarında (FCM) kullanmıştır [9]. Rodriguez ve arkadaşları sınıflandırma kural çıkarımı için ada modeli tabanlı bir dağıtık iyileştirilmiş genetik algoritma sunmuştur [10]. Sarkar ve arkadaşları karar verme ağacı ve genetik algoritma (DTGA) olarak isimlendirdikleri bölge, büyüklük, boyutluluk ve sınıf dağılımından bağımsız herhangi bir sınıflandırma problemleri üzerinde kestirim doğruluğunu iyileştirmeyi amaçlayan bir doğruluk tabanlı eğitim sistemi kurmuşlardır [11]. DTGA iki aşamadan oluşmaktadır. İlk aşamada bir temel bir sınıflandırıcı yöntem olan C4.5 (bir karar ağacı tabanlı kural çıkarıcı) eğitim veri setinden kural üretimi için kullanılırken, bir sonraki aşamada GA kullanılarak kestirim için daha doğru ve yüksek performanslı kural çıkarımı sağlamayı amacıyla elde edilmiş kuralları sadeleştirilmiştir. Özbakır ve ark. Kural çıkarma amacıyla DIFACONN-miner olarak adlandırdıkları yeni bir yöntem geliştirmişlerdir. Bu yöntemde önce YSA diferansiyel gelişim algoritması kullanarak eğitilmiş, bir sonraki adımda YSA'dan karınca kolonisi algoritması kullanılarak kurallar çıkarılmıştır [12]. Baykasoğlu ve ark. gezgin satıcı probleminin modellenmesi ve çözümü için YBS algoritmalarından klonal seçim algoritması kullanılmışlar [13].

Çalışma aşağıdaki gibi organize edilmiştir. İkinci bölümde yapay bağışıklık sistemleri açıklanmıştır. Üçüncü bölümde sunulan yöntem açıklanmıştır. Deney sonuçları ve bu sonuçların değerlendirilmesi dördüncü bölümde sunulmuştur. En son bölümde çalışma sonuçları yorumlanmıştır.

2.YAPAY BAĞIŞIKLIK SİSTEMLERİ
(ARTIFICIAL IMMUNE SYSTEMS)

Bağışıklık sistemi yaşayan organizmalarda yabancı oluşumları fark etme ve üstesinden gelme ile sorumlu bir doğal mekanizmadır [14]. Bağışıklık sisteminin ana amacı vücuttaki bütün hücreleri (veya molekülleri) tanıması ve bu hücreleri kendinden yada kendinden olmayan (yabancı) şekilde ayırması, patojen olarak isimlendirilen hastalık yapıcı hücrelere karşı koruması ve doğru çalışmayan hücrelerin eliminasyonudur [15]. Bağışıklık sistemi tarafından tanınan bütün elementler antijenolarak isimlendirilir [16]. İki tip antijen vardır; kendinden olan ve yabancı. Yabancı antijenler hastalık yapıcı elementlerken, kendinden olan antijenler vücuda zararsızdır [17]. B-hücreleri ve T-hücreleri olarak adlandırılan iki ana bağışıklık hücresi grubu vardır. Bunlar sınırsız denebilecek kadar çok antijenikpatern (antijenik örüntü) tanınmasına yardımcı olurlar. Hastalıklara karşı aşılanmış insanlarda belirli hastalık yapıcı etmenlere bir şekilde bağlanan bazı özel etmenlerin var olduğu bulunmuştur. Bu etmenler antikor olarak isimlendirilirler [18]. YBS sınıflandırma, anomali bulma, bilgisayar güvenliği, nümerik fonksiyon optimizasyonu, optimizasyon, öğrenme, biyo-enformatik, görüntü işleme, robotik, kontrol, virüs saptama ve web madenciliği gibi birçok farklı alanda

problem çözmekte kullanılır [19]. YBS üç ana bağışıklık teorisine odaklanmıştır: klonal seçim, bağışıklık ağları ve negatif seçim [20].

2.1. CLONALG algoritması(CLONALG algorithm)

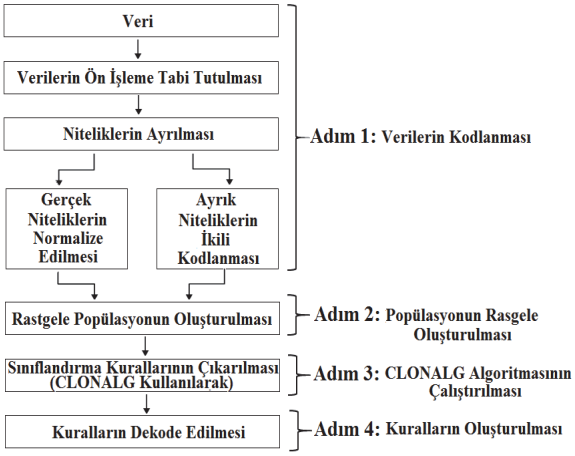
CLONALG algoritmasının genel hatlarıyla adım adım gösterimi [21]:

1. **Başla** – Algoritmanın ilk adımıdır. Bu adımda sabit bir N boyutlu antikor havuzu oluşturulur. Bu havuz daha sonra iki parçaya bölünür. Bunlar hafıza antikor havuzu *m* ve kalan antikor havuzu *r*. Hafıza antikorları en sonunda algoritmanın çözümünü gösterir. Kalan antikor havuzu ise sisteme çeşitlilik eklemek için kullanılır.

2. **Döngü** – Algoritma daha sonra bilinen tüm antijenler için sistemi belirli bir sayıda iterasyona maruz bırakarak çalışmasını devam ettirir. Tek bir tur çalıştırma veya bir iterasyon işlemi, bir jenerasyon olarak ifade edilir. Sistemde probleme özel bir sonlanma koşulu kullanılabilmesine rağmen, sistem jenerasyon sayısı*G,* kullanıcı tanımlı olarak çalıştırılabilir.
   a. **Antijen Seçimi** –Antijen havuzundan rastgele bir antijen seçilir.
   b. **Maruz Bırakma** – Sistem seçilen antijen için çalıştırılır. Duyarlılık değerleri, antijene karşı tüm antikorlar için hesaplanır. Duyarlılık bir benzerlik ölçüsüdür ve probleme bağlıdır. Yaygın olarak Hamming mesafesi kullanılır.
   c. **Seçim** – *n* tanelik antikor kümesi, antijen ile en yüksek duyarlılığa sahip tüm antikor havuzundan seçilir.
   d. **Klonlama** – Seçilen antikorların kümesi duyarlılıklarıyla orantılı olarak klonlanır.
   e. **Duyarlılık Olgunlaşması (Mutasyon)** – Klonlar (kopya antijenlerin kümesi) söz konusu *m* antijeniyle daha iyi bir eşleşme sağlamak için bir duyarlılık olgunlaştırma sürecine tabi tutulur. Burada olgunlaşma derecesi ebeveynlerinin duyarlılığıyla ters orantılıdır. Yani daha yüksek duyarlılık daha az mutasyon demektir.
   f. **Klon Çalıştırma** – Daha sonra klonların duyarlılıkları hesaplanır.
   g. **Adaylık** – Klonlar içinde en yüksek uygunluklu antikor veya antikorlar aday hafıza antikorları olarak seçilir. Eğer bir aday hafıza hücresinin duyarlılığı, hafıza hücresindeki en yüksek uyarılmış antijenin duyarlılığından yüks"ek ise bunlar yer değiştirilir. Grup yer değiştirmeleri de benzer şekilde meydana gelir.
   h. **Yer Değiştirme** – Son olarak en düşük duyarlılığa sahip *d* sayıda birey yeni rastgele antikorlar ile değiştirilir.
3. **Bitir** –Antijen havuzunun elemanlarından biri olan *m* sayıda hafıza elemanı çözüm olarak alınır. Probleme bağlı olarak, tek bir çözüm olabileceği gibi, havuzdaki tüm antijenler de çözüm olabilir.

3. ÖNERİLEN YÖNTEM(PROPOSED METHOD)

Verilerden kural çıkarmak için önerilen yaklaşımda sınıflandırma doğruluk fonksiyonu CLONALG kullanılarak optimize edilmektedir. Optimizasyon algoritmasının seçiminde CLONALG, genetik algoritma, parçacık sürü algoritması, guguk kuşu algoritmaları denenmiş olup, sonuçlarda kayda değer bir fark gözlenmediği için daha önceki çalışmalarımızla bütünlük açısından CLONALG tercih edilmiştir. Önerilen algoritmanın temel adımları ve algoritmanın akış diyagramı Şekil 1'de gösterilmektedir.



Şekil 1. Önerilen algoritmanın temel adımları ve algoritmanın akış diyagramı (The basics teps and flowchart of the proposed method algorithm)

3.1 Veri Kodlanması (Data Coding)

Veri kümesindeki nitelikler ayrık ve gerçek olarak iki gruba ayrılmıştır ve farklı yöntemlerle kodlanmıştır. Ayrık nitelikler ikili olarak kodlanmıştır. Bunun için aşağıdaki yöntem kullanılmıştır. Veri *N* sayıda kesin niteliğe sahip olsun. Her A_n {*n=1,2,...,N*} niteliği m_n sayıda $\{a_1, a_2, ..., a_{m_n}\}$ biçimli alt katara (string) bölünür ve $\{b_{n1}, b_{n2}, ..., b_{nm_n}\}$ olarak kodlanır. Eğer A_n niteliği a_i *(i=1,2,....m_n)* alt katarına aitse, b_{nj} denklem (1)'deki,kesin niteliklerin kodlanmış hali denklem (2)'deki,*X* vektörünün uzunluğu denklem (3)'deki gibidir [1].

$$b_{nj} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} j=1,2,...,m_n \qquad (1)$$

$$X = \bigcup_{n=1}^{N} \bigcup_{i=1}^{m_n} b_{ni} \quad (2)$$

$$m = \sum_{n=1}^{N} m_n \quad (3)$$

Gerçek niteliklerin kodlanması:
Veri *K* adet gerçek niteliğe sahip olsun. Her bir L_k {*k = 1,2, ...,K}*niteliği iki sayı ile temsil edilir. İlk *K*adet $\{l_{11}, l_{21}, \dots, k_{K1}\}$ sayıları aralıkların orta noktalarını, ikinci *K*adet $\{l_{12}, l_{22}, \dots, k_{K2}\}$ sayıları genişlemelerini denklem (4)belirtmektedir.

$$l_i \in (l_{i1} - l_{i2}, l_{i1} + l_{i2}) i=1,2,\dots,K \quad (4)$$

Gerçek özelliklerin kodlanması denklem (5) ile hesaplanır. *XR* vektörünün uzunluğu *t=2*K'*dır.

$$XR = \bigcup_{n=1}^{K} \bigcup_{i=1}^{2} l_{ki} \quad (5)$$

Sonuç olarak veri kodlandıktan sonra oluşan vektör denklem (6)'deki gibi elde edilir.

$$V = X \cup XR = \left(\bigcup_{n=1}^{N} \bigcup_{i=1}^{m_n} b_{ni} \right) \cup \left(\bigcup_{n=1}^{K} \bigcup_{i=1}^{2} l_{ki} \right) \quad (6)$$

Vektörlerin toplam uzunluğu ise denklem (7) ile hesaplanır.

$$uz = m + t = \sum_{n=1}^{N} m_n + t \quad (7)$$

3.2. Kullanılan Duyarlılık Fonksiyonu(The Used Fitness Function)

Bu çalışmada sınıflandırma kural çıkarımı için CLONALG algoritması kullanılmıştır. Duyarlılık fonksiyonu için denklem (8) kullanılmaktadır [22]:

$$F = \frac{TP * TN}{(TP + FN) * (FP + TN)} \quad (8)$$

Burada *N* örneklerin toplam sayısıdır. ***TP*** (True Positives/Doğru Pozitifler) kural tarafından sınıfları doğru tahmin edilmiş örneklerin toplam sayısıdır. ***FP*** (FalsePositivies/Yanlış Pozitifler) kural tarafından sınıfları yanlış tahmin edilmiş örneklerin toplam sayısıdır. ***FN*** (FalseNegativies/Yanlış Negatifler) kuralın ait olduğu sınıfa ait ama bu kural tarafından belirlenememiş örneklerin toplam sayısıdır. ***TN*** (True Negativies/Doğru Negatifler) kural tarafından belirlenememiş ve başka sınıfa ait olan örneklerin toplam sayısıdır.

3.3. Kural Çözme (Rule Decoding)

Son aşama olarak bulunan en iyi antikorun kodu çözülür. Kural çözme işlem adımları kaba kodu Şekil 2'de gösterilmektedir.

Adım 1. Optimizasyon sonucu elde edilmiş antikorların kodu çözülür.
Adım 2. Her bir antikor iki alt kümeye ayrılır; Her antikorun ilk *m* elemanı ayrık nitelikleri belirtir ve *X* ile gösterilir. Diğer *t* eleman gerçek nitelikleri belirtir ve XR ile gösterilir.
- ***Adım 2.1.*** Ayrık nitelikleri belirten *X*'in kodu çözülür.
  - *Adım 2.1.1*. *X* kümesi *N* parçaya bölünür. Her parça A_n niteliğini belirtir ve uzunluğu m_n'dir.
  - *Adım 2.1.2*. Denklem (1)'e göre kodu çözülür.
  - *Adım 2.1.3*. Aynı niteliğin farklı değerlerini birleştirmek için "VEYA" kullan. Farklı nitelikleri birleştirmek için "VE" kullanılır.
- ***Adım 2.2.*** Gerçek nitelikleri belirten *XR*'nin kodu çözülür.
  - *Adım 2.2.1.XR*' yi iki parçaya böl. İlk parça aralıkların orta noktalarını, ikinci parça genişlemelerini temsil eder.
  - *Adım 2.2.2*. (4)'e göre kodu çözülür.
  - *Adım 2.2.3*.Farklı özelliklerini birleştirmek için "VE" kullanılır.

Adım 3. *Adım 2.1.3* ve *Adım 2.2.3*'te bulunmuş alt kuralları "VE" ile birleştirerek kural tabanını oluşturulur.

Şekil 2. Kural çözme işleminin kaba kodu (Pseudocode of the rule decoding prosess)

4. DENEYSEL SONUÇLAR (EXPERIMENTAL RESULTS)

Önerilen yöntem dört farklı veri kümesine uygulanmıştır. Kullanılan veri kümeleri deniz kabuğu, şarap, süsen çiçeği ve cam kimliklendirmedir.

Uygulama için ilk adım olarak gerçek özellikler [0-1]aralığındadenklem (9)'a göre normalize edilmiştir.

$$x_n = \frac{x_r - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

Burada x_r niteliğin orijinal değerini, x_{min} bütün örneklerin bu özelliğine ait minimum değeri,

x_{max} bütün örneklerin bu özelliğine ait maksimum değeri, x_n niteliğin normalize edilmiş değerini ifade etmektedir. Daha sonra ayrık veriler denklem(1) kullanılarak kodlanmıştır.

Algoritmanın ikinci adımında popülasyon oluşturulmuştur. Popülasyondaki her bireyin uzunluğu denklem(7) kullanılarak hesaplanmıştır. Üçüncü adımda, CLONALG her sınıf için ayrı ayrı çalıştırılmış, uygunluk fonksiyonu olarak (denklem (8)) kullanılmıştır. (denklem (8)) sınıflandırma başarısını belirlediği için CLONALG maksimizasyon amaçlı kullanılmıştır. Durdurma kriteri olarak sonuçların iyileşmesinin durması belirlenmiştir.

Farklı sınıflara ait örnek sayıları arasında ciddi fark olan veri tabanlarında (örneğin Bölüm 4.4. Cam kimliklendirme) sınıf vektör içine gömüldüğü zaman optimizasyon sonucu genellikle örneği daha fazla olan sınıflara ait kurallara oluşmaktadır. Bunu önlemek amacıyla her sınıf için ayrı ayrı optimizasyon yapılmıştır. Sonuç olarak üretilen her vektör bir kural belirtmektedir. Kuralları görmek için bulunan vektörlerin kodu çözülmelidir.

4.1. Deniz Kabuğu Veri Kümesi(Abalone Dataset)

Uygulama amacıyla kullanılan ilk veri kümesi, UCI sitesinden alınan deniz kabukları verileridir. Deniz kabuğunun yaşını hesaplamak için kabuk konisi dikey olarak kesildikten sonra boyanarak halkalar sayılmaktadır. Halka sayımı işlemi mikroskop kullanılarak yapılmaktadır. Halkalarının sayısına1,5 eklendikten sonra deniz kabuğunun yaşı elde edilir. Bu iş pahalı ve zaman alıcı bir işlemdir. Dolaysıyla yaşların belirlenmesi için bilgisayarlı yöntemlerin kullanılmasının çok daha etkili, hızlı ve ucuz olacağı öngörülmektedir. Deniz kabuğu veri kümesi 8 nitelik ve 29 sınıftan oluşmaktadır. Bu sınıflar Clark ve arkadaşlarının çalışmaları kullanılarak birleştirilmiştir [23]. Sınıf 1 (1'den 8'e kadar olan sınıflar), Sınıf 2 (Sınıf 9 ve sınıf 10), ve Sınıf 3 (11'den 29'a kadar olan sınıflar) olarak belirlenmiştir. Veri kümesi 4177 örnekten oluşmaktadır. Deniz kabuğu veri kümesinin nitelikleri ve değer aralıklarıTablo 1'de gösterilmiştir.

Tablo 1. Deniz kabuğuveri kümesinin nitelik ve aralık değerleri(Abalone dataset attribute and range values)

Nitelik	Aralık Değerleri
Cinsiyet– (Erkek–Dişi–Çocuk)	M,F,I
Uzunluk– (LG)	0,075 – 0,815
Çap– (DM)	0,055 – 0,650
Yükseklik– (HE)	0 – 1,130
Toplam Ağırlık– (WW)	0,002 – 2,825
Kabuksuz Ağırlık– (SUW)	0,001 –1,488
İç OrganAğırlığı – (VW)	0,0005 – 0,760
Kabuk Ağırlığı– (SHW)	0,0015 – 1,005

Önerilen yöntem deniz kabuğu veri kümesine uygulanmıştır. Bunun için önce ilk popülasyon oluşturulmuştur. Bu veri kümesinde sadece cinsiyet ikili değer (M, F, I), diğer 7 özellik gerçel değerler aldığı için vektörün ikili kısmının uzunluğu 3, gerçel kısımının uzunluğu 14'tür. İlk popülasyonu oluşturmak için cinsiyet özelliğine rastgele bir değer atamak amacıyla ikili kısımdaki değerlerden rastgele birinin değeri 1 yapılmıştır, gerçel kısım "random" fonksiyonu kullanılarak üretilmiştir. Deniz Kabuğu veri kümesi için oluşturulan ilk popülasyondaki vektörlerden biri aşağıdaki gibi oluşmuştur: [1; 0; 0, 0,6551; 0,3112; 0,9027; 0,8116; 0,5211; 0,0424; 0,6663; 0,5985; 0,4624; 0,8699; 0,6834; 0,9296; 0,0426; 0,3478]. Bir sonraki aşamada CLONALG çalıştırılmıştır.

Deniz kabuğu veri kümesi için Sınıf 1 için geçerli olmak üzere ilk kural {1;0; 0;-0,5229, -0,0475; -0,3153; -0,0377; -0,3003; -0,0024; -0,2508; 0,7149; 0,4724; 0,5999; 5,4553; 0,7939; 0,7301; 0,4138}olarak bulunmuştur. İlk üç sayı {1,0,0} cinsiyeti belirler, burada yalnızca ilk sayı 1 olduğundan cinsiyet erkek anlamındadır. Elde edilen vektörün 4-10 elemanları nitelikler için aralıkların orta noktalarını, 11-17 elemanları ise aralıkların genişlemelerini temsil etmektedir. Vektörlerin kodlarının çözülmesiyle "Eğer Cinsiyet=Erkek & LG ϵ (-0,5229; 0,7149) & DM ϵ (-0,0475; 0,4724) & HE ϵ (-0,3153; 0,5999) & WW ϵ (-0,0377; 5,4553) & SUW ϵ (-0,3003; 0,7939) & VW ϵ (-0,0024; 0,7301)& SHW ϵ (-0,2508; 0,4138) O halde Sınıf 1" kuralı elde edilir.

Veri kümesinde LG niteliği[0,075; 0,815] aralığında değerler aldığından ve -0,5229<0,075 olduğundan LG ϵ (-0,5229; 0,7149) yerine LG<0,7149 yazılabilir. Benzer bir mantıkla DM<0,4724; HE<0,5999; SUW<0,7939; VW<0,7301 ve SHW<0,4138 olarak yazılabilir. WW[0,002; 2,825] aralığında değerler aldığından ve [0,002; 2,825]$\subset$(-0,0377; 5,4553) olduğundan WW ϵ (-0,039669; 5,457325) ifadesinin bir anlamı kalmamaktadır ve bu koşulun kuraldan kaldırılması başarıyı etkilemeyeceği gibi karar verme hızı artıracaktır. Sonuç olarak kodları çözülmüş kural kısaca "Eğer Cinsiyet = Erkek & LG<0,7149& DM<0,4724& HE<0,5999 &SUW <0,7939& VW <0,7301& SHW<0,4138 O halde Sınıf1" şeklinde ifade edilir.Kuralın ilk temsili yanlış değildir ama okunması zordur. Bununla beraber ikincisi daha anlaşılır ve kısadır.

Sonuç olarak 54 kural (Her bir sınıf için 18 kural) oluşturulmuştur. Bu veri kümesinde sınıflandırma başarısı %62,59 olarak elde edilmiştir. Literatürde bu veri kümesi üzerinde herhangi bir kural çıkarım çalışmasına rastlayamadığımızdan dolayı sonuçlar karşılaştırılamamıştır. Veri kümesinden çıkarılan kurallar (her sınıf için 2 kural) Tablo 2'de verilmiştir.

Tablo 2. Deniz kabuğuveri kümesi kuralları (Abalone dataset rules)

Kural Numarası	Kurallar
1	**Eğer**Cinsiyet=Erkek &LG < 0,7149 & DM < 0,4724 & HE < 0,5999 & SUW < 0,7939 & VW < 0,7301 & SHW < 0,4138 **O HaldeSınıf 1.**
2	**Eğer** LG < 0,7823 & DM < 0,5504 & WW < 1,9368 & SUW < 1,0222 & SHW < 0,1724 **O HaldeSınıf 1.**
3	**Eğer**LG < 0,7418 & DM ϵ (0,4252; 0,6395) & HE < 0,9937 & WW ϵ (0,1917; 1,5756) & SUW < 1,0693 & VW < 0,5329**O HaldeSınıf 2.**
4	**Eğer**LG > 0,0995 & DM < 0,5132 & WW > 0,0119 & SUW > 0,1214 & VW > 0,1025 & SHW > 0,2038 **O HaldeSınıf 2.**
5	**Eğer**DM > 0,0672 & HE > 0,1445 & WW > 0,4614 & SUW > 0,1635**O HaldeSınıf 3.**
6	**Eğer**Cinsiyet ≠ Erkek &DM > 0,2144 & WW > 0,6124 & SUW > 0,1480& SHW < 0,5461 **O HaldeSınıf 3.**

4.2. Şarap Veri Kümesi (Wine Dataset)

Uygulama için kullanılan ikinciveri kümesi, UCI sitesinden alınan şarap veri kümesidir. Bu veri kümesinin on üç niteliği vardır. Veri kümesi Stefan Aeberhard, M. Forina ve arkadaşları tarafından oluşturulmuştur. Bu veriler aslında İtalya'nın aynı yöresinde yetişen üzümlerden yapılmış şarapların kimyasal analizlerinin sonuçlarıdır ve üç farklı yetiştiriciden elde edilen verilerdir. Analizlerde üç tip şarap için bulunan 13 bileşenin miktarı belirtilmiştir [23]. Sınıflar, sınıf 1, sınıf 2 ve sınıf 3 olarak kodlanmıştır. Veri kümesi 178 örneğe sahiptir. Şarap veri kümesinin nitelik ve aralık değerleri Tablo 3'de gösterilmiştir.

Tablo 3. Şarap veri kümesinin nitelik ve aralık değerleri(Wine dataset attribute and range values)

Nitelik	Aralık Değerleri
Alkol - (ALC)	11,03 – 14,83
Malik Asit - (MAC)	0,74 – 5,80
Ash - (ASH)	1,36 – 3,23
Bazik Ash - (ALA)	10,60 – 30,00
Magnezyum - (MAG)	70,00 – 162,00
Toplam Phenols – (TPH)	0,98 – 3,88
Flavanoids - (FLA)	0,34 – 5,08
Nonflavanoid phenols - (NPH)	0,13 – 0,66
Proanthocyanins - (PRT)	0,41 – 3,58
Renk Yoğunluğu – (CIN)	1,28 – 13,00
Renk - (HUE)	0,48 – 1,71
OD280/OD315 ile Sulandırılmış - (ODW)	1,27 – 4,00
Proline - (PRL)	278,00 – 1680,00

Önerilen yöntem bu verilere uygulanmış ve sonuç olarak 15 kural (Her sınıf için 5 kural) oluşturulmuştur. Bu veri kümesinden çıkarılmış olan kurallar (Örnek olarak her sınıf için 2 kural) Tablo 4'de verilmiştir. Bu veri kümesindeki örneklerin %99,44'u doğru olarak sınıflandırılmıştır.

Tablo 4. Şarapveri kümesi kuralları (Wine dataset rules)

Kural Numarası	Kurallar
1	**Eğer** ALC > 11,7181 & ASH > 1,9503 & FLA > 2,0453 & CIN ϵ (3,3683; 9,7618) & HUE ϵ (0,784; 1,5591) **O HaldeSınıf 1.**
2	**Eğer** ALC > 12,3359 & MAC < 5,7084 & ALA < 21,8451 & MAG > 76,3762 & FLA > 0,7239 & NPH < 0,6055 & PRT < 2,9912 & CIN < 9,5682 & HUE > 0,7964 & ODW > 2,1572 & PRL > 685,9059 **O HaldeSınıf 1.**
3	**Eğer** ALC < 13,0541 & ALA > 12,8089 & MAG < 158,7653 & FLA > 0,7 & CIN < 9,5164 & HUE > 0,5674 & ODW > 1,6852 **O HaldeSınıf 2.**
4	**Eğer** ALA < 28,5601 & TPH > 1,0261 & FLA < 2,6819 & HUE ϵ (0,7921; 1,6072) & ODW > 1,4223 & PRL < 1016,9881 **O HaldeSınıf 2.**
5	**Eğer**ALC ϵ (12,0835; 14,7932) & MAG ϵ (74,4699; 138,1886) & TPH < 2,26 & FLA < 3,8919 & ODW < 3,2486 **O HaldeSınıf 3.**
6	**Eğer** MAC > 1,1863 & ASH > 2,0526 & MAG < 142,4527 & TPH < 3,8601 & FLA < 2,4885 & PRT > 0,522 & CIN > 3,6158 & ODW < 3,3188 **O HaldeSınıf 3.**

Önerilen yöntemin sınıflandırma sonuçları literatürdeki çalışmalarla karşılaştırılmış ve Tablo 5'te verilmiştir. Tablo 5'te de görüldüğü üzere, önerilen yöntem kural çıkarımı için en iyi sonucu vermiştir. Ayrıca önerilen yöntemin hesaplama karmaşıklığı *O(n)* olarak bulunmuştur. Tablo 5'te 2. Sırada bulunan C4.5 algoritmasının karmaşıklığı ise *O(nlogn)*'dir. Bu da yöntemin başarılı sınıflandırma yapmasının yanı sıra, daha düşük karmaşıklığa sahip olduğunu göstermektedir.

Tablo 5. Şarapveri kümesine ait deneysel sonuçlar(Wine dataset experimental results)

Yöntem	Doğruluk%	Referans
Önerilen Yöntem	**99,44**	**Köklü ve ark.**
C4.5	97,12	Sarkar ve ark. [11]
RISE	96,80	Rijnbeek and Kors [25]
RBC using RS	96,78	Sarkar ve ark. [11]
PDGFR	95,18	Nojima ve ark. [26]
EXPLORE	94,10	Rijnbeek and Kors [25]
PART	94,00	Rijnbeek and Kors [25]
C4.5R	93,70	Rijnbeek and Kors [25]
CN2	93,30	Rijnbeek and Kors [25]
SL^2	92,60	Rijnbeek and Kors [25]
RIPPER	92,10	Rijnbeek and Kors [25]
CBA	86,70	Rijnbeek and Kors [25]
IR	86,40	Rijnbeek and Kors [25]
CDF	86,10	Fu and Wang [27]

4.3. Süsen Çiçeği Veri Kümesi (Iris Dataset)

Kullanılan üçüncü veri kümesi, UCI sitesinde yer alan R.A.Fisher'in oluşturduğu süsen çiçeği veri kümesidir. Süsen çiçeği veri kümesi3 sınıflıdır ve4 niteliği ölçülmüştür.Bu veri kümesi her sınıftan 50 örnek olmak üzere toplam 150 örneğe sahiptir. Sınıf 1 (IrisSetosa), Sınıf 2 (IrisVersicolor) ve Sınıf 3 (IrisVirginica) olarak sınıflandırılmıştır [23]. Süsen çiçeği veri kümesinin nitelik ve aralık değerleri Tablo 6'da gösterilmiştir.

Tablo 6. Süsen çiçeği veri kümesinin nitelik ve aralık değerleri(Iris dataset attribute and range values)

Nitelik	Aralık Değerleri
Çanak YaprağınınUzunluğu - (SL)	4,3 – 7,9
Çanak YaprağınınGenişliği- (SW)	2 – 4,4
Taç Yaprağının Uzunluğu - (PL)	1 – 6,9
Taç Yaprağının Genişliği -(PW)	0,1 – 2,5

Önerilen yöntem süsen çiçeği veri kümesine uygulanmıştır ve toplam 7 kural (1 kural Sınıf 1 için, 3 kural Sınıf 2 için ve 3 kural Sınıf 3 için) elde edilmiştir.Çıkan kurallar Tablo 7'de verilmiştir.

Tablo 7. Süsen çiçeği veri kümesi kuralları (Iris dataset rules)

Kural Numarası	Kurallar
1	**Eğer**SW > 2,2961& PL < 2,996 & PW < 1,4778 **O HaldeSınıf 1.**
2	**Eğer**SW <4,094 & PL ϵ (2,9529; 5,5484) & PW < 1,6001 **O HaldeSınıf 2.**
3	**Eğer**SW ϵ (2,8232; 3,3694) & PL ϵ (4,6651; 5,0689) **O HaldeSınıf 2.**
4	**Eğer**SW ϵ (2,3993; 2,7015) & PL < 5,2246 & PW ϵ (0,6250; 1,6500) **O HaldeSınıf 2.**
5	**Eğer**PL >4,7428**O HaldeSınıf 3.**
6	**Eğer**SL < 7,6554 & SW < 2,9679 & PL > 3,3016 & PW > 1,4345 **O HaldeSınıf 3**
7	**Eğer** SL ϵ (5,9947; 6,7277) & SW < 3,7341 & PW > 1,7034 **O HaldeSınıf 3.**

Bu veri kümesiiçin örneklerin %100'ü doğru olarak sınıflandırılmıştır. Önerilen yöntemin sınıflandırma sonuçlarıliteratürdeki çalışmalarla karşılaştırılmış ve Tablo 8'de verilmiştir. Tablo 8'dende görüldüğü üzere önerilen yöntem kural çıkarımı için en iyi sonuçlardan birisi olmuştur.

Tablo 8. Süsen çiçeğiveri kümesine ait deneysel başarılar(Iris dataset experimental results)

Yöntem	Doğruluk%	Referans
Önerilen Yöntem	**100,00**	**Köklü ve ark.**
Taco-Miner	100,00	Özbakır ve ark. [27]
PSO-Miner	100,00	Özbakır and Delice [28]
Grobian (rough)	100,00	Browne [29]
PVM 2 Rules	98,00	Weiss and Kapouleas [30]
C-MLP2LN	98,00	Duch ve ark. [31]
SSV	98,00	Duch ve ark. [31]
PVM 1 Rule	97,30	Weiss and Kapouleas [30]
Minerva	97,00	Huysmans ve ark. [41]
NEFCLASS	96,67	Nauck ve ark. [32]
C4.5	96,64	Sarkar ve ark. [11]
CART	96,00	Weiss and Kapouleas [30]
IRID3 Rule Base	96,00	Pal and Chakraborty [33]
HREM	95,78	Wang ve ark. [34]
RBC using RS	95,37	Sarkar ve ark. [11]
RS	95,10	Wang ve ark. [34]
Ripper	95,00	Huysmans ve ark. [41]
ADERF	94,67	Castellanno ve ark. [35]

4.4. Cam Kimliklendirme Veri kümesi (Glass Identification Dataset)

Uygulama amaçlı kullanılan son veri kümesi, UCI sitesinden alınan cam sınıflandırma verileridir. Veri kümesi on niteliğe sahiptir. Bu veri kümesiDiagnostic Products Corporation'dan Vina Spiehler tarafından oluşturulmuş ve kriminolojik inceleme amaçlı kullanılmıştır. Cam kimliklendirme veri kümesi 9 nitelik, 7 sınıftan ve 214 örnekten oluşmaktadır.Fakat 4. sınıfta veri bulunmadığından uygulamada kullanılmamıştır.Cam tipleri, sınıfları ve örneklem sayıları aşağıdakilerdir[23]:

- Sınıf 1 bina penceresi yayılarak işlenmiş (70 adet),
- Sınıf 2 bina penceresi yayılmadan işlenmiş (76 adet),
- Sınıf 3 araç camı yayılarak işlenmiş (17 adet),
- Sınıf 4 araç camı yayılmadan işlenmiş
- (0adet, veri kümesinde hiç yoktur bu sebeple kural çıkarılmamıştır),
- Sınıf 5 kavanozlar (13 adet),
- Sınıf 6 sofra takımları (9 adet),
- Sınıf 7gece lambaları (29 adet).

Cam kimliklendirme veri kümesinin nitelik ve aralık değerleri Tablo 9'da gösterilmiştir.

Tablo 9. Cam kimliklendirmeveri kümesinin nitelik ve aralık değerleri(Glass identification dataset attribute and range values)

Nitelik	Aralık Değerleri
Kırılma Katsayısı - (RI)	1,5112 – 1,5339
Sodyum - (NA)	10,73 – 17,38
Magnezyum - (MG)	0 – 4,49
Aliminyum- (AL)	0,29 – 3,5
Silikon - (SI)	69,81 – 75,41
Potasyum - (K)	0 – 6,21
Kalsiyum - (CA)	5,43 – 16,19
Baryum- (BA)	0 – 3,15
Demir - (FE)	0 – 0,51

Önerilen yöntem bu veri kümesine uygulanmıştır. Sonuç olarak 78 kural (her bir sınıf için13 kural, veri örneği olmaması nedeniyle sınıf dört için kural oluşturulmamıştır) oluşturulmuştur. Çıkarılmış kurallar her sınıf için bir örnek olmak üzere Tablo 10'da verilmiştir. Üretilen kural kümesi %77,1 oranında doğru sınıflandırma yapmaktadır.

Tablo 10. Cam kimliklendirme veri kümesi kuralları (Glass identification dataset rules)

Kural Numarası	Kurallar
1	**Eğer** RI > 1,5152 & NA < 17,2839 & AL < 1,4478 & SI < 74,8362 & CA < 10,2991 & BA < 0,9945 & FE < 0,2756 **O HaldeSınıf 1.**
2	**Eğer**RI > 1,5144 & NA > 12,0406 & MG ϵ (1,9662; 4,3776) & AL > 1,0244 & SI < 73,5057 & K < 5,9072 & CA ϵ (7,632; 8,6153) & BA < 1,9393 & FE < 0,3321**O HaldeSınıf 2.**
3	**Eğer**RI ϵ (1,5134; 1,5195) & NA < 17,1445 & MG ϵ (2,8642; 4,4483) & SI ϵ (71,8428; 73,0494) & K < 2,7148 & CA < 9,0695 & BA < 3,1004**O HaldeSınıf 3.**
4	**Eğer** RI < 1,5332 & MG < 2,5352 & AL ϵ (1,2216; 3,1499) & SI > 70,1034 & CA < 13,1166 & BA < 0,1974**O HaldeSınıf 5.**
5	**Eğer** NA > 13,3564 & MG < 2,9461 & SI > 71,8245 & K < 5,9972 & BA < 1,3884 & FE < 0,3049 **O HaldeSınıf 6.**
6	**Eğer** RI < 1,5305 & NA > 11,2075 & AL ϵ (1,0141; 3,2659) & BA > 0,1908 & FE < 0,1854 **O HaldeSınıf 7.**

Önerilen yöntemin sınıflandırma sonuçları literatürdeki çalışmalarla karşılaştırılmış ve Tablo 11'de verilmiştir. Tablo 11'dende görüldüğü üzere önerilen yöntem kural çıkarımı için en iyi sonucu vermiştir.

Tablo 11. Cam kimliklendirme veri kümesine ait deneysel başarılar(Glass identification dataset experimental results)

Yöntem	Doğruluk%	Referans
Önerilen Yöntem	**77,10**	**Köklü ve ark.**
C4.5	72,85	Sarkar ve ark. [11]
RBC using RS	71,87	Sarkar ve ark. [11]
HREM	63,89	Wang ve ark. [35]
RS	60,42	Wang ve ark. [35]
FRBC	58,28	Kim ve ark. [37]
IRID3 Rule Base	52,70	Pal and Chakraborty [34]
ADERF	50,95	Castellanno ve ark. [36]
GGP	35,20	Pappa and Freitas [38]

5. SONUÇLAR VE ÖNERİLER (CONCLUSIONS AND RECOMMENDATIONS)

Literatürde birçok kural çıkarım çalışması vardır. Bu kural çıkarım çalışmalarının çoğunluğu YSA veya SVM tabanlı çalışmalardır. Ek olarak, araştırmacılar bulanık sistemleri veya verinin isteğe bağlı olarak standart aralıklara ayrıldığı ikili veri kodlamayı kullanmışlardır. Bu çalışmada, önceki çalışmalardan farklı olarak kurallar veriden doğrudan çıkartılmaktadır. Ayrıca literatürde veri kodlama işleminde genellikle ikili kodlama sistemi kullanıldığı görülmektedir. Bunun için her nitelik aralığı belirli bir mantıkla önceden belirlenmiş *n* sayıda alt aralıklara bölünerek *n* uzunluğunda bir ikili katar (string) oluşturularak, verinin değeri hangi alt aralığa aitse ona karşılık gelen değer 1, diğer (n-1) değer 0 olarak kodlanmaktadır. Bu çalışmada ise alt aralıklar optimizasyon aracılığıyla belirlendiği için, daha etkin şekilde oluşturulmaktadır. Böylece bu yöntem her nitelik için aralıkları optimizasyon kullanarak hassasiyetle belirlemektedir. Yöntem aşağıdaki adımlardan oluşur;

- Niteliklerin ayrık ve gerçekdeğerli olarak ayrılması
- Gerçekdeğerli özelliklerin 0 ile 1 arasında normalize edilmesi.
- Ayrık özelliklerin ikili kodlanması
- Popülasyonun rastgele oluşturulması
- Sınıflandırılma kurallarının optimize edilerek çıkarılması
- Kurallarınkodunun çözülmesi

Optimizasyon amacıyla CLONALG algoritması kullanılmıştır. Sunulmuş yaklaşım, dört gerçek durum sınıflandırma problemine uygulanmıştır. Veriler UCI sitesinden alınmıştır[23]. Önerilen yöntemilesüsen çiçeği (iris) için %100, şarap (wine) için %99,44, cam kimliklendirme (glassidentification) için %77,10 ve deniz kabuğu (abalone) için %62,59 doğrulukla sınıflandırma yapılmıştır. Önceki araştırmalarla karşılaştırıldığında, önerilen metot ile daha başarılı sonuçlar elde edildiği görülmüştür. Ayrıca önerilen yöntemin hesaplama karmaşıklığı *O(n)*'dir [39]. Diğer yöntemlerin karmaşıklığı incelendiğinde örneğin C4.5 algoritmasının hesaplama karmaşıklığının *O(nlogn)* olduğu görülmektedir [40]. Bu da önerilen yöntemin karmaşıklık açısından da üstün olduğunu gösterir.

Bu çalışmadaki bütün problemlerde çok sınıflı veri kümesi kullanılmıştır. Bilindiği gibi çok sınıflı veri kümeleri, kural çıkarımı için en zor veri kümeleridir. Çok sınıflı kural çıkarımı, için literatürde birçok yöntem önerilmiştir. Ancak tarafımızdan önerilen yöntem literatürdeki yöntemlerden daha iyi sonuçlar verdiği gözlemlenmiştir.

Teşekkür: Bu çalışma Selçuk Üniversitesi Bilimsel Araştırma Projeleri Koordinatörlüğü tarafından desteklenmektedir.

6. KAYNAKLAR (REFERENCES)

1. Kahramanlı H., and Allahverdi N., "Rule Extraction from Trained Adaptime Neural Networks Using ArtificialİmmuneSystems", **Expert Systems With Applications**, Sayı 36, pp, 1513-1522, 2009.
2. Gallant, S. I., "Connection Expert Systems",**Communications of the ACM**, 31(2), pp.152–169, 1998.
3. Towell, G. G., and Shavlik, J., "Extracting Refined Rules From Knowledge-Based Neural Networks". **Machine Learning**, 13, pp.71–101, 1993.
4. Lu, H., Setiono, R., and Liu, H., "Effective Data Mining Using Neural Networks",**IEEE Transactions on Knowledge and Data Engineering**, 8(6), pp.957–961, 1996.
5. Keedwell, E., Narayanan, A., and Savic, D., "Evolving rules from neural networks trained on

continuous data",**Evolutionary Computation Congress,**Proceedings Of The Congress On Evolutionary Computation, 1, pp.639-645, 2000a.

6. Keedwell, E., Narayanan, A., and Savic, D., "Creating Rules From Trained Neural Networks Using Genetic Algorithms"**International Journal of Computers Systeming Signals (IJCSS)**, 1(1), pp.30–42, 2000b.
7. Aliev R.A., Aliev R.R., Guirimov B. and Uyar K., "Dynamic Data Mining Technique for Rules Extraction in a Process of Battery Charging", **Applied Soft Computing,** 8, pp. 1252–1258, 2008.
8. Ang J.H., Tan K.C. and Mamun A.A., "An Evolutionary Memetic Algorithm For Rule Extraction", **Expert Systems with Applications**, Volume 37, 2, pp.1302-1315, 2010.
9. Papageorgiou E.I., "A New Methodology for Decisions in Medical Informatics Using Fuzzy Cognitive Maps Based on Fuzzy Rule-Extraction Techniques", **Applied Soft Computing,** 11, pp.500–513, 2011.
10. Rodríguez M., Escalante D. M., and Peregrín A., "Efficient Distributed Genetic Algorithm for Rule Extraction", **Applied Soft Computing**, Vol.11, 1, pp. 733-743,2011.
11. Sarkar, B.K., Sana S.S. and Chaudhuri K., "A Genetic Algorithm-Based Rule Extraction System",**Applied Soft Computing** 12,pp.238–254, 2012.
12. Özbakır L., Baykasoğlu A. and Kulluk S., "A Soft Computing-Based Approach for Integrated Training and Rule Extraction From Artificial Neural Networks: DIFACONN-Miner", **Applied Soft Computing** 10,pp.304–317, 2010.
13. Baykasoglu A., Saltabaş A., Taşan A.S. and Subulan K., "Yapay Bağışıklık Sisteminin Çoklu Etmen Benzetim Ortamında Realize Edilmesi ve Gezgin Satıcı Problemine Uygulanması" **Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi,** Cilt 27, No 4, 901-909, 2012.
14. Seredynski, F., and Bouvry, P., "Anomaly Detection in TCP/IP Networks Using Immune Systems Paradigm",**Computer Communications**, 30, pp.740–749, 2007.
15. Kalinli, A., and Karaboga, N., "Artificial Immune Algorithm for IIR Filter Design",**Engineering Applications of Artificial Intelligence**, 18, pp.919–929, 2005.
16. Musilek, P., Lau, A., Reformat, M., & Wyard-Scott, L., "Immune Programming",**Information Sciences**, 176, pp.972–1002, 2006.
17. Kumar, A., Prakash, A., Shankar, R., & Tiwari, M. K., "Psychoclonal Algorithm Based Approach to Solve Continuous Flow Shop Scheduling Problem",**Expert System with Applications,** 31, pp.504–514, 2006.
18. De Castro, L. N., & Timmis, J., "Artificial Immune Systems: ANew Computational Intelligence Approach". UK: **Springer**, 2002.
19. Hart, E. and Timmis, J., "Application Areas of AIS: The past, The Present and The Future", **International Conference on Artificial Immune Systems (ICARIS)**, Canada, LNCS 3627, pp. 483-497, 2005.
20. Timmis, J., Hone, A., Stibor, T. and Clarck, E., "Theoretical advances in artficial immune systems", **Theoretical Computer Science,** 403, pp. 11-32, 2008.
21. Brownlee, J., "Clonal Selection Theory & CLONALG the Clonal Selection Classification Algorithm (CSCA)", **Tecnical Report**, 2005.
22. Parpinelli, R.S., Lopes, H.S., Freitas, A.A., "An Ant Colony Based System For Data Mining: Applications To Medical Data",**Proceedings of the Genetic and Evolutionary Computation Conference,** San Francisco, California, pp.791–798, 2001.
23. Frank, A. and Asuncion, A., "UCI Machine LearningRepository",[http://archive.ics.uci.edu/ml], I**rvine, CA: University of California, School of Information and Computer Science**, Last accessed:February, 2011.
24. Clark D., Schreter Z. and Adams A., "A Quantitative Comparison of Dystal and Backpropagation", **Submitted to the Australian Conference on Neural Networks,**1996.
25. Rijnbeek P.R. and Kors J.A., "Finding a Short and Accurate Decision Rule in Disjunctive Normal Form by Exhaustive Search", **Machine Learning,** 80: 33–62, DOI 10.1007/s10994-010-5168-9, 2010.
26. Nojima Y., Ishibuchi H. And Kuwajima I., "Parallel Distributed Genetic Fuzzy Rule Selection (FDGFRS)", **Soft Computing,** 13, pp.511–519, 2009.
27. Ful X. and Wang L., "A Rule Extraction System with Class-Dependent Features", **Studies in Fuzziness and Soft Computing,** Volume 163/2005, pp.79-99, DOI: 10.1007/3-540-32358-9_5, 2005.
28. Özbakır, L., Baykasoğlu, A., Kulluk, S., Yapıcı, H., "TACO-miner: An Ant Colony Based Algorithm for Rule Extraction from Trained Neural Networks", **Expert Systems with Applications,** 36, 12295-12305, 2009.
29. Özbakır, L.,Delice,Y., "Exploring comprehensible classification rules from trained neural Networks integrated with a time-varying binary particle swarm optimizer", **Engineering Applications of Artificial Intelligence,** 24, 491-500, 2011.
30. Browne C., Duntsch I. and Gediga G., "IRIS Revisited: A Comparison Of Discriminant and Enhanced Rough Set Data Analysis". **In: L. Polkowski and A. Skowron, eds. Rough Sets in Knowledge Discovery,** vol. 2. Physica Verlag, Heidelberg, 345-368, 1998.
31. Weiss S.M. and Kapouleas, I., "An Empirical Comparison of Pattern Recognition, Neural Nets

and Machine Learning Classification Methods", in: J.W. Shavlik and T.G. Dietterich, **Readings in Machine Learning**, Morgan Kauffman Publ, CA 1990.

32. Duch W, Adamczak R, Grąbczewski K., "A New Methodology of Extraction, Optimization and Application of Crisp and Fuzzy Logical Rules". **IEEE Transactions on Neural Networks,** 12, pp.277-306, 2001.

33. Nauck D., Nauck U. and Kruse R., "Generating Classification Rules with the Neuro-Fuzzy System NEFCLASS",**Proc. Biennial Conf. of the North American Fuzzy Information Processing Society (NAFIPS'96)**, Berkeley, 1996.

34. Pal N. R. and Chakraborty S., "Fuzzy Rule Extraction From ID3-Type Decision Trees for Real Data", **IEEE Trans. Syst.,** Man, Cybern. B. 31 pp. 745-754, 2001.

35. Wang S., Wu G. and Pan J., "A Hybrid Rule Extraction Method Using Rough Sets and Neural Networks", **Lecture Notes in Computer Science,** 2, 4492, pp. 352–361, 2007.

36. Castellano G., Fanelli A. M., and Mencar C., "An empirical risk functional to improve learning in a neuro-fuzzy Classifier",**IEEE Trans. Syst.,** B. 34 pp. 725-731, 2004.

37. Kim D.W., Park J.B., and Joo Y.H., "Design of Fuzzy Rule-Based Classifier: Pruning and Learning", **Fuzzy Systems and Knowledge Discovery Lecture Notes in Computer Science**, 3613, pp. 416–425, Springer, 2005.

38. Pappa G.L. and Freitas A. A., "Evolving Rule Induction Algorithms with Multi-Objective Grammar-Based Genetic Programming", **Knowledge and Information Systems,** 19, pp.283–309. DOI 10.1007/s10115-008-0171-1, 2009.

39. De Castro, L. N.,, "Learning and Optimization Using the Clonal Selection Principle",**IEEE Transactions On Evolutionary Computation,** Vol. 6, no. 3, June, 2002.

40. C 4.5 Lecture Notes, [http://www.sts.tu-harburg.de/teaching/ss-09/ml-sose-09/03-Decision-Tree-c45.pdf], **Hamburg University of Technology, Lecture Notes in Electronic Science,**Last accessed,: Semptember, 2013.

41. Huysmans J., Baesens B. and Vanthienen J, "Using Rule Extraction to Improve the Comprehensibility of Predictive Models", **Open Access publications from Katholieke Universiteit** Leuven with number urn:hdl:123456789/121060, KBI-0612, 2006.7