

## Diyabet Hastalıkları İçin GA-DFÇ-UÖM Tabanlı Uzman Tanı Sistemi

Akif DOĞANTEKİN<sup>1\*</sup>, Cafer BAL<sup>2</sup>

<sup>1</sup> İç Hastalıkları Bölümü, Gaziantep Emek Hastanesi, Gaziantep, Türkiye  
<sup>2</sup> Mekatronik Mühendisliği, Teknoloji Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye  
<sup>1</sup> akifdahiliye@gmail.com, <sup>2</sup> cbal@firat.edu.tr

(Geliş/Received: 01/12/2021;

Kabul/Accepted: 12/12/2021)

**Öz:** Uç Öğrenme Makinesi (UÖM), regresyon ve sınıflandırma problemleri için yeni bir alandır. Bu çalışmada diyabet teşhisi için Genetik Algoritma-Dalgacık Çekirdek Fonksiyonu-Uç Öğrenme Makinesi (GA-DFÇ-UÖM) yöntemi kullanılmıştır. GA, UÖM' nin gizli nöron sayısını (GNS) ve Dalgacık Çekirdek Fonksiyonu (DFÇ)' nin q, t, j parametre değerlerini optimize etmek için kullanılır. Ayrıca DFÇ-UÖM' nin sınıflandırma performansını artırmak için Genetik Algoritma (GA) kullanılmaktadır. Diyabet Veri Seti (DVS) toplam 768 vaka içermektedir. Bu deneysel çalışmada kullanılan veri seti, UCI veri setinden alınan gerçek diyabet verilerinden oluşmaktadır. Veri seti, DFÇ-UÖM' nin eğitimi ve testi için kullanılır. Sağlıklı ve diyabetik hasta bilgilerinin özellik vektörü, DFÇ-UÖM sınıflandırıcısına girdi olarak sağlanır. Önerilen GA-DFÇ-UÖM yönteminin en başarılı sınıflandırma doğruluğu %98,3 olarak bulunmuştur. Bu başarıya dayalı olarak dalgacık çekirdek fonksiyonunun (DFÇ) q, t, j parametrelerinin değerleri 8, 9 ve 7 olarak bulunmuş ve GNS 140 olmuştur.

**Anahtar kelimeler:** Uç Öğrenme Makinesi, Dalgacık Çekirdeği, Genetik Algoritma, Diyabet Hastalıkları Veri Kümesi, Teşhis Sistemi.

## GA-WKF-ELM Based Expert Diagnosis System for Diabetes Diseases

**Abstract:** The Extreme Learning Machine (ELM) is a new field for regression and computation. For this modeling function, it was designed using the Genetic Algorithm (GA)-Wavelet Kernel (WKF)- Extreme Learning Machine (ELM) method. GA is used for the optimization of Hidden Neuron Number (HNN) and q, t, j parameters of Wavelet Kernel Function (WKF). In addition, Genetic Algorithm (GA) is used for the comparative magnification of CFD-NPM. The Diabetes Dataset (DVS) contains a total of 768 cases. The data set used in this training consists of UCI data. The dataset is used for training and testing of the DFÇ-NPM. The behavior towards healthy and dietetic disease is submitted as input to the CFC-NPM classifier. I think that the GA-DFÇ-NPM method has the most successful results with 98.3%. Based on this structure, the values of the wavelet core (ql, t, j frame were used as 8, 9 and 7 and the GNS was 140).

**Key words:** Edge Learning Machine, Wavelet Core, Genetic Algorithm, Diabetes Diseases Dataset, Diagnostic System.

### 1. Giriş

Kan aynı zamanda ömür boyu şeker hastalığına (diyabet) neden olan ve yüksek şeker değerlerine neden olan kronik bir hastalıktır. Vücudumuzdaki kan şekerini dengelemek için pankreas tarafından insülin adı verilen bir hormon üretilir [1]. Diyabet, vücudun insülini gerektiği gibi üretememesi ve vücudun insülini gerektiği gibi kullanamamasından kaynaklanır [2]. Vücudun şekerini (glikoz) enerji olarak depolanmak üzere kas, yağ ve karaciğer hücrelerine taşımadıkları için şeker hastalarında kan şekeri seviyeleri yüksektir. Diyabet çevresel ve genetik faktörlerden etkilenir. Şu anda dünyadaki yetişkinlerin %6,4'ü diyabetiktir [3]. 2030 yılına kadar diyabetli sayısının 438 milyona ulaşacağı ve yetişkinler için bu oranın %7,4 olacağı tahmin edilmektedir.

Diyabet iki gruba ayrılır [4]. Bunlar yetişkin ve çocuk diyabetidir. Obezite yetişkin diyabetine neden olur. Diyabet hastaları kesin olarak tedavi edilmezler ancak diyet ve egzersizle kontrol altına alınabilirler.

Literatürde diyabetle ilgili birçok çalışma yapılmıştır. Levenberg-Marquardt, BFGS yarı Newton, Gradient, GRNN, RBF diyabet teşhisi için literatürde kullanılmıştır. Diyabet için Gradient iniş [5], GDA-LS-SVM [4], SVM [6], Semi-Naive Bayes, Naive Bayes [7] gibi yöntemler önerilmiştir. Önerilen akıllı sistem, doktorlara diyabet teşhisi konusunda yardımcı oluyor.

Bu çalışmada, diyabet tanısı için GA-DFÇ-UÖM tabanlı sistem önerilmektedir. Bu makalede sınıflandırıcı olarak tek katmanlı sinir ağı (SLNN) sınıflandırıcı, sınıflandırıcı eğitimi için DFÇ-UÖM öğrenme yöntemi kullanılmıştır. DFÇ-UÖM yapısı, DFÇ' nin üç adet düzenlenebilir parametresine sahiptir. q, t, j parametreleri UÖM' nin performansını etkiler. Bu parametrelere ek olarak GNS' nin belirlenmesinde önemlidir. Bu nedenle q, t,

\* Sorumlu yazar: [akifdahiliye@gmail.com](mailto:akifdahiliye@gmail.com). Yazarların ORCID Numarası: <sup>1</sup> 0000-0001-6078-540X, <sup>2</sup> 0000-0002-1199-2637

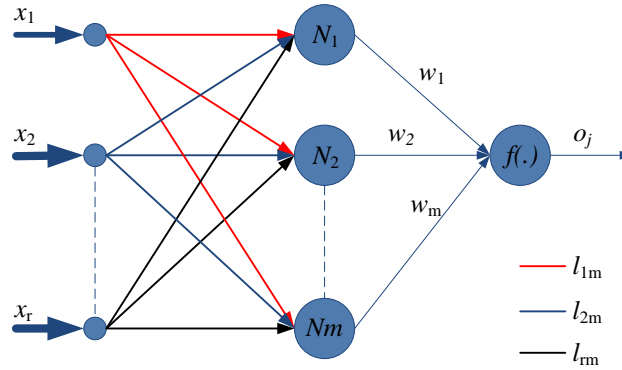
$j$  ve gizli nöronların (GN) değerlerinin doğru olması gerekir. Bu çalışmada, UÖM' nin bu parametrelerin ve GNS' lerinin optimal değerlerini bulmak için GA uygulanmıştır [8].

Bu çalışmada kullanılan veriler UCI Dataset' te elde edilen gerçek DVS' den alınmıştır. DVS, sağlıklı ve hasta olmak üzere iki gruptan oluşur. Önerilen GA-DÇF-UÖM sisteminin performansını değerlendirmek için ROC eğrileri, sınıflandırma doğruluğu, duyarlılık ve özgüllük analizi hesaplanır. GA, UÖM' nin GNS' sı ve DÇF' nin parametrelerini optimize etmek için kullanılır. GA' da bir birey toplam 20 bittten oluşur. Bu bireylerin her birinin 1., 2., 3. ve 4. bitleri, DÇF' lerin  $q$  parametre değerlerini (1 ila 16) temsil eder. Bu bireylerin her birinin 5., 6., 7. ve 8. bitleri, DÇF' lerin  $t$  parametre değerlerini (1 ila 16) temsil eder. Bu bireylerin her birinin 9., 10., 11. ve 12. bitleri, DÇF' lerin  $j$  parametre değerlerini (1 ila 16) temsil eder. Kalan 8 bit HNN'yi gösterir (5 ile 259 arasında). Birinci popülasyonda rastgele 60 kişi seçilmektedir. DDS toplam 768 vaka içerir. WK-ELM sınıflandırıcı, iyi bir genelleme performansı elde etmek için 576 veri seti ile eğitilmiştir. Kalan 192 veri seti, WK-ELM'yi test etmek için kullanılır. Sonuç olarak sınıflandırıcının hedefleri, hasta ve sağlıklı vakalar vardır. GA-DÇF-UÖM yöntemi diyabet tanısı için literatürde kullanılan yöntemlerle karşılaştırılır [8].

Bölüm 2' de UÖM, Bölüm 3' te GA, Bölüm 4' te DÇF-UÖM, Bölüm 5' te Kullanılan veritabanının vadeli işlemleri, Bölüm 6' da GA-DÇF-UÖM' nin diyabet teşhisi için uygulanması, Bölüm 7' de tartışma ve sonuç sırasıyla sunulmaktadır.

## 2. Uç Öğrenme Makinesi (UÖM)

Yapay Sinir Ağları (YSA' lar) problemlerde ve örüntü tanıma uygulamalarında kullanılır. Öğrenme parametrelerinin zor olması, hızlı olmayan öğrenme ve zayıf eğitim performansı YSA' ların dezavantajlarıdır [9],[10]. Huang ve diğerleri, UÖM' yi önerdiler. Gradient tabanlı öğrenme yöntemlerinin dezavantajlarının üstesinden gelmek için [10]. UÖM algoritmaları, tek bir gizli katman (GK) ileri besleme ağının (SLFN) hedef ağırlıklarının yinelemeli öğrenme şemasını kullanmaz. Bunun yerine UÖM, Moore-Penrose'u (MP) genelleştirilmiş tersini kullanarak analitik olarak hesaplar. Şekil 1, UÖM kullanılarak bir SLFN' nin yapısını göstermektedir.  $l_{1m}$ ,  $l_{2m}$  ve  $l_{rm}$ , GN ve giriş nöronlarını birleştiren ağırlık vektörleridir.  $w$ , GN ve hedef nöronun ağırlık fonksiyonudur,  $f(.)$  ÇF' dur.



Şekil 1. UÖM Mimarisi.

UÖM' nin özellikleri aşağıda verilmiştir.

- UÖM, hızlı öğrenme yeteneği nedeniyle sınıflandırma problemde kullanılır.
- UÖM, en azından eğitim hatasını ve en küçük ağırlığı elde etmeyi amaçlar.
- Bu nedenle UÖM, sinir ağları için iyi performans gösterir.
- UÖM' nin öğrenmesi hızlıdır. Bu nedenle SLFN, UÖM ile klasik yöntemlerden daha hızlı eğitilebilir.
- UÖM öğrenme algoritmaları, SLFN' leri eğitmek için kullanılır. UÖM, türevlenemeyen ÇF' lar ile kullanılabilir.
- UÖM, cevaba doğrudan ulaşmayı amaçlar [9]. Aşağıda,  $m$  tane nöron ve  $f$  tane ÇF' a sahip bir SLFN' nin çıkışı bulunmaktadır.

$$o_j = \sum_{i=1}^m \beta_i f(l_i x_r + b_i) \quad (1)$$

Burada, standart SLFN m adet GN ve f adet ÇF' a sahiptir, SLFN çıkışı m tane örneği kullanarak, herhangi bir hata olmadan hesaplar, burada w ve l ağırlıklar, b biastır [9], [11].

$$\sum_{j=1}^m (o_j - t_j) = 0$$

$$t_j = \sum_{i=1}^m w_i f(l_i x_r + b_i), \quad j = 1, 2, \dots, m \quad (2)$$

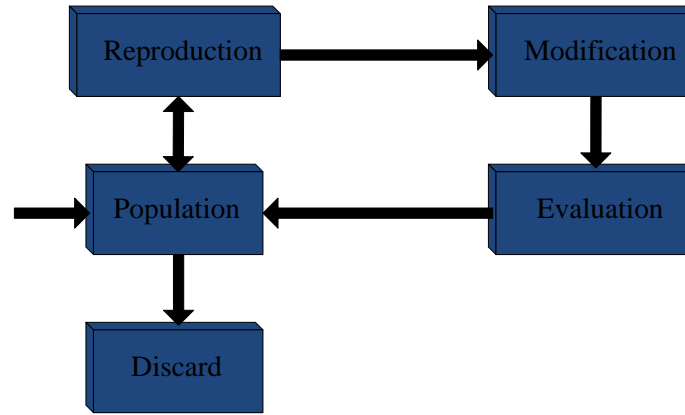
$$H \times w = T \quad (3)$$

Burada H, SLFN' nin hedef matrisidir.

$$H(l_1, \dots, l_m, b_1, \dots, b_m, x_1, \dots, x_m) = \begin{bmatrix} f(l_1 x_1 + b_1) & \dots & f(l_m x_1 + b_m) \\ \vdots & \dots & \vdots \\ f(l_1 x_m + b_1) & \dots & f(l_m x_m + b_m) \end{bmatrix} \quad (4)$$

### 3. Genetik Algoritma

GA, doğal seleksiyon ilkesine dayanan sayısal bir optimizasyon yöntemidir. GA, optimizasyon ve makine öğrenimi uygulamaları için bazı parametreleri ayarlamak için etkili bir yöntemdir [12]. GA tıpta, bilimde ve mühendislikte yaygın olarak kullanılmaktadır. Şekil 2, bir GA'nın yapısını göstermektedir. GA' nın yapısı, kodlama tekniğini, başlatma prosedürünü, değerlendirme fonksiyonunu, ebeveyn seçimini, genetik operatörleri içerir [8].



Şekil 2. Bir GA' nın bileşenleri.

GA' da bireyler kromozomlarla temsil edilir ve kromozomlu bir başlangıç popülasyonu rastgele oluşturulur [13]. Bireyler, permütasyonlar (P11, P3, P7, ..., P14, P15), kurallar listesi (K1, K2, K3, ..., K22, K23), bit dizileri (0101 ... 1100), gerçek sayılar (56.3-78.6 ... 1.4 91.15 ) gibi veri yapıları ile temsil edilir. Aşağıdaki işlem adımlarını kullanarak GA, üreme döneminde popülasyonun bireylerini, popülasyon değerlerine göre rastgele seçer. GA, değerlendirme aşamasında bir kromozomu çözer ve onu bir uygunluk fonksiyonundan geçirerek, uygunluk değerini hesaplar [8]. Aşağıda GA içerisinde yapılan bu işlemler adım adım verilmiştir:

**Adım 1:** Problem için uygun çözüm olan rastgele n tane kromozomdan oluşmuş bir popülasyon elde edilir.

**Adım 2:** Popülasyondaki her bir x bireyi için uygunluk fonksiyonu  $f(x)$  hesaplanır. Bu çalışmada, popülasyondaki her bir birey, rastgele oluşturulur.

**Adım 3:** Popülasyondaki bireyler, en yüksek uygunluk değerlerine göre sıralanarak, istenilen sayıda uygunluk değeri yüksek birey seçilerek, bu bireyler kendi arasında çaprazlanır. Böylelikle, popülasyonun bir sonraki nesilinde de birey sayısının sabit tutabilmesi için eksik birey sayısı tamamlanır.

**Adım 4:** Çaprazlama operatörü uygulanarak, yeni bireyler elde edilir. Eğer bu adımda çaprazlama operatörü uygulanmasaydı, bireyler bir önceki popülasyondaki bireylerle aynı kalırdı.

**Adım 5:** Kromozom üzerindeki bazı diziler, spesifik mutasyon oranına göre yer değiştirir.

**Adım 6:** Yeni elde edilen popülasyon, eski popülasyon ile değiştirilir. Böylece yeni bireyler elde edilir.

**Adım 7:** Nihai koşullar uygunsa, genetik algoritma durdurulur ve mevcut popülasyon en iyi sonucu verir.

**Adım 8:** 2. adım geçilir. Yeni oluşturulan popülasyon, algoritmada [14] döndürülür.

#### 4. DÇF-UÖM

Öğrenme parametrelerinin zor olması, hızlı olmayan öğrenme ve zayıf eğitim performansı YSA' ların dezavantajlarıdır. Son zamanlarda UÖM kullanımı artmıştır [11], [16]. UÖM algoritmasında gizli katmanın başlangıç parametrelerinin ayarlanmasına gerek yoktur. Bu algoritma doğrusal olmayan fonksiyonlarda GN olarak kullanılır. Farklı örnek veri kümeleri  $\{(p_j, m_j) | p_j \in Q^l, m_j \in Q^k, j = 1, \dots, M\}$  için k, bir GN olarak UÖM' nin hedef işlevidir.

$$u_k(p) = \sum_{j=1}^K N_j h_j(p) = h(p)N, \quad (5)$$

$h = [h_1(p), h_2(p), \dots, h_k(p)]$  p girişine göre Gizli Katman (GK)' nin çıkış vektörüdür.  $N = [N_1, N_2, \dots, N_k]$  çıkış nöronları ve GK nöronları arasındaki çıkış ağırlıklarının vektörüdür. h vektörü, giriş alanını UÖM özellik alanına dönüştürür [16]. UÖM' nin yapımında eğitim hataları minimuma indirilmelidir. Bu nedenle eğitim hatası ve çıkış ağırlıkları aynı anda en düşük noktaya indirilmelidir. UÖM kullanımı ile sinir ağlarının genelleştirme yeteneği artmaktadır.

Minimize denklemleri:  $\|TN - W\|, \quad \|N\|.$

$$N = T^T \left( \frac{1}{R} + TT^T \right)^{-1} W, \quad (6)$$

T, GK çıkış matrisini, W, örneklerin beklenen çıkış matrisini ve R, regresyon katsayısını temsil eder. Aşağıda, UÖM öğrenme algoritmasının çıkış fonksiyonu gösterilmektedir.

$$u(p) = h(p)T^T \left( \frac{1}{R} + TT^T \right)^{-1} W \quad (7)$$

$h(p)$  öznitelik vektörü bilinmediği sürece, Mercer'in koşullarına göre UÖM' nin çekirdek matrisi aşağıdaki gibidir:

$$K = TT^T : k_{jz} = h(p_j)h(p_z) = b(p_j, p_z), \quad (8)$$

Böylece,  $u(p)$  UÖM çıkış fonksiyonuna dayalı DÇF aşağıdaki gibidir:

$$u(p) = [b(p, p_1), \dots, b(p, p_M)] \left( \frac{1}{R} + K \right)^{-1} W \quad (9)$$

$K = TT^T$  çekirdek matrisidir.  $b(p, g)$  UÖM' nin ÇF' sidir. Doğrusal çekirdek, polinom çekirdeği, Gauss çekirdeği ve üstel çekirdek gibi bazı ÇF' ları UÖM literatüründe Mercer koşuluna göre fonksiyonlardır. Bu çalışmada Denk. (10), DÇF-UÖM [25]' in performans testi için uygulanmıştır.

$$b(p, g) = \cos \left( q \frac{\|p - g\|}{t} \right) \exp \left( - \frac{\|p - g\|^2}{j} \right) \quad (10)$$

Denklem 11' de gösterilen DÇF' nin eğitim ve test başarısı, sırasıyla doğrusal çekirdek, polinom çekirdeği, Gauss çekirdeği ve üstel klasik ÇF' lerin performansları ile karşılaştırıldığında daha iyidir. DÇF' nin  $q$ ,  $t$  ve  $j$  düzenlenebilir parametreleri, sınır ağlarının performansını etkiler. Bu nedenle bu parametrelerin değerleri probleme göre dikkatli bir şekilde ayarlanmalıdır. GK özelliklerinin ve GNS' nin DÇF-UÖM algoritmalarında bilinmesine gerek yoktur. Ayrıca, DÇF-UÖM öğrenme algoritması, diğer benzer algoritmalarından daha iyi genelleme yeteneğine sahiptir. DÇF-UÖM' nin klasik UÖM' den daha tutarlı olduğu açıktır. DÇF-UÖM diğer sınıflandırıcılardan daha hızlı sonuca ulaşır [16], [17].

## 5. Kullanılan Veritabanının Özellikleri

Bu deneysel çalışmada kullanılan veri seti, UCI veri setinden alınan gerçek diyabet verilerinden oluşmaktadır. DÇF-UÖM' nin eğitimi ve testi için veri seti uygulanır. Bu veritabanı aşağıdaki özelliklerden oluşmaktadır.

- a) Örnek Sayısı: 768' dir.
- b) Özellik Sayısı: 8' dir.

Özellikler, oral glukoz tolerans testinde 2 saatlik plazma glukoz konsantrasyonu, diyastolik kan basıncı, gebelik sayısı, triceps deri kıvrım kalınlığı, saat serum insülini, vücut kitle indeksi, diyabet soy fonksiyonu ve yaştan oluşur. Tablo 1' de veri setinin istatistiksel analizi verilmektedir. Sütun 9, 0 veya 1 değerlerinden birini alır. Sınıf değeri 1, "diyabet için test pozitif" olarak yorumlanır ve sınıf değeri 0, diyabet için "diyabet için test negatif" olarak yorumlanır.

Örneklerin Sınıflara göre dağılımı

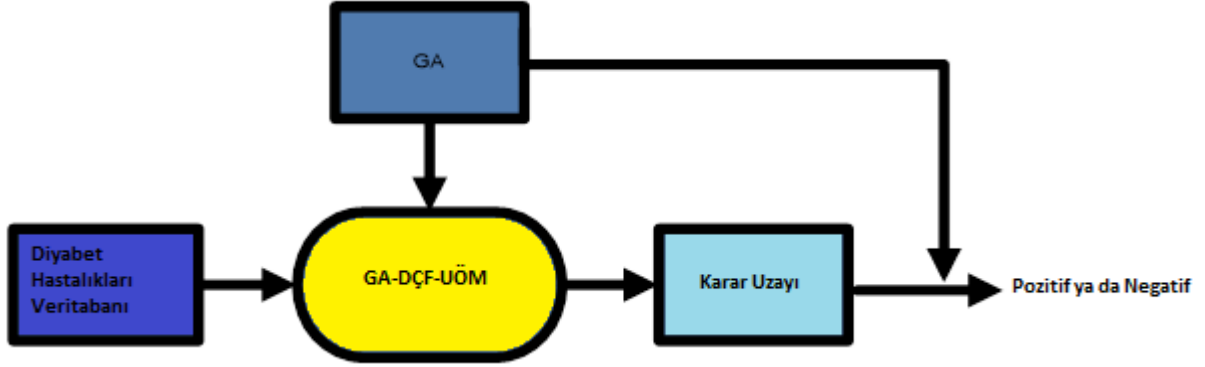
0	500
1	268

**Tablo 1.** Veritabanının istatistiksel analizi

Özellik Numarası	Ortalama	Standart Sapma
1	3.8	3.4
2	120.9	32.0
3	69.1	19.4
4	20.5	16.0
5	79.8	115.2
6	32.0	7.9
7	0.5	0.3
8	33.2	11.8

## 6. GA-DÇF-UÖM Yönteminin Diyabet Hastalıkları Veritabanına Uygulanması

Diyabet hastalıklarının teşhisi için GA-DÇF-UÖM yöntemi tabanlı akıllı teşhis sisteminin akış şeması aşağıda Şekil 3' te gösterilmektedir.



Şekil 3. GA-DÇF-UÖM' nin akış şeması

GA-DÇF-UÖM' nin üç katmanı vardır. İlk katman DVS içerir. DVS toplam 768 örnek içerir. İkinci katmanda GA, UÖM' nin GNS' sini ve DÇF' nin parametrelerini optimize eder. GA yapısında bir birey toplam 20 bitten oluşur.

- Bu bireylerin her birinin ilk dört biti (1., 2., 3. ve 4. bitler), DÇF'lerin q parametre değerlerini (1 ila 16) temsil eder.
- Bu bireylerin her birinin ikinci dört biti (5., 6., 7. ve 8. bit), DÇF'lerin t parametre değerlerini (1 ila 16) temsil eder.
- Bu bireylerin her birinin (9., 10., 11. ve 12. bitler) üçüncü dört biti, DÇF'lerin j parametre değerlerini (1 ila 16) temsil eder.
- Kalan 8 bit GNS' yi (5 ile 259 arasında) gösterir [8].

İlk popülasyon için bireyler rastgele seçilir. DÇF-UÖM, sınıflandırıcının en iyi performansını elde etmeyi amaçlar. DÇF-UÖM sınıflandırıcısı 576 veri seti ile eğitilmiştir. Kalan 192 veri seti, DÇF-UÖM' yi test etmek için kullanılır. Sağlıklı ve hastalıklı durumları sınıflandırıcının çıkışı temsil eder.

DÇF parametresinin optimal değerleri ve UÖM' nin gizli nöron sayısı, optimum değerleri öneren GA-DÇF-UÖM yöntemi ile hesaplanmıştır. GA-DÇF-UÖM, bu alandaki önceki çalışmalardan farklıdır. Sonuç olarak, GA-DÇF-UÖM diyabet teşhisi için çok güçlü ve gerçek zamanlı bir sistem olarak görünmektedir. Bu çalışmada, GA kullanılarak üçlü çapraz doğrulama şeması uygulanmıştır. GA-DÇF-UÖM performansını bulmak için ortalama değerler hesaplanmıştır. İlk olarak, rastgele başlangıç popülasyonu olarak toplam 20 bitten oluşan 60 birey seçilir. DÇF'nin q, t, j parametrelerinin kodlama değerleri Tablo 2' de, GNS' nin kodlaması ise Tablo 3' te verilmiştir.

**Tablo 2.** DÇF' lerin parametreleri için kodlama

q, t, j parametrelerinin değerleri	Kodlama
1	0 0 0 0
2	0 0 0 1
3	0 0 1 0
4	0 0 1 1
5	0 1 0 0
6	0 1 0 1
7	0 1 1 0
8	0 1 1 1
9	1 0 0 0
10	1 0 0 1
11	1 0 1 0
12	1 0 1 1
13	1 1 0 0
14	1 1 0 1
15	1 1 1 0
16	1 1 1 1

**Tablo 3.** GNS' nin Kodlanması

q, t, j parametrelerinin değerleri	Kodlama
5	0 0 0 0 0 0 0 0
6	0 0 0 0 0 0 0 1
7	0 0 0 0 0 0 1 0
8	0 0 0 0 0 0 1 1
.	.
.	.
.	.
.	.
259	1 1 1 1 1 1 1 1

Bireylerin her birinin (1., 2., 3. ve 4. bitler) ilk dört biti, DÇF' lerin q parametre değerlerini (1 ila 16) temsil eder. Bireylerin her birinin ikinci dört biti (5., 6., 7. ve 8. bit), DÇF' lerin t parametre değerlerini (1 ila 16) temsil eder. Bireylerin her birinin üçüncü dört biti (9., 10., 11. ve 12. bit) DÇF' lerin j parametre değerlerini (1 ila 16) temsil eder ve kalan 8 bit GNS' yi (5 ila 259) temsil eder [8].

Bu çalışmada, sınıflandırma doğruluğu, Alıcı Operasyonel Karakter (AOK) eğrileri, duyarlılık ve özgüllük analizi kullanılarak GA-DÇF-UÖM' nin performansı hesaplanmıştır. Veri setleri için sınıflandırma doğrulukları aşağıdaki denklem kullanılarak hesaplanır [20-22].

$$DO(C) = \frac{\sum_{k=1}^{|C|} d(c_k)}{|C|}, \quad c_k \in C \quad (11)$$

$$d(C) = \begin{cases} 1 & \text{if } C(c) = c.m \\ 0 & \text{aksihalde} \end{cases}$$

$$d(c) = \begin{cases} 1, & \text{if } C(c) = c.d \\ 0, & \text{aksihalde} \end{cases} \quad (12)$$

Burada C sınıfların tamamıdır. c.m, m. c sınıfıdır. Duyarlılık ve özgüllük analizinin hesaplanması aşağıda gösterilmiştir.

$$sensitivity(\%) = \frac{TP}{TP + FN} \quad (13)$$

$$specificity(\%) = \frac{TN}{FN + TN} \quad (14)$$

Burada; TP, gerçek pozitifleri temsil eder. TN, gerçek negatifleri temsil eder. FP, yanlış pozitifleri temsil eder. FN, yanlış negatifleri temsil eder [21].

## 7. Sonuçlar ve Karşılaştırmalar

Önerilen GA-DÇF-UÖM' nin performansı, sınıflandırma doğruluğu, AOK eğrisi, duyarlılık ve özgüllük analizi kullanılarak değerlendirilmiştir. Farklı GNS ve farklı ÇF' lara sahip klasik UÖM sınıflandırıcılarının performansı, optimum GNS ve optimal parametreler q, t, j2 nin kullanıldığı GA-DÇF-UÖM yönteminin performansı ile karşılaştırılmıştır. Önerilen GA-DÇF-UÖM' nin en iyi sınıflandırma doğruluğu % 98,3 olarak bulunmuştur. Bu başarıya dayalı olarak DÇF' nin q, t, j parametrelerinin değerleri 8, 9 ve 7 olarak bulunmuş ve GNS 140 olarak bulunmuştur. Tablo 4, önerilen GA-DÇF-UÖM'nin en iyi dört performansını ve parametre değerlerini göstermektedir.

**Tablo 4.** Önerilen GA-DÇF-UÖM' nin en iyi dört performans ve parametre değeri.

Kullanılan Yöntem	ÇF Tipi	DÇF' nin q parametre tipi	DÇF' nin t parametre tipi	DÇF' nin j parametre tipi	GNS	Doğruluk (%)
GA-DÇF-UÖM	Dalgacık	8	9	7	140	98.31
GA-DÇF-UÖM	Dalgacık	6	9	11	66	97.65
GA-DÇF-UÖM	Dalgacık	11	15	13	115	96.58
GA-DÇF-UÖM	Dalgacık	14	2	15	9	95.86

UÖM' nin optimum parametrelerini bulmak için farklı işlemler gerçekleştirilebilir. Gizli Katman Nöron Sayısı (GKNS)' nı kademeli olarak 1'den 100'e 10' ar adım çıkararak, eğitim ve test süreçlerinde farklı aktivasyon fonksiyonları kullanılmıştır [23]. Tablo 5' te UÖM' nin performansı gösterilmiştir.



**Tablo 5.** Farklı parametreler için UÖM' nin performansı.

Aktivasyon Fonksiyonu	GKNS	Sınıflandırma Doğruluğu (%)	
		Eğitim	Test
<i>Logaritmik Sigmoid</i>	10	79.69	76.02
	20	78.65	75.00
	30	77.60	76.50
	40	77.08	76.04
	50	77.08	76.04
	60	80.73	75.52
	70	77.60	75.52
	<b>80</b>	<b>80.73</b>	<b>79.17</b>
	90	80.73	72.40
	100	81.25	74.48
<i>Sinüsoidal</i>	10	78.99	75.52
	20	80.21	77.60
	30	81.60	75.52
	40	80.73	75.00
	50	82.47	76.04
	60	84.03	76.56
	70	82.29	76.04
	80	83.33	76.56
	<b>90</b>	<b>83.16</b>	<b>78.65</b>
	100	83.85	75.00
<i>Tanjant Sigmoid</i>	10	78.47	78.13
	20	81.03	80.21
	30	82.05	81.25
	40	80.73	77.08
	<b>50</b>	<b>85.02</b>	<b>84.61</b>
	60	80.73	80.21
	70	82.64	76.56
	80	84.03	76.56
	90	84.03	75.52
	100	84.55	79.69

Tablo 4 ve Tablo 5 karşılaştırıldığında, GNS' nin optimum değeri GA kullanılarak belirlenerek sınıflandırma performansı iyileştirilmiştir. Ayrıca DÇF-UÖM' nin q, t, j parametreleri GA kullanılarak optimize edilmiştir. Böylelikle, GA-DÇF-UÖM' nin sınıflandırma performansı artırılmıştır.

Tablo 4' de GA-DÇF-UÖM tabanlı sistemin sınıflandırma doğruluğu önceki yöntemler [5], [7], [15], [18], [19], [24] ile karşılaştırılmıştır. Tablo 4 ve Tablo 5 karşılaştırıldığında, GNS' nin GA kullanılarak belirlenen optimum değeri kullanılarak, sınıflandırma performansının iyileştirildiği görülmüştür. Ayrıca DÇF-UÖM' nin q, t, j parametreleri GA kullanılarak optimize edilmiştir. Böylelikle, GA-DÇF-UÖM' nin sınıflandırma performansı artırılmıştır. Tablo 6' da GA-DÇF-UÖM tabanlı sistemin sınıflandırma doğruluğu önceki yöntemler [5], [7], [15], [18], [19], [24] ile karşılaştırılmıştır.

**Tablo 6.** GA-DÇF-UÖM tabanlı sistemin doğru tanıma performansının önceki yöntemlerle karşılaştırılması

<u>Yöntem</u>	<u>Doğruluk %</u>	<u>Kaynak</u>
Logdisc	77.7	Statlog
IncNet	77.6	Norbert Jankowski
DIPOL92	77.6	Statlog
Linear Discr. Anal.	77.5-77.2	Statlog; Ster & Dobnikar
SMART	76.8	Statlog
GTO DT (5xCV)	76.8	Bennet and Blue
kNN, k=23, Manh, raw, W	76.7±4.0	WD-GM, feature weighting 3CV
kNN, k=1:25, Manh, raw	76.6±3.4	WD-GM, most cases k=23
ASI	76.6	Ster & Dobnikar
Fisher discr. analysis	76.5	Ster & Dobnikar
MLP+BP	76.4	Ster & Dobnikar
MLP+BP	75.8± 6.2	Zarndt
LVQ	75.8	Ster & Dobnikar
RBF	75.7	Statlog
NB	75.5-73.8	Ster & Dobnikar; Statlog
kNN, k=22, Manh	75.5	Karol Grudziński
MML	75.5± 6.3	Zarndt
SNB	75.4	Ster & Dobnikar
BP	75.2	Statlog
kNN, k=18, Euclid, raw	74.8± 4.8	WD-GM
CART DT	74.7± 5.4	Zarndt
CART DT	74.5	Statlog
DB-CART	74.4	Shang & Breiman
ASR	74.3	Ster & Dobnikar
Bayes	72.2± 6.9	Zarndt
C4.5 (5xCV)	72.0	Bennet and Blue
CART	72.8	Ster & Dobnikar
Kohonen	72.7	Statlog
kNN	71.9	Ster & Dobnikar
C4.5 rules	67.0± 2.9	Zarndt
QDA	59.5	Ster, Dobnikar
BFGS quasi Newton	77.08	(Yildirim et al. 2003)
Levenberg-Marquardt	77.08	(Yildirim et al. 2003)
RBF	68.23	(Yildirim et al. 2003)
GRNN	80.21	(Yildirim et al. 2003)
GDA-LS-SVM	79.16	(K. Polat et al., 2008)
Gradient descent	77.60	(Yildirim et al. 2003)
SVM (5xCV)	77.6	Bennet and Blue
C4.5	76.0±0.9	Friedman, 5xCV
Semi-Naive Bayes	76.0±0.8	Friedman, 5xCV
Naive Bayes	74.5±0.9	Friedman, 5xCV
GA-DÇF-UÖM	98.31	Bu çalışmada

Önerilen GA-DÇF-UÖM' nin geçerliliğini göstermek için önceki çalışmalarla karşılaştırılan sonuçlar da Tablo 4' de gösterilmektedir. Tablo 6' da gösterildiği gibi, diyabet için en iyi performans %98,31 ile önerilen GA-DÇF-UÖM kullanılarak elde edilmiştir. Çünkü UÖM' nin GNS' sının optimum değerleri ve DÇF' nun q, t, j parametreleri GA kullanılarak elde edilmiştir.

GA-DÇF-UÖM' ye dayalı olarak önerilen tanı sisteminin etkinliğini göstermek için duyarlılık ve özgüllük analizi yapılır. Diyabet için GA-DÇF-UÖM yöntemi kullanılarak elde edilen duyarlılık ve özgüllük değerleri Tablo 7' de gösterilmiştir. Tablo 8' de, elde edilen duyarlılık ve özgüllük değerlerinin bu alanda yapılan diğer çalışmalarla karşılaştırılması gösterilmiştir.

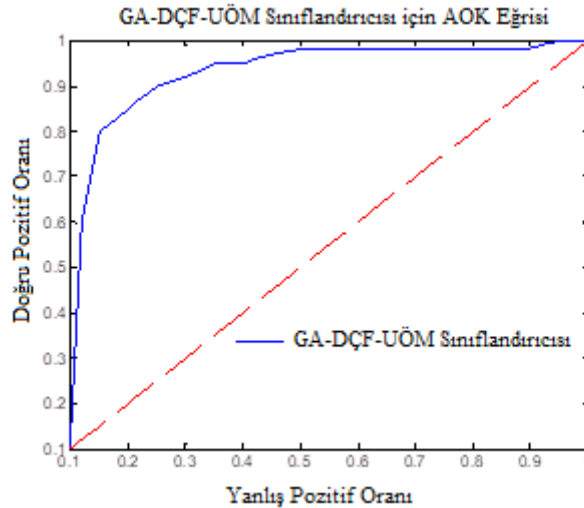
**Tablo 7.** Diyabet için GA-DÇF-UÖM yöntemi kullanılarak elde edilen duyarlılık ve özgüllük değerleri

Yöntem	Sınıflandırma Doğruluğu (%)
Duyarlılık Analizi	97.72
Özgüllük Analizi	98.9
<b>Ortalama</b>	<b>98.31</b>

**Tablo 8.** Duyarlılık ve özgüllük değerlerinin bu alanda yapılan diğer çalışmalarla karşılaştırılması

	Duyarlılık (%)	Özgüllük (%)
LS-SVM (K. Polat ve arkadaşları, 2008)	73.91	80
GDA-LS-SVM (K. Polat ve arkadaşları, 2008)	79.16	83.33
GA-DÇF-UÖM (Bu çalışmada kullanılan yöntem)	97.72	98.9

Bu çalışmada TP, TN, FP ve FN kullanılarak AOK eğrileri ve AUC değerleri bulunmuştur. TP, gerçek pozitifleri temsil eder. TN, gerçek negatifleri temsil eder. FP, yanlış pozitifleri temsil eder. FN, yanlış negatifleri temsil eder. GA-DÇF-UÖM'nin AOK eğrisi Şekil 5'te gösterilmiştir. AOK eğrilerinin AUC (AOK Eğrisi Altında Kalan Alan) değeri, diyabet verileri için GA-DÇF-UÖM kullanılarak 0,983 olarak hesaplanmıştır.



**Şekil 4.** Diyabet teşhisi için önerilen GA-DÇF-UÖM yönteminin AOK eğrisi.

## 8. Tartışma ve Sonuç

Bu çalışmada, GA-DÇF-UÖM diyabet teşhisi için akıllı bir teşhis sistemi önerilmektedir. Özellik vektörünün doğrudan kullanımı, hızlı eğitim ve test süresi ve geleneksel sinir ağları üzerinden genelleme yapabilme yeteneği, GA-DÇF-UÖM diyabet tanı sisteminin avantajlarıdır. Diyabet veri tabanından alınan özellik vektörü, DÇF-UÖM

sınıflandırıcılarına girdi olarak verilir. DÇF-UÖM yapısı, DÇF' nin üç adet düzenlenebilir parametresine sahiptir. DÇF-UÖM sınıflandırıcısının içerisindeki UÖM' nin GNS değerleri, q, t, j parametreleri UÖM' nin performansında önemli bir rol oynamaktadır. Bu nedenle bu parametrelerin ve GNS değerlerinin dikkatli bir şekilde ayarlanması gerekir. Bu çalışmada, GA ile DÇF-UÖM' deki UÖM' nin GNS değerlerinin ve q, t, j parametrelerinin optimum değerleri elde edilmiştir. GA kullanılarak elde edilen optimal değerler, DÇF-UÖM sınıflandırıcısına verilmiştir. GA-DÇF-UÖM tabanlı akıllı diyabet tanı sistemi, olası diyabet hastalarının iyi düzeyde teşhis edilmesini sağlamaktadır. Bu tatmin edici sonuçlar sayesinde doktorlar, hastalar hakkında nihai karara varabileceklerdir. Böylece, doktorlar diyabet teşhisi konusunda daha doğru karar verebilmek için önerilen yöntemi kullanabileceklerdir.

### Kaynaklar

- [1] American Diabetes Association. (2014). Diagnosis and classification of diabetes mellitus. *Diabetes care*, 37(Supplement 1), S81-S90.
- [2] American Diabetes Association. (2014). Standards of medical care in diabetes—2014. *Diabetes care*, 37(Supplement 1), S14-S80.
- [3] Johnson, R. J., Nakagawa, T., Sanchez-Lozada, L. G., Shafiu, M., Sundaram, S., Le, M., ... & Lanaspas, M. A. (2013). Sugar, uric acid, and the etiology of diabetes and obesity. *Diabetes*, 62(10), 3307-3315.
- [4] K. Polat, S. Gunes and A. Arslan, (2008). A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine, *Expert Systems with Applications* 34, p.p.482-487.
- [5] K., Kayaer, & T. Yıldırım, (2003). Medical diagnosis on pima indian diabetes using general regression neural networks, artificial neural networks and neural information processing (ICANN/ICONIP) (pp. 181-184), Istanbul, Turkey, June 26-29.
- [6] K.P. Bennett, J. Blue, A Support Vector Machine Approach to Decision Trees, R.P.I Math Report No. 97-100, Rensselaer Polytechnic Institute, Troy, NY, 1997.
- [7] Friedman N, Geiger D, Goldszmit M (1997). Bayesian networks classifiers. *Machine Learning* 29: p.p.131-163.
- [8] Avci, D. (2016). An Automatic Diagnosis System for Hepatitis Diseases Based on Genetic Wavelet Kernel Extreme Learning Machine. *Journal of Electrical Engineering & Technology*, 11(4), 993-1002.
- [9] G. B. Huang, Q.-Y. Zhu and C.-K. Siew, "Extreme Learning Machine: Theory and Applications", *Neurocomputing*, vol. 70, pp. 489-501, 2006.
- [10] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A Fast and Accurate On-line Sequential Learning Algorithm for Feedforward Networks", *IEEE Transactions on Neural Networks*, 17 (2006) 1411-1423.
- [11] Al-Shayea, Qeethara Kadhim. "Artificial neural networks in medical diagnosis", *International Journal of Computer Science Issues* 8.2 (2011): 150-154.
- [12] Whitley, D. (2014). An executable model of a simple genetic algorithm. *Foundations of genetic algorithms*, 2(1519), 45-62.
- [13] Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., ... & Morris, Q. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science*, 347(6218), 1254806.
- [14] Goldberg, D. E. (2006). *Genetic algorithms*. Pearson Education India.
- [15] Bin Li, Xuewen Rong and Yibin Li, "An Improved Kernel Based Extreme Learning Machine for Robot Execution Failures", *Hindawi Publishing Corporation The Scientific World Journal*, Volume 2014, Article ID 906546, pp. 7, <http://dx.doi.org/10.1155/2014/906546>.
- [16] Peng Guan, De-Sheng Huang, Bao-Sen Zhou, Forecasting model for the incidence of hepatitis A based on artificial neural network, *China World Journal of Gastroenterol*; 10(24), 2004, pp. 3579-3582.
- [17] Polat K., Gunes S., Hepatitis disease diagnosis using a new hybrid system based on feature selection (FS) and artificial immune recognition system with fuzzy resource allocation, *Digital Signal Processing* 16 (2006), pp. 889-901.
- [18] J. Beyer, J. Schrezenmeir, G. Schulz, T. Strack, E. Küstner, G. Schulz, The influence of different generations of computer algorithms on diabetes control, *Computer Methods and Programs in Biomedicine*, Vol. 32, Issues 3-4, July-August 1990, Pages 225-232.
- [19] <http://www.phys.uni.torun.pl/kmk/projects/datasets.html> ((last accessed: April 18, 2011).
- [20] Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*, 12(1), 77.
- [21] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874.
- [22] Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, 44(1), 83-98.
- [23] Matlab 7.7.0, MATLAB Company, 2011.
- [24] Ertam, F., & Avci, E. (2017). A new approach for internet traffic classification: GA-DÇF-UÖM. *Measurement*, 95, 135-142.
- [25] Ding, S., Zhang, J., Xu, X., & Zhang, Y. (2016). A wavelet extreme learning machine. *Neural Computing and Applications*, 27(4), 1033-1040.