

Genetik Algoritmalar Kullanarak EKG Vuru Sınıflandırması için Öznitelik Seçimi

Feature Selection for ECG Beat Classification using Genetic Algorithms

Çağla Sarvan¹, Nalan Özkurt¹, Korhan Karabulut²

¹Elektrik-Elektronik Mühendisliği Bölümü, Yaşar Üniversitesi, İzmir, Türkiye
caglasarvan@gmail.com, nalan.ozkurt@yasar.edu.tr

² Yazılım Mühendisliği Bölümü, Yaşar Üniversitesi, İzmir, Türkiye
korhan.karabulut@yasar.edu.tr

Özetçe— Bu çalışmada, kalp vurularının aritmi tipine göre sınıflandırılmasında en uygun öznitelik setinin seçilmesi için genetik algoritma yöntemi kullanılmıştır. Sınıflandırmada MIT-BIH kalp aritmi veri tabanından elde edilen elektrokardiyografi (EKG) sinyallerinin normal, sağ dal bloku, sol dal bloku ve pace ritmi örnekleri kullanılmıştır. Yaygın kullanımı olan kesikli dalgacık dönüşümü (KDD) katsayılarının ortalaması, standart sapması, enerjisi ve entropisi sınıflandırmada kullanılabilecek öznitelik ana kümesi olarak önerilmiştir. Önerilen KDD yöntemi ile 13 farklı dalgacık tipi için elde edilen 208 adetlik öznitelik seti içerisinde yüksek sınıflandırma doğruluğu sağlayan 16 adetlik öznitelik seti genetik algoritma yöntemi uygulanarak elde edilmiştir. Genetik algoritma yöntemi ile sınıflandırma doğruluğunu yükselten özniteliklerin tespit edilebildiği ve farklı tip dalgacıkların farklı seviyelerde seçilen katsayılarından elde edilen öznitelik setinin standart olan tek-tip dalgacıktan elde edilen katsayılara göre EKG aritmi sınıflandırmasında daha yüksek başarımla gösterdiği gözlemlenmiştir.

Anahtar Kelimeler— EKG vuru sınıflandırması, aritmi, kesikli dalgacık dönüşümü, dalgacık öznitelikleri, öznitelik seçimi, yapay sinir ağları, genetik algoritmalar.

Abstract— In this study, genetic algorithm method was used to select the most suitable set of features for classification of arrhythmia types of heart beats. Normal, right branch block, left branch block and pace rhythm samples of electrocardiography (ECG) signals which obtained from the MIT-BIH cardiac arrhythmia database were used in the classification. Mean, standard deviation, energy and entropy of discrete wavelet transform (DWT) coefficients were proposed as the features for the classification. By using the proposed DWT method, 16 features which have high classification accuracy were obtained among the 208 feature sets constructed from 13 different wavelet types by applying the genetic algorithm method. It was observed that the features that increase accuracy can be detected by the genetic

algorithm and the feature set obtained from the coefficients of the different types of wavelets selected at different levels show higher performance than the coefficients obtained from the standard individual wavelet in the ECG arrhythmia classification.

Keywords— ECG heart beat classification, arrhythmia, discrete wavelet transform, wavelet features, feature selection, neural network, genetic algorithm.

I. GİRİŞ

Günümüzde birçok probleme çözüm bulacak otomatik veya yarı-otomatik sistemlerin geliştirilmesi daha doğru sonuçlara varılmasında önem taşımaktadır. Bu bağlamda biyomedikal uygulamalarda uzmanlara yardımcı olmak açısından başarımla yüksek sistemler geliştirilmektedir. Bu sistemlerden elde edilen sonuçların değerlendirilmesi hastalık teşhisinin doğru konulabilmesinde önemlidir. Uzmanlarca, değerlendirmelerinde kullanılan sistemlerden biri olan elektrokardiyografi (EKG) kalbin çalışması esnasında meydana gelen elektriksel faaliyetin takip edilebilmesi ve kayıt altına alınabilmesini sağlar. EKG cihazlarından elde edilen kalp sinyalleri, kalbin çalışması sırasında oluşan aritmi bozuklukları ve bunlardan kaynaklanabilecek kalp rahatsızlıkları hakkında önemli bilgiler taşımaktadır. EKG kayıtları incelenerek kardiyak aritmilerinin, başarımla yüksek olarak otomatik veya yarı-otomatik yöntemlerle ayrıştırılabilmesi kardiyak döngü bozukluklarının tespiti ve teşhisi konusunda uzmanlara yardımcı olmaktadır.

Kardiyak döngüsü sırasında oluşan elektriksel aktivite, P, Q, R, S ve T olarak adlandırılan beş farklı dalgalanma yapıyla karakterize edilir. Bu aktivite incelenerek kardiyak döngüsünde oluşan aritmi bozuklukları belirlenebilmektedir. Farklı tip aritmilerin birbirinden

ayrıştırılması QRS kompleksinden elde edilen bilgiler ile yapılabilmektedir. Bu ayrıştırıcı bilgilere sinyallerden öznitelik seti çıkarımı ile ulaşılabilmektedir. Sınıflandırmanın başarısı açısından sinyalden elde edilen öznitelik setinin uygunluğu önem taşımaktadır.

EKG sınıflandırma için farklı türden öznitelik çıkarımı yöntemleri önerilmiştir. Öznitelik seti olarak morfolojik yöntemlerden [1-4], Fourier dönüşümü kullanılan yöntemlerden [5-7] ve dalgacık dönüşümü kullanılan yöntemlerden [8-10] elde edilen öznitelikler yaygın olarak kullanılmaktadır. Morfolojik yöntemlerde EKG vurusunda belirgin olan P, Q, R, S ve T elektriksel noktalarının arasında geçen süreler ile sinyalin morfolojik yapısı matematiksel olarak elde edilerek sınıflandırmada kullanılacak olan öznitelik seti oluşturulur[1-4]. Fourier dönüşümü ile elde edilen frekans bileşenlerinin spektrumundan zaman alanı gibi maksimum nokta, minimum nokta, ortalama, varyans ve standart sapma gibi istatistiksel öznitelikler çıkartılır [11]. Literatürde yaygın kullanımı olan dalgacık dönüşümünde ise dalgacık tipine bağlı olarak elde edilen detay katsayılarından standart sapma, ortalama, enerji ve entropi gibi öznitelikler elde edilmektedir.

Dalgacık dönüşümü (DD), yüksek frekansları değerlendirmek için kısa bir zaman aralığı ve düşük frekansları değerlendirmek için uzun bir zaman aralığı kullanır. Bu özellik sayesinde kısa zaman aralığında yüksek frekans bileşenleri DD uygulanarak başarıyla gözlemlenebilmektedir. Dalgacık dönüşümünün avantajlarından bir tanesi de, sinyalleri çeşitli çözünürlüklerde ayrıştırabilmesidir; bu da, EKG gibi durağan olmayan sinyallerden doğru öznitelik çıkarılmasına olanak tanır. Her dalgacık dönüşümü, sinyalin zamansal genişliği ve sinyalin frekans spektrumuyla ilgili bilgi verir. Bu noktada sinyal şekline uygun dalgacık fonksiyonunun belirlenmesi, elde edilen özniteliklerde daha başarılı sonuçlar verebilmektedir. Belirlenen dalgacık fonksiyonu kullanılarak sinyal seviyelere ayrıştırılarak sinyalde lokal analiz yapılır. Bu seviyelerden elde edilen detay katsayılarının bilgisi öznitelik çıkarımında kullanılmaktadır.

Farklı yöntemler ile elde edilen öznitelik setleri Fatemeh Molaei Vaneghi ve arkadaşlarının yapmış olduğu çalışmada da yer aldığı gibi farklı sınıflandırma başarımı gösterebilmektedir [12]. Bu nedenle sinyale uygun öznitelik setinin seçimi başarımı etkileyen önemli unsurlardan biridir. Öznitelik seti seçiminde kullanılan yaygın yöntemlerden biride genetik algoritmalar [13-17]. Genetik ve evrimsel algoritmalar birçok problemin optimizasyonunda doğru veya kabul edilebilir bir çözüm elde etmek açısından yaygın olarak kullanılmaktadır. Özellikle yüksek boyutlu arama alanlarında etkili oldukları için çok boyutlu problem

optimizasyonuna yaklaşımlarından kaynaklı öznitelik seçimlerinde yaygın olarak kullanılmaktadır.

Bu çalışmada Physiobank MIT-BIH Aritmi veri tabanından alınan EKG sinyalleri kullanılmıştır [18-20]. Sınıflandırılacak aritmi tipleri normal, sağ dal bloku, sol dal bloku ve pace ritmi olarak belirlenmiştir. Öznitelik ana kümesinin oluşturulmasında dalgacık dönüşümü yöntemi kullanılmıştır. 13 farklı tip dalgacıktan elde edilen 4 seviyedeki detay katsayılarının ortalama, standart sapma, enerji ve entropi değerleri kullanılarak 208 adetlik öznitelik havuzu oluşturulmuştur. GA kullanılarak bu havuz içerisinde 16 adetlik, yapay sinir ağı testlerinde sınıflandırma başarımını yükselten öznitelik alt kümesi elde edilmiştir.

II. DALGACIK ANALİZİ

Sürekli dalgacık dönüşümü (SDD) sinyal hakkında yüksek çözünürlüklü zaman ve frekans bilgisi elde etmek için geliştirilmiş bir yöntemdir. Sürekli Dalgacık Dönüşümünde, Kısa Süreli Fourier Dönüşümü yerine değişken boyutlu pencere kullanılmaktadır. Sıkıştırılmış ve genişletilmiş dalgacıklar ile sağlanan çoklu çözünürlük özelliği sayesinde sinyale ait farklı özellikler elde edilebilmektedir [21].

SDD, zaman-frekans atomları $\Psi(t)$ ile $s(t)$ sinyalinin ilişkilendirilmesi ile hesaplanarak, $C(a,b)$ ile ifade edilen dalgacık katsayılarını üretir:

$$C(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \Psi^* \left(\frac{t-b}{a} \right) dt \quad (1)$$

b kaydırma, a ise ölçek katsayısıdır ve frekans ile ters orantılıdır ($a>0$), “*” karmaşık sayının eşleniğini ve $\Psi(t)$ ana dalgacık fonksiyonudur.

Kesikli Dalgacık Dönüşümü (KDD) sıkıştırma ve kaydırma aralıklarını kesikli hale getirir. Sayısal filtreleme teknikleri kullanılarak hızlı algoritmalar ile dalgacık katsayıları hesaplanabilir. KDD şu şekilde elde edilebilir

$$C_{mn} = \int_{-\infty}^{+\infty} s(t) \Psi_{mn}(t) dt \quad (2)$$

İşaret analizinde sıklıkla kullanılan dyadic skalada sıkıştırma katsayısı 2, kaydırma katsayısı 1 alınır ve dalgacıklar şöyle elde edilir

$$\Psi_{mn}(t) = 2^{-m/2} \Psi(2^{-m}t - n) \quad (3)$$

Ortogonalite özelliklerini sağlayan Daubechies (DB), Symlets (SYM), Coiflets (COIF) ve Bior (BIOR) biyomedikal sinyal işlemede yaygın kullanımı olan dalgacık aileleridir [21].

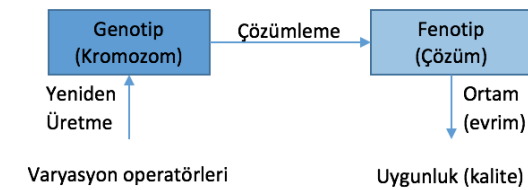
III. GENETİK ALGORİTMALAR

Genetik Algoritmalar (GA) evrimsel algoritmaların alt sınıfı olup analogisini Darwin'in doğal seleksiyonundan temel alır. GA birçok probleme uyarlanabilir optimum veya kabul edilebilir çözümü arama metodudur. GA' da genel olarak iki çözüm üreten çaprazlama yöntemleri ön plana çıkarken, çözümler arasındaki çeşitliliğin artmasını bireyin içeriğini rastgele şekilde değiştiren mutasyon yöntemleri sağlamaktadır.

GA tasarımında kullanılan bazı genel yapılar bulunmaktadır. Bu yapılarda kendi içerisinde özelleştirilebilmektedir.

A. Gösterim

GA' da kodlanmış çözüme kromozom adı verilirken, bir çözüm içindeki karar değişkenleri genlerdir. Şekil 1'de gösterildiği üzere genotip kodlamayı temsil ederken, fenotip olası çözümü temsil eder. Dolayısıyla, fenotipi üretmek için genotip kodunun çözülmesi gerekir. Varyasyon operatörleri, genotip seviyesinde hareket ederken, uygunluk fonksiyonu ilgili kişinin fenotipini kullanır. Doğrudan kodlamanın kullanıldığı durumda, genotip fenotipe benzer fakat dolaylı bir kodlama kullanılırsa, genotip ve fenotip farklı yapılara sahip olurlar [22].



Şekil 1. Genotipe karşılık gelen Fenotip

B. Popülasyon

Popülasyon olası çözümlerin yer aldığı alt kümeyi ifade etmektedir. Başlangıç popülasyonu iki şekilde oluşturulabilir. Tamamen rasgele oluşturulmuş çözümlerden rasgele başlangıç şeklinde veya problemin bilinen bir çözümüyle bulgusal başlangıç şeklinde oluşturulabilmektedir [22].

C. Uygunluk Gösterimi

GA' da uygunluk fonksiyonu, tanımlanan probleme yönelik aday bir çözümü girdi olarak alıp çıkış üreterek

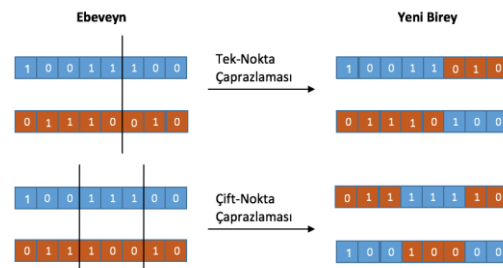
çözümün ne kadar iyi olduğunu, probleme ne kadar uyumlu olduğunu göstererek oluşturulan fonksiyon yapısıdır. Uygunluk değerinin hesaplanması GA' da art arda yapılır ve bu nedenle hesaplamada hızlı olmalıdır. Probleme en uygun çözümün üretilmesi için çözümün ne kadar uygun olduğunun veya verilen çözümden kişilerin nasıl üretilebileceğinin nicel olarak ölçülmesi gereklidir.

D. Seçim Yöntemleri

Genetik algoritmalarda sonraki nesillere kopyalanacak bireylerin seçilmesinde kullanılan birçok farklı yöntem bulunmaktadır. Seçkin seçim, rulet-çarkı seçilimi, turnuva seçilimi ve orantılı seçim kullanılan yöntemlerden bazılarıdır. Bu yöntemlerden bir kısmı kombinasyon halinde kullanılabilir. Seçkin seçimde her nesilden en uygun olan bireylerin seçilmesi garanti edilirken her bir neslin en iyi veya en iyi birkaç bireyi, gelecekte daha iyi bir birey ortaya çıkmaması durumunda yeni nesillere kopyalanır. Orantılı seçimde diğerlerine göre daha uygun bireylerin seçilmesi olasıdır, fakat kesin değildir. Rulet-çarkı seçilimi orantılı seçilimin bir formu olup, yeni nesle seçilecek olan bireyin uygunluğunun rakiplerinden daha yüksek veya düşük olması baz alınarak belirlenir. Turnuva seçilimin de ise bireylerin alt grupları daha büyük olan popülasyonlardan oluşturulur ve tüm alt grup bireyleri birbirleriyle yarışır. Her alt gruptan sadece bir birey yeni popülasyonu oluşturmak üzere çaprazlama işlemi için aday olarak seçilir [22].

E. Çaprazlama Yöntemleri

Çaprazlamanın yaygın biçimleri arasında, tek-nokta, n-nokta ve bir-örnek çaprazlaması bulunmaktadır. Tek-nokta çaprazlamada, Şekil 2' de gösterildiği üzere çaprazlama noktası olan k değeri rastgele seçilir ve ebeveynlerin kesimlerinin değiştirilmesi ile iki kök oluşturulur, ebeveynlerin bu bölümlerini değiş tokuş ederek yeni iki birey yaratılır. Aynı şekilde birden fazla k noktasının rasgele belirlenmesiyle yeni iki bireyin üretilmesi n-nokta çaprazlaması olarak adlandırılmaktadır [22].



Şekil 2. Tek-Nokta ve Çift-Nokta Çaprazlaması

Bir-örnek çaprazlamasında kullanan iki ebeveyn, kromozomlarının boyutlarına göre yeniden birleştirilebilir. Kromozomun her elemanı rasgele seçilir. Şekil 3’ te belirtildiği üzere her ebeveyn, yeni bireyleri üretmek için eşit olarak katkıda bulunur [22].



Şekil 3. Bir-Örnek Çaprazlaması

F. Mutasyon Yöntemleri

Seçilim yöntemleri kullanılarak birey seçimi yapıldıktan sonra yeni nesle aktarılacak olan bireyin uygunluğunun artması umuduyla rasgele değişimler uygulanmaktadır. Bunlardan en basiti mutasyon olarak adlandırılır. Tıpkı canlılardaki mutasyonların bir geni değiştirdiği gibi, genetik bir algoritmada mutasyon, bireyin kodundaki tek noktalarda küçük değişikliğe neden olması olarak ifade edilir.

Mutasyon yöntemlerinden sık kullanımı olanlar bit çevirme, rasgele sıfırlama, takas mutasyonudur. Rasgele sıfırlama, rastgele seçilen bir gene izin verilen değerler grubundan rastgele bir değer atanması ile yapılır. Takas mutasyonunda ise rasgele olarak kromozom üzerinde iki konum seçer ve değerlerin yerini değiştiririz.

İkili sistem gösterimlerinde bit çevirme mutasyonu sık kullanımı olan mutasyon yöntemi olup çevirme operatörü olarak tanımlanır [22].

IV. YÖNTEM

Bu çalışmada genetik algoritma tabanlı dalgacık dönüşümünden elde edilen öznitelik ana kümesinden alt küme seçimi uygulanmıştır. Dalgacık dönüşümü kullanılarak elde edilen 4 seviyedeki detay katsayılarından ayrı ayrı olacak şekilde hesaplanan ortalama, standart sapma, enerji ve entropi değerleri genetik algoritmada öznitelik uzay seti olarak kullanılmıştır. GA sonuçları, aşağıdaki parametre ayarlarıyla elde edilmiş seçilime dayanmaktadır:

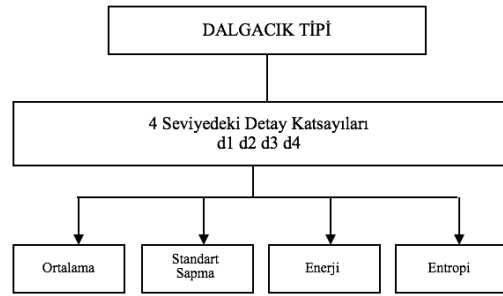
- **Çaprazlama Olasılığı:** 0.7
- **Mutasyon Olasılığı:** 0.05
- **Mutasyon Yöntemi:** Rasgele seçimle iki gen değiştirme
- **Ebeveyn Seçilimi:** Turnuva Seçilimi
- **Seçkinlik:** Var

- **Çaprazlama Yöntemleri:** Tek-nokta, Çift-nokta ve Bir-örnek
- **Çaprazlama Yöntem Seçilimi:** Rulet-Çarkı Seçilimi

Önerilen GA tabanlı öznitelik seti seçimi ve ikili sistem gösterimi, popülasyon oluşumu, uygunluk fonksiyonu ve sistem mimarisi aşağıda alt başlıklar halinde tanımlanmaktadır.

A. İkili Sistem Gösterimi

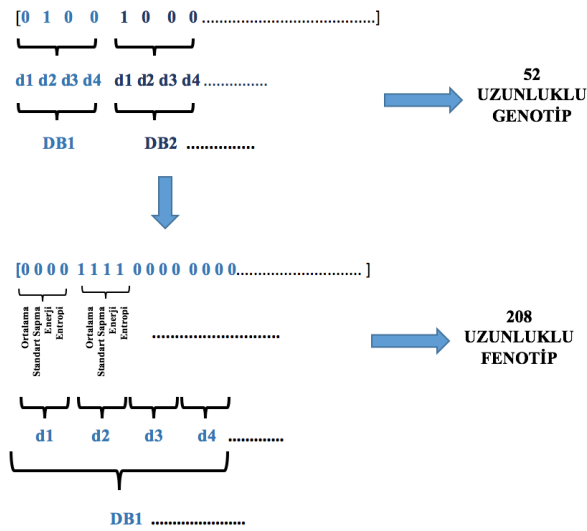
Bu çalışmada öznitelik ana kümesi, Şekil 4’de görüldüğü üzere 13 adet farklı dalgacık tipinden dalgacık ayrıştırması yapılarak 4 seviyede elde edilen detay katsayılarından ortalama, standart sapma, enerji ve entropi değerleri hesaplaması yapılarak oluşturulmuştur.



Şekil 4. Öznitelik Ana Küme Oluşturma Yöntemi

Bireylerin genotip ve fenotip kodlamaları 1 ve 0’lardan oluşan ikili sistem dizilimi ile gösterilmektedir. Şekil 5’te görüldüğü üzere genotip içinde ilgili dalgacığın hangi seviyesinin seçildiği “1” ile gösterilirken, diğer seviyeler “0” ile ifade edilir. Fenotipte ise her seviye için hangi özniteliklerin seçildiği gösterilir.

Bu çalışmada 52 uzunluklu genotipten, 208 uzunluklu fenotip elde edilmiştir. Fenotip uygunluk fonksiyonu hesaplamalarında kullanılmıştır.



Şekil 5. Genotip ve Fenotip Yapısı

B. Popülasyon Oluşumu

Rasgele oluşturulan başlangıç popülasyonunda 2 seviyeli turnuva seçilimi uygulanmıştır. Popülasyondan rasgele seçilen birey sayısı kendi içerisinde ikiye bölünmüştür. Oluşturulan iki alt grup uygunluk fonksiyonundan dönen değerlere göre yarıştırmıştır, kazanan bireyler final havuzuna atılmıştır. Final havuzundan alınan çiftlerden yeni birey oluşturmak için çaprazlama uygulanmıştır. Çaprazlama yöntemi oranları 0.7 bir-örnek, 0.2 çift-nokta ve 0.1 tek-nokta olarak belirlenmiş olup seçiminde rulet-çarkı seçilimi kullanılmıştır. Her çaprazlamadan oluşan 2 adet birey yeni popülasyon havuzuna atılmıştır. Eski popülasyon ile yeni popülasyonda seçkinlik değişimi uygulanmıştır. Eski popülasyonun en iyi bireyinin uygunluk değeri ile yeni popülasyonda bulunan en iyi bireyin uygunluk fonksiyon değeri karşılaştırılıp yüksek olan yeni popülasyonda tutulmuştur. Böylelikle bulunan en iyi birey tüm nesiller boyunca popülasyonda tutulmuştur. Seçkinlik değişiminden sonra en iyi birey kayıp olmayacak şekilde mutasyon oranına bağlı olarak rasgele zar atımı yapılır, gelen zar değeri mutasyon oranından düşükse bireye iki gen değişimi yapılarak mutasyon uygulanır, bu değer büyükse mutasyon uygulanmaz. Bu çalışmada popülasyondaki bireyler %5'lik ihtimal ile mutasyona girebilmektedir.

C. Uygunluk Fonksiyonu

Bu çalışmada öznelilik setinin aritmleri sınıflandırmada yapay sinir ağı testleri sonucunda yüksek başarımlı oranın göstermesi beklenmektedir. Uygunluk fonksiyonunda

popülasyondaki bireylerin fenotipleri kullanılmaktadır ve 208 adetlik öznelilik ana kümesinden 16 adet öznelilik fenotipe göre seçilerek yapay sinir ağına 16x27454' lük giriş matrisi verilmektedir. Yapay sinir ağı eğitim fonksiyonu olarak Levenberg-Marquardt (trainlm) algoritması seçilmiştir [23]. Popülasyondaki bireylerin oluşturduğu 16x27454 lük matrislerden %70 eğitim, %15 test ve %15 doğrulama verisi yapay sinir ağına rasgele dizilimle seçilmiş olup sınıflandırma algoritmasında 10 saklı katmanda eğitilip test edilmiştir. Yapay sinir ağından elde edilen test ve doğrulama verilerinin doğruluk oranı (4) numaralı denklem kullanılarak hesaplanmıştır.

Doğruluk A yüzde oranı gerçek pozitif TP ve gerçek negatif TN toplamının, gerçek pozitif TP, yanlış pozitif FP, gerçek negatif TN ve yanlış negatif FN toplamına oranı ile belirlenir. Doğruluk A yüzde oranı şu şekilde ifade edilebilir

$$A = \frac{TP + TN}{TP + FP + TN + FN} \cdot 100\% \quad (4)$$

GA' de kullanılan uygunluk değeri yapay sinir ağına normal, sol dal bloku, pace ritmi ve sağ dal bloku sinyallerinden elde edilen test ve doğrulama verilerinin ayrı ayrı doğruluk oranlarının, sırasıyla ANormal, ALBBB, APaced ve ARBBB , toplanıp ortalamasının alınması ile elde edilmiştir. Uygunluk şu şekilde ifade edilebilir

$$Uygunluk = \frac{A_{Normal} + A_{LBBB} + A_{Paced} + A_{RBBB}}{4} \quad (5)$$

Hesaplanan bu değer popülasyondaki bireylerin uygunluk değeri olarak alınmış olup, genetik algoritmada maksimize edilmeye çalışılmaktadır.

D. GA Tabanlı Sistem Mimarisi

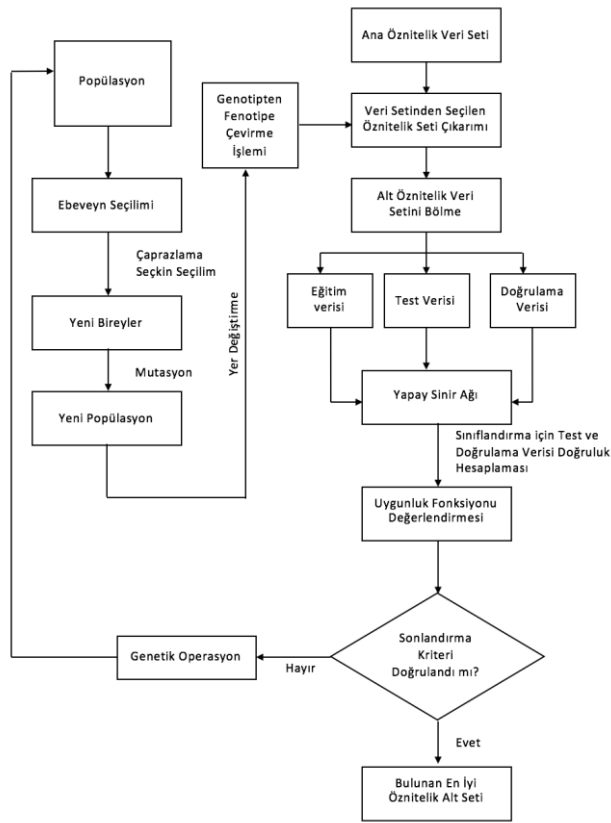
Bu çalışmada uygulanan GA sistem mimarisi Şekil 6'da gösterilmektedir. Sistem mimarisinden yola çıkılarak çalışma yapısı aşağıda belirtilen adımlarda açıklanmaktadır.

- (1) *Ana Öznelilik Veri Seti.* Bu aşamada MIT-BIH veri tabanından alınan 4 farklı aritmi tipi örneklerinden dalgacık dönüşümü kullanılarak 208x27454'lük ana öznelilik veri seti elde edildi.
- (2) *Genotipten Fenotipe Çevirme İşlemi.* Bu adım, her parametreyi ve özellik kromozomunu genotipinden bir fenotipe dönüştürür.
- (3) *Veri Setinden Seçilen Öznelilik Seti Çıkarımı.* Genotipten fenotipe çevirme işleminden sonra ana öznelilik setinden seçilen 16 adet öznelilik ayrıştırılır.

(4) *Uygunluk Fonksiyonu Değerlendirmesi*. Alt öznelik veri seti %70 eğitim, %15 test ve %15 doğrulama verisi olarak yapay sinir ağına kullanılmak üzere bölünür. Yapay sinir ağı eğitimi sınıflandırma doğrulama hesaplamak için test ve doğrulama veri setini kullanılır. Sınıflandırma doğruluğu elde edildiğinde her birey uygunluk fonksiyonu ile değerlendirilir.

(5) *Sonlandırma Kriteri Doğrulandı mı?* Sonlandırma kriteri doğrulandıysa bulunan en iyi öznelik setine geri dönülür aksi halde genetik operasyon çalışmaya bir sonraki nesilden devam eder. Bu çalışmada sonlandırma kriteri atanan nesil sayısı olarak belirlenmiştir. Verilen nesil sayısı tamamlandığında sonlandırma kriteri doğrulanmış olur.

(6) *Genetik Operasyon*. Bu adımda, sistem, seçim, çaprazlama, mutasyon ve yer değiştirme dahil olmak üzere genetik işlemlerle daha iyi çözümler aramaya devam eder.

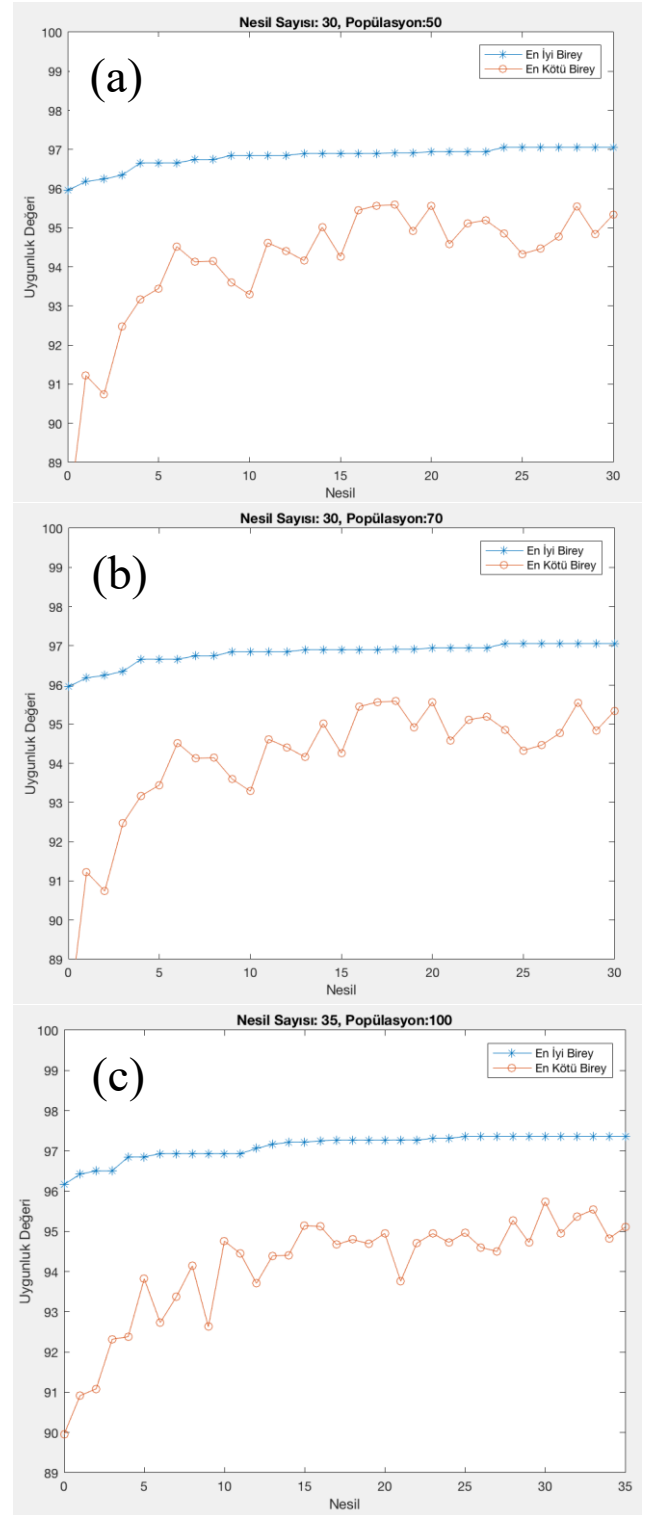


Şekil 6. GA Sistem Mimarisi

V. SONUÇLAR VE TARTIŞMA

Çalışmada MIT-BIH Aritmi veri tabanından alınan örneklerden elde edilen 7198 adet normal, 6902 adet sol dal bloku, 6332 adet sağ dal bloku ve 7022 adet pace ritmi verisi kullanılmıştır. Toplamda 27454 örnek kullanılarak 13 adet aday dalgacık için 4 seviyedeki dalgacık katsayılarının ortalama, standart sapma, enerji ve entropisinden oluşan 16x27454 öznelik setleri

birleştirilerek 208x27454'lük ana öznelik seçim kümesi elde edilmiştir.



Şekil 7. Nesillere göre En İyi ve En Kötü Birey

GA yöntemi kullanılarak 208 adetlik 27454 örnekten 4 seviyedeki detay katsayılarının seçimi yapılarak seçilen katsayıların ortalama, standart sapma, enerji ve entropi değerlerinden oluşan 16 adetlik alt öznitelik seti elde edilmiştir.

Nesil ve popülasyon sayılarında değişiklik yapılarak tespit edilen en iyi ve en kötü bireylerin uygunluk fonksiyonu karşılaştırması Şekil 7’de gösterilmektedir. Şekil 7 (a)-(c) incelendiğinde nesiller boyunca üretilen yeni bireylerin uygunluğu artmış olup mutasyon sayesinde lokal optimumda takılan bireylerin optimum sonuca yaklaştığı gözlemlenmiştir. Şekil 7 (c)’ de görüldüğü üzere 6. ile 11., 16. ile 22. nesil arasında lokal optimumda kalan birey, sırasıyla 11. ve 22. nesilde mutasyona uğrayıp lokal optimumdan kurtulmuştur.

Tablo I ’de 3 farklı nesil ve popülasyon denemesi sonucunda elde edilen en iyi ve en kötü öznitelik seti karşılaştırmaları verilmiştir. COIF2 ve COIF5 dalgacık tiplerinin içerisinde yer aldığı öznitelik alt kümesi düşük başarımlı gösterirken, DB2, BIOR1.5, BIOR2.6 ve BIOR3.5 dalgacık katsayılarından oluşan alt öznitelik seti hem 30 nesil sayısı 70 popülasyonda hem de 35 nesil sayısı 100 popülasyonda bulunan en başarılı öznitelik seti olarak tespit edilmiştir.

GA Elde Edilen Sonuçlar	GA Sonuçları		
	Nesil Sayısı:30 Popülasyon:50	Nesil Sayısı:30 Popülasyon:70	Nesil Sayısı:35 Popülasyon:100
En İyi Sonuç için Uygunluk Değeri %	97,06	97,32	97,36
Bulunan En İyi Öznitelik Seti	DB2 (d1) + COIF2(d4) + BIOR1.5 (d2) + BIOR2.6 (d4)	DB2 (d1) + BIOR1.5 (d1) + BIOR2.6 (d4) + BIOR3.5 (d4)	DB2 (d1) + BIOR1.5 (d1) + BIOR2.6 (d4) + BIOR3.5 (d4)
En Kötü Sonuç için Uygunluk Değeri %	87,39	89,05	89,95
Bulunan En Kötü Öznitelik Seti	DB4 (d2) + SYM4(d1) + COIF5(d1+d2)	DB2 (d2) + COIF2 (d1) + COIF5(d2) + BIOR5.5 (d2)	SYM4 (d2) + COIF2 (d1) + BIOR3.5 (d3) + BIOR5.5 (d2)
Seed Değeri	1408513609	1406764815	1411520240

Tablo I. GA Sonuçları

GA tarafından bulunan en iyi sonuç, dalgacık tiplerinden elde edilen standart öznitelik setlerinden elde edilen test sonuçlarıyla karşılaştırılmıştır. Karşılaştırma Tablo II de gösterilmektedir.

Dalgacık 16 lı Öznitelik Setleri	Yapay Sinir Ağı Doğruluk Hesaplama Yüzdeleri	
	Eğitim Verisi Doğruluk Oranı %	Test ve Doğrulama Verisi Doğruluk Oranı %
DB2 (d1) + BIOR1.5 (d1) + BIOR2.6 (d4) + BIOR3.5 (d4)	97,39	97,36
DB1 (d1+d2+d3+d4)	94,77	94,28
DB2 (d1+d2+d3+d4)	94,64	94,48
DB4 (d1+d2+d3+d4)	95,09	95,09
DB6 (d1+d2+d3+d4)	93,97	93,75
DB10 (d1+d2+d3+d4)	94,06	93,9
SYM4 (d1+d2+d3+d4)	94,98	94,74
SYM6 (d1+d2+d3+d4)	94,88	94,5
COIF2 (d1+d2+d3+d4)	94,12	93,75
COIF5 (d1+d2+d3+d4)	93,15	93,15
BIOR1.5 (d1+d2+d3+d4)	93,45	93,6
BIOR2.6 (d1+d2+d3+d4)	94,47	94,52
BIOR3.5 (d1+d2+d3+d4)	92,86	92,63
BIOR5.5 (d1+d2+d3+d4)	94,35	93,76

Tablo II. Yapay Sinir Ağı Doğruluk Hesaplama Yüzdeleri

13 adet farklı dalgacık tipinden, her bir dalgacıktan ayrı ayrı 4 seviyede katsayılar elde edilerek bu katsayıların ortalama, standart sapma, enerji ve entropi değerleri elde edilerek 16x27454 lük öznitelik setleri oluşturulmuştur. Bu setler karşılaştırmanın birebir olabilmesi adına GA’da kullanılan uygunluk fonksiyonundaki yapay sinir ağı parametreleri baz alınarak değerlendirilmiştir. 13 adet dalgacık tipi için eğitim, test ve doğruluk oranları

hesaplanmıştır. Bu oranlar GA’ da bulunan en iyi sonuçla karşılaştırıldığında, GA yöntemiyle tespit edilen öznelik setinin yapay sinir ağındaki başarımları diğerlerine göre daha yüksek olduğu gözlenmiştir.

VI. VARGI

EKG sinyal ayrıştırmasında kullanılan farklı öznelik setleri farklı sınıflandırma başarımları gösterebilmektedir. Bu yüzden de bu çalışmada sinyale uygun öznelik setinin seçiminin genetik algoritma yöntemi kullanılarak yapılması önerilmiştir. Farklı dalgacık tiplerinden elde edilen öznelikler seçim havuzu olarak kullanılmıştır. Sonuçlar da genetik algoritma yöntemi ile sınıflandırma doğruluğunu yükselten özneliklerin tespit edilebildiğini ve farklı tip dalgacıkların farklı seviyelerde seçilen katsayılarından elde edilen öznelik setinin standart olan tek-tip dalgacıktan elde edilen katsayılara göre EKG aritmi sınıflandırmasında başarımları yükselttiğini göstermiştir. Sonraki çalışmalarda öznelik seçim kümesine farklı yöntemlerden elde edilen öznelikler dahil edilerek genetik algoritmada seçimin başarımlarını artırma yönünde daha fazla parametre testi yapılması hedeflenmektedir.

KAYNAKÇA

- [1] Chazal, P., O'Dwyer, M., Reilly, R.B., Automatic classification of heart beats using ECG morphology and heart beat interval features. *IEEE Tran Biomed Eng.* 51(7):1196-206, 2004.
- [2] Osowski, Stanislaw, and Tran Hoai Linh. "ECG beat recognition using fuzzy hybrid neural network." *IEEE Transactions on Biomedical Engineering* 48, no. 11, 1265-1271, 2001.
- [3] Mazomenos, E.B., Chen, T., Acharyya, A., Bhattacharya, A., Rosengarten, J., Maharatna, K., A time domain morphology and gradient based algorithm for ECG feature extraction. In: *Proceedings of IEEE international conference on industrial technology (ICIT)*, pp. 117-122, 2012.
- [4] Ince, Turker, Serkan Kiranyaz, and Moncef Gabbouj. "A generic and robust system for automated patient-specific classification of ECG signals." *IEEE Transactions on Biomedical Engineering* 56, no. 5, 1415-1426, 2009.
- [5] Mironovova, Martina, and Jiri Bila. "Fast fourier transform for feature extraction and neural network for classification of electrocardiogram signals." In *Future Generation Communication Technology (FGCT), 2015 Fourth International Conference on*, pp. 1-6. IEEE, 2015.
- [6] Uslu, Erkan, and Gökhan Bilgin. "Arrhythmia classification by Local Fractional Fourier Transform." In *Signal Processing and Communications Applications Conference (SIU), 2013 21st*, pp. 1-4. IEEE, 2013.
- [7] Uslu, Erkan, and Gökhan Bilgin. "Exploiting locality based Fourier transform for ECG signal diagnosis." In *Applied Electronics (AE), 2012 International Conference on*, pp. 323-326. IEEE, 2012.
- [8] Cohen, Leon. "Time-frequency analysis." *Wavelets and Signal Processing*, 2012.
- [9] Castro, B., D. Kogan, and A. B. Geva. "ECG feature extraction using optimal mother wavelet." In *Electrical and Electronic Engineers in Israel, 2000. The 21st IEEE Convention of the*, pp. 346-350. IEEE, 2000.
- [10] Gao, Robert X., and Ruqiang Yan. *Wavelets: Theory and applications for manufacturing*. Springer Science & Business Media, 2010.
- [11] Shufni, Shazwani Ahmad, and Mohd Yusoff Mashor. "ECG signals classification based on discrete wavelet transform, time domain and frequency domain features." In *Biomedical Engineering (ICoBE), 2015 2nd International Conference on*, pp. 1-6. IEEE, 2015.
- [12] Vaneghi, Fatemeh Molaei, Maysam Oladazimi, F. Shiman, Afshan Kordi, M. J. Safari, and F. Ibrahim. "A comparative approach to ECG feature extraction methods." In *Intelligent Systems, Modelling and Simulation (ISMS), 2012 Third International Conference on*, pp. 252-256. IEEE, 2012.
- [13] Yang, Jihoon, and Vasant Honavar. "Feature subset selection using a genetic algorithm." *IEEE Intelligent Systems and their Applications* 13, no. 2 (1998): 44-49.
- [14] Tian, David. "A Multi-Objective Genetic Local Search Algorithm for Optimal Feature Subset Selection." In *Computational Science and Computational Intelligence (CSCI), 2016 International Conference on*, pp. 1089-1094. IEEE, 2016.
- [15] Kim, Ho-Duck, Chang-Hyun Park, Hyun-Chang Yang, and Kwee-Bo Sim. "Genetic algorithm based feature selection method development for pattern recognition." In *SICE-ICASE, 2006. International Joint Conference*, pp. 1020-1025. IEEE, 2006.
- [16] Oreski, Stjepan, and Goran Oreski. "Genetic algorithm-based heuristic for feature selection in credit risk assessment." *Expert systems with applications* 41, no. 4 (2014): 2052-2064.
- [17] Huang, Cheng-Lung, and Chieh-Jen Wang. "A GA-based feature selection and parameters optimization for support vector machines." *Expert Systems with applications* 31, no. 2 (2006): 231-240.
- [18] Moody, George B., and Roger G. Mark. "The impact of the MIT-BIH arrhythmia database." *IEEE Engineering in Medicine and Biology Magazine* 20, no. 3 (2001): 45-50.
- [19] Moody, George B., Roger G. Mark, and Ary L. Goldberger. "PhysioNet: a web-based resource for the study of physiologic signals." *IEEE Engineering in Medicine and Biology Magazine* 20, no. 3 (2001): 70-75.
- [20] MIT-BIH arrhythmia database, <http://www.physionet.org/>
- [21] Addison, P.S., *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, CRC Press, 2002
- [22] Talbi, El-Ghazali. *Metaheuristics: from design to implementation*. Vol. 74. John Wiley & Sons, 2009.
- [23] Sapna, S., A. Tamilarasi, and M. Pravin Kumar. "Backpropagation learning algorithm based on Levenberg Marquardt Algorithm." *Comp Sci Inform Technol (CS and IT)* 2 (2012): 393-398.