

İkili Guguk Kuşu Arama Algoritması ile Öznitelik Seçimi

Feature Selection Using Binary Cuckoo Search Algorithm

Yasin KAYA
Bilgisayar Mühendisliği
Karadeniz Teknik Üniversitesi
Trabzon, Türkiye
yasin@ktu.edu.tr

Özetçe—Guguk kuşu arama algoritması son yıllarda geliştirilmiş doğadan esinlenen algoritmalarından bir tanesidir. Algoritmaya guguk kuşlarının ilginç üreme stratejileri esin kaynağı olmuştur. İkili guguk kuşu arama algoritması ise arama uzayının sadece ikili sisteme indirildiği özel bir sürümüdür. Bu çalışmada EKG’de bulunan erken karıncık kasılması verisinin sınıflandırılması için öznitelik seçim işlemi için ikili guguk kuşu algoritması kullanılmıştır. Öznitelik olarak sinyalin 200 genlik değeri kullanılmış ve bunlardan sinyali daha iyi temsil edecek olanlar seçilmiştir. Seçilen öznitelikler kullanılarak erken karıncık kasılması vuruları %99,67 doğrulukla algılanmıştır.

Anahtar Kelimeler — *Guguk Kuşu Arama Algoritması; Lévy uçuşu; EKG; Öznitelik Seçimi; Aritmi Tespiti; k-NN; YSA.*

Abstract—The cuckoo search algorithm is one of the nature-inspired algorithms developed in recent years. Interesting breeding strategies of cuckoo birds have inspired the algorithm. The binary cuckoo search algorithm is a special version that reduces the search space to binary. In this study, we used binary cuckoo search algorithm in the feature selection process to classify the premature ventricular contraction beats in ECG. We used 200 amplitude values of the signal to represent one beat and selected those that represent the signal best. Using selected features, premature ventricular contraction beats were perceived with 99.67% of accuracy.

Keywords — *Cuckoo search algorithm; Lévy Flight; ECG; feature selection; arrhythmia classification; k-NN; Neural Networks.*

I. GİRİŞ

Guguk kuşu arama (GKA) algoritması Yang ve Deb tarafından 2009 yılında, guguk kuşlarının ilginç üreme stratejilerinden esinlenerek optimizasyon problemlerinin çözümü için geliştirilmiştir [1]. GKA yazarlar tarafından mühendislik tasarımı optimizasyonu için kullanılmıştır [2]. Li ve Yin iki yeni mutasyon kuralı geliştirerek algoritmayı düzenlemiş ve kendi kendine parametre optimizasyonu yapan sürümünü önerdi [3]. Yine mühendislik tasarımı ve optimizasyonu için Zheng ve Zhou gaus dağılımı temelli GKA yöntemi üzerinde çalıştı [4]. Rodrigues ve ark. İkili guguk kuşu algoritmasını

önerdi ve güç sistemleri dağıtımı üzerindeki kaçakları yakalamak için kullandı [5].

Bu çalışmada GKA algoritmasını EKG’de bulunan erken karıncık kasılması (EKK) verisinin sınıflandırılması için öznitelik seçimi aşamasında kullandık. EKK çeşitli aritmler arasında en önemli aritmlerden bir tanesidir [6]. EKK karıncıklardaki bir odaktan yayılan bir tetikleme ile karıncıkların zamanından önce kasılması sonucu meydana gelmektedir. Genellikle yapısal kalp rahatsızlıkları ile birlikte görülür ve ani ölüm riskini artırır [7].

Çalışmada yapılan deneylerde MIT-BIH aritmi veritabanından [8] alınan normal vuru (N) ve EKK vuruları kullanılmıştır. Çalışmanın amacı herhangi bir matematiksel öznitelik hesaplamadan, sinyalin genlik değerlerini öznitelik olarak kullanarak yüksek başarımla elde etmektir. Bu amaçla bir vuruya ait genlik değerleri arasından vuruyu daha iyi temsil edecek olanlar GKA algoritması tarafından seçilmiştir. Seçilen öznitelikler k-NN, yapay sinir ağları (YSA) ve Karar ağaçları (KA) sınıflandırıcıları ile sınıflandırılmış ve sonuçlar karşılaştırılmıştır. Sınıflandırma işlemi sonucunda, k-NN sınıflandırıcı kullanılarak %99,67 sınıflandırma doğruluğu elde edilmiştir.

II. MATERYAL VE YÖNTEM

A. Veri seti

MIT-BIH Aritmi Veritabanı yaklaşık 30 dk. süreyi içeren iki kanallı 48 kayıttan oluşmaktadır. Kayıtlar 1975-1979 yılları arasında BIH aritmi laboratuvarı tarafından 47 denek üzerinden elde edilmiştir. Veritabanındaki sinyaller 360 Hz. ile örneklenmiştir. Veritabanının ilk bölümü (100-124) yaygın vuru türlerini ve gürültüsüz sinyalleri içermektedir. İkinci bölümü ise (200-234) önemli klinik olayları ve gürültülü kayıtları içermektedir [8], [9].

Çalışmada MLII derivasyonu içeren kayıtlardan N ve EKK vurularının içeren 43 kayıttan alınan vurular kullanılmıştır [6]. Çalışmada kullanılan vuruların dağılımı Tablo 1’de gösterilmiştir.

TABLO I. ÇALIŞMADA KULLANILAN VURULARIN DAĞILIMI [6].

Dosya	Vurular		
	<i>N</i>	<i>EKK</i>	<i>Toplam</i>
100,101,103,205	401	42	443
106,107,108,109	99	622	721
111,112,113,115	300	1	301
116,117,118,119	300	568	868
121,122,123,124	300	51	351
200,201,202,203	400	456	856
205,207,208,209	300	296	596
210,212,213,214	300	669	969
215,217,219,220	400	389	789
221,222,223,228	400	400	800
230,231,234	300	6	306
Toplam	3500	3500	7000

B. Guguk Kuşu Arama Algoritması

Bu çalışmada GKA temelli bir arama algoritması kullanılmıştır. Yang ve Deb, guguk kuşlarının ilgi çekici üreme taktiklerinden ilham alarak, GKA adı verilen evrimsel bir optimizasyon yöntemi önerdi [1]. Guguk kuşların bazı türleri yumurtalarını diğer kuşların yuvalarına yerleştirerek üremeye parazitlik kullanır. Yumurtaları yuvadaki diğer yumurtaların şekil ve renk gibi morfolojik özelliklerini taklit edebilir. Plan başarısız olursa, ev sahibi kuş guguk kuşu yumurtasını yuvadan atabilir veya yuvayı tamamen terk edebilir ve başka bir yerde yeni bir yuva oluşturabilir. Bu bağlamda, GKA aşağıdaki üç kuralla özetlenebilir:

- 1) Yumurtlamak için her guguk kuşu rasgele bir yuva seçer.
- 2) Yuva sayısı sabittir ve yüksek kaliteli yumurta içeren yuva gelecek nesillere aktarılır.
- 3) Ev sahibi kuş, bir guguk yumurtasını (belirlenen bir olasılıkla) bulursa, ya yumurta atılır ya da ev sahibi kuş yuvayı terk eder [1].

Algoritmik olarak, bir yuvadaki her yumurta bir çözümü temsil eder ve bir guguk yumurtası yeni bir çözümü karakterize eder. Burada amaç yuvada iyi olmayan çözümlerin yerine yeni ve muhtemelen daha iyi çözümleri kullanmaktır.

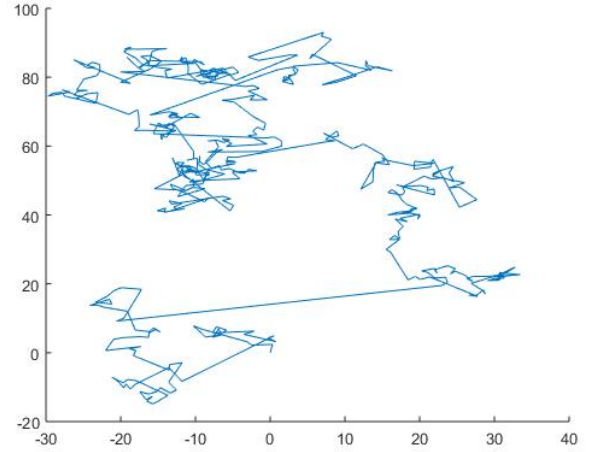
Bir guguk kuşu için bir sonraki çözümü ($x_i^{(t+1)}$) üretmek için Lévy uçuşu (1) ve (2) eşitliklerindeki gibi kullanılır [10].

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \otimes L(s, \lambda) \quad (1)$$

$$L(s, \lambda) \sim \frac{\lambda \Gamma(\lambda) \sin(\pi\lambda/2)}{\pi} \frac{1}{s^{1+\lambda}}, (s \gg 0) \quad (2)$$

Burada s adım uzunluğunu ve $\alpha > 0$ adım uzunluğu ölçek parametresini $x_i^{(t)}$ i. yuvadaki t . yumurtayı göstermektedir. $i (\forall i = 1, \dots, m)$ $j (\forall j = 1, \dots, d)$. Lévy uçuşu, bir Lévy dağılımından üretilen rasgele bir adım uzunluğunu kullanır ve eşitlik (2)'de gösterilmiştir. Lévy uçuşu, bazı hayvanlar tarafından yiyecek arama ve üreme stratejisi olarak kullanılan bir uçuş biçimidir. Bu tür bir davranış optimizasyon ve en

uygunu arama yöntemi olarak uygulanmıştır [1]. İki boyutlu uzayda 500 adımlık Lévy uçuşu şekil 1'de gösterilmiştir.



Şekil 1. İki boyutlu uzayda Lévy uçuşu örneği.

Algoritma 1: GKA algoritması temelli öznelik seçimi algoritması

```

GİRİŞ: Eğitim ve Test verisi  $E, T$ , algoritma parametresi  $p$ 
ve  $\alpha$  değeri, yuva sayısı  $m$ , boyut  $d$ , tekrar sayısı  $T$ .
ÇIKIŞ: En iyi sonucu veren öznelik vektörü  $BEST$ .
1.  $n$  adet yuva için başlangıç popülasyonunu oluştur  $x_i^j$ ,
2.  $f_i \leftarrow -\infty$ ,  $Globalfit \leftarrow -\infty$ 
3. For  $t (\forall t = 1, \dots, T)$ , do
4.   For  $n_i (\forall i = 1, \dots, m)$ , do.
5.      $n_i$  yuva için test ve eğitim verisini filtrele  $E', T'$ 
6.      $dog \leftarrow Sınıflandır(E', T')$ 
7.     if ( $dog > f_i$ ), then.
8.        $f_i \leftarrow dog$ 
9.       For  $j (\forall j = 1, \dots, d)$ ,
10.         $\hat{x}_i^j \leftarrow x_i^j(t)$ 
11.      end for
12.    end if
13.  end for
14.   $[maxfit, maxindex] \leftarrow \max(f)$ 
15.  if ( $maxfit > globalfit$ ), then.
16.     $globalfit \leftarrow maxfit$ ,  $BEST \leftarrow x_{maxindex}$ 
17.  end if
18.  For  $n_i (\forall i = 1, \dots, m)$ , do.
19.     $p_a \in [0, 1]$  olasılığı ile en kötü yuvaları seç ve
20.    onları yeni çözümler ile değiştir
21.  end for
22.  For  $n_i (\forall i = 1, \dots, m)$ , do.
23.    For  $j (\forall j = 1, \dots, d)$ , do.
24.       $x_i^j(t) = \hat{x}_i^j(t-1) + \alpha \otimes Lévy(\lambda)$ 
25.    End for
26.  End for
27. End for

```

Algoritmada son olarak, en düşük kalitede yumurtalara sahip yuvalar, $p_a \in [0,1]$ olasılığıyla yenileriyle değiştirilir. Öznitelik seçimi için kullanılan yöntemin algoritması Algoritma 1'de gösterilmiştir.

III. YAPILAN ÇALIŞMALAR

Bu çalışma MIT-BIH aritmi veritabanından alınan kayıtlar kullanılarak deneysel sonuçlar gerçekleştirilmiştir. Tablo 1'de N vurusundan 3750 ve EKK vurusundan 3750 vuru olmak üzere deneylerde toplam rasgele seçilmiş 7500 vuru kullanılmıştır. Matlab paket programında ön işleme adımları gerçekleştirilmiştir. Bir kalp döngüsünü gösteren sinyal değerlerini elde etmek için vuru ayrıştırma işlemleri yapılmıştır. Bu değerlerden vuruyu temsil edebilecekler GKA ile seçilmiş ve üç ayrı sınıflandırıcıda giriş verisi olarak kullanılarak sınıflandırma başarımları irdelenmiştir.

A. Sinyal ön işleme

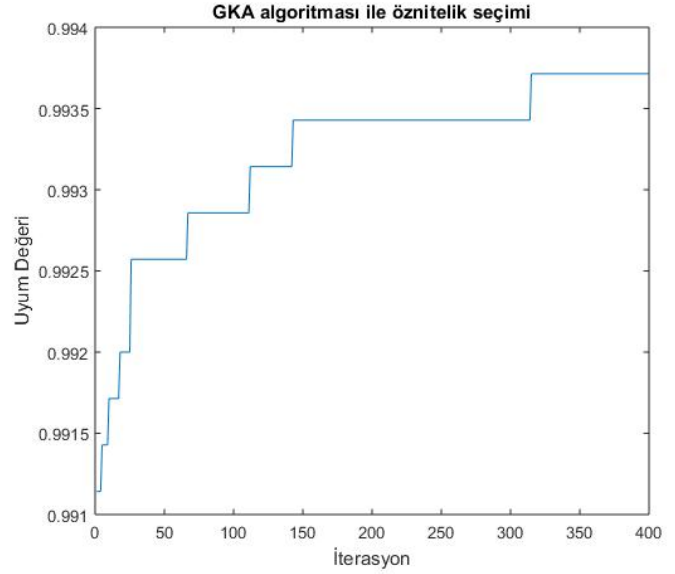
Sinyal üzerinde bulunan gürültüler ve dalgalanmalar sinyal işleme uygulamalarının başarımını olumsuz yönde etkilemektedir [11], [12]. Hastanın nefes alması, kıvrılması ve öksürmesi gibi günlük olağan davranışlarını sergilemesi esnasında kaydedilen sinyallerde taban hattı sapması sıklıkla gözükmemektedir. Taban hattı sapmasının frekansı genellikle 2 Hz. altındadır [13]. Bu nedenle sinyalde yapılacak ilk işlem bu gürültü ve dalgalanmaların giderilmesi olacaktır. Çalışmada sinyale ait genlik değerleri öznitelik olarak kullanıldığı için sinyalin taban hattının sıfıra indirilmesi uygun olacaktır. Bu amaçla sinyalden ilk olarak sinyal ortalaması çıkarılmış ve sinyal sıfır ortalamalı bir sinyal haline dönüştürülmüştür. İkinci adımda 2 Hz altında kalan frekans bileşenleri sinyalden yüksek geçiren filtre ile temizlenmiş ve taban hattı sapması giderilmiştir. Son aşamada ise her bir vuruyu temsil etmesi için 200 örnek genişliğinde pencere genişliği belirlenmiş, R tepesini ortalayacak şekilde bölütlenmiş ve bir sonraki aşamaya yani sınıflandırma adımına hazır hale getirilmiştir. R tepesinin yerinin tespiti için veritabanındaki notlardan faydalanılmıştır.

B. GKA Algoritması ile öznitelik seçimi

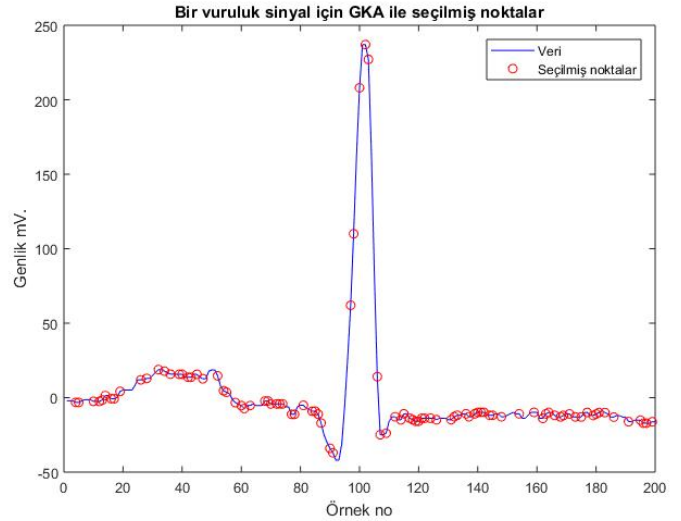
Bu çalışmada GKA algoritmasını temel alan bir öznitelik seçim yöntemi önerilmiştir. Önerilen yaklaşım Algoritma 1'de gösterilmiştir. Algoritmada ilk olarak başlangıç popülasyonu oluşturulmaktadır. Bu popülasyondaki değerler kullanılarak her bir yumurta için uyum değeri hesaplanmaktadır. Uyum değerlerinin hesaplanmasında Öklid uzaklık fonksiyonunu kullanan k-NN sınıflandırıcı kullanılmıştır. İlk popülasyon için en iyi uyum değeri ve bu değeri verecek öznitelik seti hesaplanır. En iyi uyum değeri korunacak şekilde ön tanımlı P_a olasılığı ile kötü sonuçlar kaldırılır ve yerine Lévy uçuşu ile yeni rasgele çözümler eklenir. Uyum değerleri ve en iyi sonucu veren değerler yeniden hesaplanır. Algoritma ön tanımlı tekrar sayısına kadar tekrarlanır. Burada uyum değeri hesaplanırken ilk olarak veri seti ilgili öznitelik listesine göre filtrelenir ve eğitim ve test seti olarak ikiye ayrılır. Sınıflandırıcı algoritma eğitim seti ile eğitilip test seti ile sınıflandırma doğruluğu yani uyum değeri hesaplanır.

Algoritmanın uygulanmasında $T=400$ tekrar, ev sahibi kuşun guguk yumurtasını bulma olasılığı $p_a = 0.25$ ve her nesildeki çözüm sayısı 25 olarak belirlenmiştir. Bu değerler

deneysel sonuçlar ile belirlenmiştir. GKA algoritması çalıştırıldığında 400 tekrar için uyum değerinin değişimi şekil 2'de gösterilmiştir. GKA algoritması ile seçilmiş noktalar şekil 3'de gösterilmiştir.



Şekil 2. GKA algoritması ile her tekrarda elde edilen uyum fonksiyonu sonucu.



Şekil 3. GKA algoritması ile seçilen noktalar.

GKA algoritması ile 200 olan öznitelik vektör boyutu 95 olarak belirlenmiştir.

C. Sınıflandırma

Bir önceki aşamada veriler GKA algoritması ile değerli olan öznitelikler belirlendi. Bu aşamada ilk olarak seçilmiş öznitelikler kalacak şekilde deney seti filtrelenir ve deneylerde kullanılacak filtrelenmiş veri seti elde edilir. Sınıflandırma sonuçlarının elde edilmesi için k-NN, YSA ve KA sınıflandırıcıları karşılaştırılmak üzere deneylerde kullanılmıştır. Deneylerde kullanılan sınıflandırma

parametreleri şu şekildedir: k-NN sınıflandırıcıda k=1 olarak belirlenmiştir. Uzaklık fonksiyonu olarak öklit uzaklık fonksiyonu kullanılmıştır. Deneylerde 20 düğümden oluşan tek gizli katmanlı geri yayılım algoritması ile eğitilen bir YSA kullanılmıştır. Öğrenme oranı 0,3 ve öğrenme tur sayısı 500 olarak belirlenmiştir. KA ile sınıflandırma yapılırken maksimum derinlik 20, minimum yaprak sayısı 2 ve minimum bölünme sayısı 4 olarak belirlenmiştir. Tüm deneylerde sınıflandırma başarımının hesaplanması için 10-kez çapraz doğrulama kullanılmıştır. 10-kez çapraz doğrulamada veri seti 10 eşit alt kümeye ayrılır ve bir alt küme test için diğer 9'u eğitim için kullanılır ve işlemler 10 kez tekrar edilir. Elde edilen sınıflandırma başarımlarının ortalaması alınır ve nihai sınıflandırma sonuçları elde edilmiş olur.

Sınıflandırma başarımının değerlendirilmesi için doğruluk, duyarlılık, kesinlik ve F1-skoru hesaplanmıştır. Bu ölçütler eşitlik (3)-(6)'da verilmiştir.

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{Kesinlik} = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

$$F1 = 2 \cdot \frac{\text{Kesinlik} \cdot \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (6)$$

IV. SONUÇLAR VE TARTIŞMA

Bu çalışmada MIT-BIH aritmi veritabanından alınan 2 sınıfa ait (Normal ve EKK) 7000 vuru sınıflandırılmıştır. Testlerde veritabanından bulunan EKK vurularının hemen hemen yarısı kullanılmışken normal vuru daha yaygın olduğu için her sinyal dosyasından yaklaşık 100 vuru rasgele seçilerek deney seti oluşturulmuştur. Her bir vuruyu temsil etmesi için vuruya ait R tepesini ortalamayan 200 örnekten oluşan pencere verileri belirlenmiştir. GKA algoritması ile bu noktalardan değerleri olanlar belirlenmiş ve sınıflandırma aşamasında bu noktalar kullanılmıştır. Sınıflandırma işlemi için 3 sınıflandırıcı kullanılmış ve elde ettikleri başarımlar tablo 2'de gösterilmiştir.

Sınıflandırma aşamasında sonuçlar hesaplanırken EKK vurusu pozitif sınıf olarak alınmış ve en yüksek başarımları %99,67 sınıflandırma doğruluğu ile k-NN sınıflandırıcı elde etmiştir. Bu sınıflandırma işlemine ait karşıtlık matrisi tablo 3'te gösterilmiştir. Bu çalışmada EKK vurusunun algılanması için GKA temelli bir hesaplama modeli önerilmiştir. GKA özneliklerin seçilmesi için kullanılmış, seçilen öznelikler üç ayrı sınıflandırıcı ile test edilmiş ve sınıflandırma başarımları hesaplanmıştır.

TABLO II. SINIFLANDIRMA SONUÇLARI (%)

Sınıflandırma metriği	Sınıflandırıcılar		
	k-NN	YSA	KA
Doğruluk	99,67	98,59	97,47
Duyarlılık	99,51	99,09	96,63
Kesinlik	99,83	98,09	98,31
F1	99,67	98,6	97,45

TABLO III. K-NN SINIFLANDIRICI İÇİN KARŞITLIK TABLOSU

Tahminler	Doğru Değerler		
	N ^a	EKK ^b	Kesinlik
N	3494	17	%99,52
EKK	6	3483	%99,83
Duyarlılık	%99,83	%99,51	

^a: Normal Vuru

^b: Erken Karıncık Kasılması Vurusu

Önerilen yaklaşımda matematiksel öznelik hesaplamaksızın sinyalin genlik değerlerinin kullanılması ve bu özneliklerin sayısının da öznelik seçim aşamasında yaklaşık yarı yarıya azaltılması ile yüksek başarımlar elde edilmesi gelecekte geliştirilecek olan aritmi algılama sistemlerine kolayca adapte edilebileceğini göstermektedir. Önerilen yaklaşımın uygulanabilirliğinin kolay olması ve iş yükünün az olması sayesinde geliştirilecek mobil uygulamalara uyarlanabilir ve giyilebilir sağlık ürünlerinde kullanılabilir. Önerilen yaklaşım gelecekte daha fazla aritmi türünü tanıyacak şekilde genişletilebilir.

KAYNAKLAR

- [1] X.-S. Yang and Suash Deb, "Cuckoo Search via Levy flights," in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, 2009, pp. 210–214.
- [2] X.-S. Yang and S. Deb, "Engineering Optimisation by Cuckoo Search," *Int. J. Math. Model. Numer. Optim.*, vol. 1, no. 4, pp. 330–343, 2010.
- [3] X. Li and M. Yin, "Modified cuckoo search algorithm with self adaptive parameter method," *Inf. Sci. (Ny)*, vol. 298, pp. 80–97, 2015.
- [4] H. Zheng and Y. Zhou, "A novel cuckoo search optimization algorithm based on Gauss distribution," *J. Comput. Inf. Syst.*, vol. 8, no. 10, pp. 4193–4200, 2012.
- [5] D. Rodrigues *et al.*, "BCS: A Binary Cuckoo Search algorithm for feature selection," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, 2013, pp. 465–468.
- [6] Y. Kaya and H. Pehlivan, "Classification of Premature Ventricular Contraction in ECG," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 7, pp. 34–40, 2015.
- [7] G. K. Lee, K. W. Klarich, M. Grogan, and Y.-M. Cha, "Premature ventricular contraction-induced cardiomyopathy: a treatable condition," *Circ. Arrhythm. Electrophysiol.*, vol. 5, no. 1, pp. 229–36, Feb. 2012.
- [8] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.
- [9] G. Moody and R. Mark, "The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it," in *[1990] Proceedings Computers in Cardiology*, 1990, pp. 185–188.
- [10] X.-S. Yang, "Mathematical Analysis of Nature-Inspired Algorithms," 2018, pp. 1–25.
- [11] Y. Kaya and H. Pehlivan, "Comparison of classification algorithms in classification of ECG beats by time series," in *23rd Signal Processing and Communications Applications Conference (SIU)*, 2015, pp. 407–410.
- [12] Y. Kaya, H. Pehlivan, and M. E. Tenekeci, "Effective ECG beat classification using higher order statistic features and genetic feature selection," *Biomed. Res.*, vol. 28, no. 17, pp. 7594–7603, 2017.
- [13] Y. Kaya and H. Pehlivan, "Feature Selection Using Genetic Algorithms for Premature Ventricular Contraction Classification," in *9th International Conference on Electrical and Electronics Engineering, ELECO*, 2015.