

Koroner Arter Hastalığı Tanısı İçin Alan Bilgisi İçeren Topluluk Öznitelik Seçim Yöntemi

An Ensemble Feature Selection Methodology That Incorporates Domain Knowledge for Cardiovascular Disease Diagnosis

BURAK KOLUKISA
Bilgisayar Mühendisliği Bölümü
Abdullah Gül Üniversitesi
Kayseri, Türkiye
burak.kolukisa@agu.edu.tr

V. CAGRI GUNGOR
Bilgisayar Mühendisliği Bölümü
Abdullah Gül Üniversitesi
Kayseri, Türkiye
cagri.gungor@agu.edu.tr

BURCU BAKIR GUNGOR
Bilgisayar Mühendisliği Bölümü
Abdullah Gül Üniversitesi
Kayseri, Türkiye
burcu.bakir@agu.edu.tr

Özetçe— Koroner Arter Hastalığı (KAH), arterlerin duvarlarında aterom denilen yağlı madde birikiminin bir sonucu olarak kalbin yeterince beslenememesi durumudur. KAH, 2016 yılında dünyadaki toplam ölümlerin %31'ine (17,9 milyon) neden olmuştur ve teşhis edilmesi zordur. 2030 yılında, yaklaşık olarak 23,6 milyon insanın bu hastalıktan öleceği tahmin edilmektedir. Makine öğrenmesi ve veri madenciliği yöntemlerinin gelişmesiyle birlikte, bazı fiziksel ve biyokimyasal değerleri inceleyerek, KAH'nı ucuz ve zahmetsiz bir şekilde teşhis etmek mümkün olabilir. Bu çalışmada, KAH sınıflandırma problemi için, uzman bilgisini içine alan yeni bir topluluk öznitelik seçim yöntemi önerilmiştir. Önerilen çözüm, UCI Cleveland KAH veri kümesi üzerinde uygulanmış, farklı sınıflandırma algoritmaları kullanılarak, farklı performans ölçütleri karşılaştırılmıştır. Gerçekleştirdiğimiz deneylerde, önerdiğimiz çözümün, MLP sınıflandırıcısı ve seçilen 9 öznitelik kullanıldığında, %85.47 doğruluk, %82.96 hassasiyet ve 0.839 F-ölçüsüne ulaştığı gösterilmiştir. Bu çalışmanın devamında, hastanelerde gerçek zamanlı veriler üzerinde, hızlı bir şekilde KAH tahminlemesi yapabilecek bir makine öğrenmesi modeli oluşturabilmeyi amaçlıyoruz.

Anahtar Kelimeler — Veri Madenciliği; Makine Öğrenmesi; Sınıflandırma Algoritması; Koroner Arter Hastalığı

Abstract— Coronary Artery Disease (CAD) is the condition where, the heart is not fed enough as a result of the accumulation of fatty matter called atheroma in the walls of the arteries. In 2016, CAD accounts for 31% (17.9 million) of the world's total deaths and its diagnosis is difficult. It is estimated that approximately 23.6 million people will die from this disease in 2030. With the development of machine learning and data mining techniques, it might be possible to diagnose CAD inexpensively and easily via examining some physical and biochemical values. In this study, for the CAD classification problem, a novel ensemble feature selection methodology that incorporates domain knowledge is proposed. Via applying the proposed methodology on the UCI Cleveland CAD dataset and using different classification algorithms, performance metrics are compared. It is shown that in our experiments, when Multilayer Perceptron classifier is used with 9 selected features, our proposed solution reached 85.47% accuracy, 82.96% accuracy and 0.839 F-Measure. As a future work, we aim to generate a machine learning model that can quickly diagnose CAD on real-time data in hospitals.

Keywords — Data mining; Machine Learning; Classification Algorithm; Coronary Artery Disease

I. GİRİŞ

Dünya Sağlık Örgütü tarafından 2016 yılında yayınlanan bir rapora göre, dünyadaki toplam ölümlerin %31'ine, koroner arter hastalığı (KAH) neden olmaktadır [1]. KAH 2016 yılında, diğer ölümcül hastalıklar arasında ilk sıradadır ve yaklaşık olarak 17.9 milyon insan, KAH nedeniyle hayatını kaybetmiştir. Dünya Sağlık Örgütü, KAH'nın 2030 yılında yaklaşık 23,6 milyon kişinin ölümüne neden olacağını tahmin etmektedir [1]. KAH, kalbi besleyen büyük arter damarların ateroskleroz olması durumudur. Ateroskleroz, arterlerin duvarlarında aterom adı verilen yağlı maddenin birikmesidir ve bu durum, damarlarda daralmaya ve tıkanıklığa neden olur. KAH zamanla ortaya çıktığı için, ilk başta teşhis edilmesi zordur, ancak damarlarda tıkanma veya kişinin kalp krizi geçirmesi halinde KAH teşhisi konulmaktadır. Kalp krizi geçiren birisi, bir tür Kardiyovasküler Hastalığa (KVH) sahiptir. KVH terimi, kalp veya kan damarlarını etkileyen her türlü hastalığı kapsayan en genel terimdir ve KAH da KVH'lara dahildir. Bu hastalığın belirlenmesi, belirli cihazlar ve uzmanlar gerektirir ve her bir yöntemin çeşitli avantajları, dezavantajları vardır. KVH'nın dörtte üçünün görüldüğü düşük ve orta gelirli ülkelerde [2], KVH'nı tanısı mümkün olmayabilir. Düşük ve orta gelirli ülkelerdeki insanlar, yüksek gelirli ülkelerdeki insanlara kıyasla, erken tanı ve tedavi programlarından sıklıkla yararlanamadığından, hastalıktan olumsuz etkilenmektedirler. Bu nedenle, bu ülkelerde KVH olan birçok kişi, geç tanı nedeniyle genç yaşta ölmektedir [3].

Gelişmiş bilgisayar sistemleri, sağlık verilerinin daha düşük maliyetle elde edilmesine yardımcı olabilir. Veri madenciliği ve makine öğrenmesi gibi, sağlık alanında popüler hale gelen yöntemler, pahalı cihazların sonuçları yerine, spesifik parametreleri inceleyerek, KVH riskini tahmin etmeyi mümkün kılabilir. Araştırmacıların KVH sınıflandırma problemi için, kamuya açık olan UCI makine öğrenme deposundaki KVH'na ait veri kümelerini kullanarak yaptığı çalışmalar, Tablo I'de karşılaştırılmıştır. Mevcut çalışmalar değerli bilgiler içermesine rağmen, genel olarak kabul edilmiş standart bir makine öğrenme modeli yoktur. Tablo I'de görülebileceği üzere, bu çalışmalar genel olarak doğruluk oranlarını iyileştirmeyi hedeflemiş ve çoğu çalışma, elde etmiş oldukları doğruluk, duyarlılık, kesinlik, F-ölçüsü ve eğrinin altındaki alan açısından ayrıntılı bir performans değerlendirmesi sunmamıştır. Gerçek hayatta KVH tanısı için, farklı performans ölçütleri de önemlidir.

TABLO I. KAH teşhisi için farklı sınıflandırma yöntemlerinin karşılaştırılması

Yazarlar	Yıl	Veri Kümesi	Yöntem	HS (%)	KS (%)	FÖ	EAA	D (%)
Kemal Polat et al [5]	2007	Cleveland	KNN	92.30	92.30	-	-	87
Sellappan et al [6]	2008	UCI	NB	-	-	-	-	95
Resul Das et al [7]	2009	Cleveland	ANN	80.95	95.91	-	-	89.01
Anbarasi et al [8]	2010	UCI	DT	-	-	-	-	99.2
Shouman et al [9]	2011	Cleveland	DT	77.90	85.20	-	-	84.1
Rajalaxmi et al [10]	2014	Cleveland	BABC	-	-	-	-	86.4
El-Biary et al [11]	2015	Cleveland	DT	-	-	-	-	78.54
Verma et al [12]	2016	Cleveland	MLR	-	-	-	-	90.28
Redd et al [13]	2019	UCI	RF	88	95.4	-	-	92.16

HS: Hassasiyet, KS: Kesinlik, FÖ: F-Ölçüsü, EAA: Eğri Altındaki Alan, D: Doğruluk

Bu çalışmada, bir uzman bazlı öznitelik seçme algoritmasını [4] ve 7 işlemsel öznitelik seçme algoritmasını birleştiren bir topluluk öznitelik seçme algoritması sunulmuştur. Sunulan yöntem, UCI Cleveland KAH veri kümesi, ve sekiz farklı sınıflandırma algoritması kullanılarak test edilmiştir. KVH tanısı için en uygun öznitelik seçim algoritmasının, en uygun özniteliklerin ve en uygun sınıflandırıcının belirlenerek, en uygun modelin geliştirilmesi hedeflenmiştir. Özellikle düşük ve orta gelirli ülkelerdeki doktorların, ellerindeki mevcut veri kümesini kullanarak, KVH'nı hızlı bir şekilde teşhis edilebilmesi amaçlanmıştır.

Bu makalenin geri kalanı şu şekilde düzenlenmiştir: Bölüm 2'de, veri kümeleri ve sunulan yöntem tanıtılmıştır. Bölüm 3'te elde edilen bulgular sunulmuştur. Son olarak, Bölüm 4'te bulgular tartışılmıştır, ve bu çalışmaya ek olarak gelecekte yapılacaklardan bahsedilmiştir.

II. VERİ KÜMESİ VE YÖNTEMLER

A. Veri Kümesi

UCI KAH veri kümesi; Cleveland, Long Beach VA, Hungarian ve Switzerland olmak üzere 4 veri kümesinden oluşmaktadır. Bu veri kümeleri toplam 76 öznitelik içermesine rağmen, birçok örnekte eksiklikler olmasından dolayı özniteliklerin sadece 14'ü kullanılmaktadır. Mevcut KAH veri kümelerindeki örneklerin etnisitesi ve örnek sayıları birbirinden farklıdır. Bu nedenle, bu çalışmada modelimizi mevcut çalışmalarla karşılaştırmak için, yaygın olarak kullanılan UCI Cleveland KAH veri kümesi seçilmiştir. UCI Cleveland veri kümesinde, 303 örnek, 13 öznitelik yer almaktadır, bu özniteliklerin detayları bir önceki çalışmamızda sunulmuştur [4]. Bu veri kümesindeki altı eksik örneğin sentetik verilerle doldurulması yerine, bu çalışmada bu örnekler veri kümesinden çıkartılmıştır.

B. Öznitelik Seçim Yöntemleri

Makine öğrenmesi işlemlerinde modelin performansı özniteliklere bağlıdır ve daha kaliteli sonuçlar elde etmek için anlamlı öznitelikler kullanmak gerekmektedir. Bu amaçla öznitelik seçim algoritmaları, ilgisi olmayan veya tahmin gücü olmayan öznitelikleri kaldırırken, tahmin gücü yüksek olan öznitelikleri öne çıkarır. Öznitelik seçim yöntemlerinin temel amaçları, modelin karmaşıklığını azaltarak kolay yorumlanmasını sağlamak, aşırı uyumlama ve gürültüyü ortadan kaldırmak, makine öğrenmesi algoritmalarının daha hızlı çalışmasını sağlamak ve performans sonuçlarını iyileştirmektir. Öznitelik seçimi üç ana kategoride gruplandırılır: i) Sınıflandırıcıdan bağımsız

filtre temelli yöntemler; ii) Sınıflandırıcı ile etkileşime giren sarıcı tabanlı yöntemler; iii) Her iki yöntemin de avantajlarını birleştiren ve öznitelik seçimi ile sınıflandırmayı aynı anda gerçekleştiren gömülü yöntemler [14,15]. Her yöntemin avantajları ve dezavantajları vardır. Veri kümesi büyük değilse, gömülü ve sarma yöntemler tercih edilmelidir. Veri kümesi büyükse, filtre yöntemi tercih edilmelidir. Günümüzde yeni öznitelik seçme yöntemleri ortaya çıkmakta ve öznitelik seçme yöntemlerinin sayısı artmaktadır. Tüm bu öznitelik seçim yöntemleri arasında, üzerinde çalışılan veri kümesi için hangi öznitelik seçim yönteminin uygun olduğuna karar vermek zorlaşmıştır.

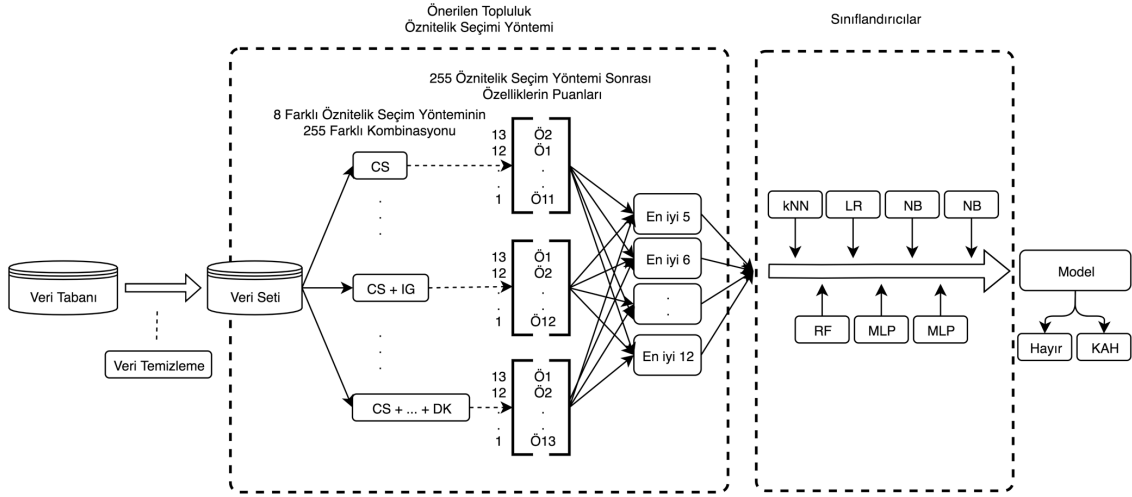
Bu çalışmada, ki-kare (chi square, CS), kazanç oranı (gain ratio, GR), bilgi kazancı (information gain, IG), rölyef f (relief F, RF), destek vektör makinesi (support vector machine, SVM), arı araması (bee search, BS), koşullu karşılıklı bilgi maksimizasyonu (conditional mutual information maximization, CMIM) yöntemleri kullanılmıştır ve bu işlemsel öznitelik seçim yöntemlerine alternatif olarak, alan bilgisi (domain knowledge, DK) tabanlı öznitelik seçim yönteminden [4] faydalanılmıştır. Önerilen alan bilgisi tabanlı öznitelik seçim yönteminde özellikler, tıp doktorunun uzmanlığına göre puanlandırılıp sıralanmıştır.

C. Sınıflandırma Yöntemleri

Bu çalışmada, yedi farklı bireysel sınıflandırıcı kullanılmıştır. Kullanılan sınıflandırma teknikleri, en yakın k komşusu (k nearest neighbour, kNN), lojistik regresyon (logistic regression, LR), doğrusal discriminant analizi (linear discriminant analysis, LDA), saf bayes (naive bayes, NB), destek vektör makinesi (support vector machine, SVM), çok katmanlı algılayıcı (multilayer perceptron, MLP) ve rastgele ormandır (random forest, RF). Bu çalışmada, Python programlama dili ve scikit-learn kütüphanesi kullanılarak sınıflandırma işlemi uygulanmıştır.

D. Önerilen Topluluk Öznitelik Seçim Yöntemi

Öznitelik seçim yöntemleri, veri kümelerinde önemli özellikleri belirler. Önemli bulunan özniteliklere yüksek puan verilirken, önemsiz özniteliklere düşük skorlar verilir. Ancak, her öznitelik seçim yöntemi aynı performansı vermez. Bu nedenle, önemli öznitelikleri belirlerken topluluk öznitelik seçim yöntemiyle belirlemek, tek bir yöntem ile belirlemekten daha iyidir. Bu bağlamda biz bu çalışmada, Şekil 1'de gösterilen topluluk öznitelik seçme yöntemini önerdik ve bu yöntemi, KAH tanısı için, farklı sınıflandırma yöntemleriyle beraber uyguladık.



Şekil 1. Önerilen topluluk öz nitelik seçim yöntemi ve sınıflandırma işleminin genel gösterimi

Daha önceki çalışmamızda, topluluk seçim yöntemi olarak, her bir öz niteliğin n farklı öz nitelik seçim yönteminde aldığı puanların ortalamasını kullanmayı önermiş ve bu yöntem ile KAH veri kümelerinde başarılı sonuçlar elde edilebileceğini göstermiştik [4,16]. Bu çalışmada ise, [4,16]'da sunulan yöntem kullanılarak, n öz nitelik seçim yönteminin tüm alt kombinasyonlarının ($2^n - 1$) her biri için, tüm öz niteliklerin (Ö) puanları belirlenir ve öz nitelik seçim sıralama listesi oluşturulur. Bu listelerde öz nitelikler 13'den 1'e kadar azalan puana göre sıralanmıştır. Her listenin en önemli k özelliği, 7 farklı sınıflandırıcıda kullanılır. Her öz nitelik seçim kombinasyonunda k değeri (seçilen öz nitelik sayısı), 5'den 12'ye kadar birer artırılarak, her sınıflandırıcıda ayrı ayrı test edilmiştir. Bahsedilen bu sekiz öz nitelik seçim yönteminin tüm olası kombinasyonları (255 adet), 7 farklı sınıflandırıcıda, tabakalı 10 kat çapraz doğrulama ile denenmiş olup, veri kümesine en uygun öz nitelik seçim algoritması belirlenmeye çalışılmıştır.

III. BULGULAR

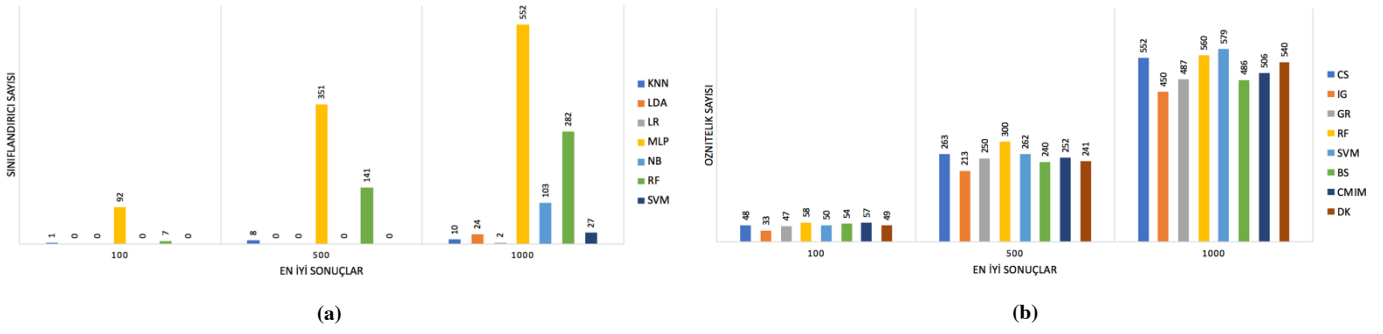
Bu çalışmada, 303 örnek ve 14 öz nitelikten oluşan halka açık Cleveland KAH veri kümesi kullanılmış olup, veri temizleme (6 örneğin eksik özellik içermesi nedeniyle çıkarılması) dışında herhangi bir ön işlem yapılmamıştır. Gerçekleştirdiğimiz deneylerde toplam 22.440 model elde edilmiştir. Tablo II'de, test edilen her bir sınıflandırıcı için, KAH veri kümesinde elde edilen en iyi performans sonuçları ve bu en iyi model için kullanılan öz nitelik seçim yöntemi kombinasyonu gösterilmiştir. En iyi performans ölçütleri, MLP sınıflandırıcısı ve CMIM öz nitelik seçim yöntemi ile 9 öz nitelik seçilerek elde edilmiştir. Bu model,

%85,47 doğruluk, %82,96 hassasiyet, %86,22 kesinlik, 0.839 F-ölçüsü ve 0.911 EAA sonucu göstermiştir. Cleveland veri kümesinde F-Ölçüsü performans ölçütüne göre ilk 100, ilk 500 ve ilk 1000 en iyi modelde en sık kullanılan sınıflandırma yöntemleri ve öz nitelik seçim yöntemleri analiz edilmiştir ve Şekil 2'de gösterilmiştir. Şekil 2a'da görüldüğü üzere, MLP sınıflandırıcısı, diğer sınıflandırıcılara göre ilk 100, ilk 500, ilk 1000 en iyi F-ölçüsü üreten modellerde, büyük ara farkla daha fazla yer almıştır. Bu analiz, Cleveland KAH veri kümesinde, MLP sınıflandırıcısı kullanıldığında diğer sınıflandırıcılara kıyasla daha iyi sonuçlar elde edilebileceğini göstermiştir. Şekil 2b'de görüldüğü üzere, öz nitelik seçim yöntemlerinin frekansları birbirine çok yakın çıkmıştır. Diğer bir deyişle, Cleveland KAH veri kümesinde kullanılan her bir öz nitelik seçim yöntemi, F-ölçüsü metriğine göre ilk 100, ilk 500, ilk 1000 modelde benzer sayılarda yer almıştır. Bu bulgu, Cleveland veri kümesinde yer alan öz nitelik sayısının (13) az olmasından veya örnek sayısının (303) az olmasından kaynaklanıyor olabilir. Şekil 3'de Cleveland veri kümesi üzerinde, farklı sayıda öz nitelikler için, MLP sınıflandırıcısının öz nitelik seçimi uygulanmadığındaki ve topluluk öz nitelik seçimi kullanıldığındaki maksimum, minimum, ve ortalama F-Ölçüsü değerleri gösterilmiştir. MLP sınıflandırıcısı ile Cleveland veri kümesinde topluluk öz nitelik seçimi uygulanarak elde edilen F-ölçüsü değerlerinin ortalaması, öz nitelik seçimi uygulanmadığındaki değerlerin biraz üzerinde çıkarken, maksimum değerleri, öz nitelik seçimi uygulanmadığındaki değerlerin oldukça üzerindedir.

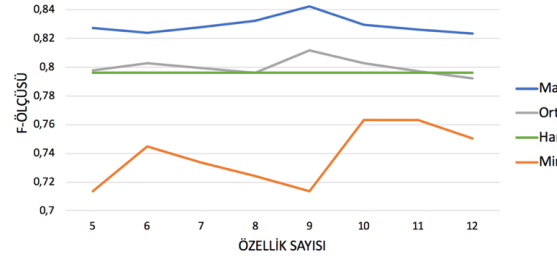
Tablo II. Test edilen her bir sınıflandırıcı için, KAH veri kümesinde öz nitelik seçim yöntemlerinin farklı kombinasyonları ile elde edilen en iyi performans sonuçları

Sınıflandırıcı	ÖSYK	DÖS	HS (%)	KS (%)	FÖ	EAA	D (%)
MLP	CMIM	9	82.96	86.22	0.839	0.911	85.47
RF	IG+RF+SVM	10	80.00	81.67	0.812	0.903	84.74
KNN	CS+RF+SVM	5	81.37	85.93	0.825	0.903	84.41
LDA	IG+SVM	11	78.51	84.52	0.809	0.901	83.41
LR	RF+SVM+DK	10	79.34	83.78	0.811	0.902	83.11
SVM	RF	10	80.05	83.13	0.811	0.911	83.10
NB	SVM	7	80.82	81.67	0.811	0.892	82.78

ÖSYK: Öz nitelik Seçim Yöntemi Kombinasyonu, DÖS: Dahil Edilen Öz nitelik Sayısı, HS: Hassasiyet, KS: Kesinlik, FÖ: F-Ölçüsü, EAA: Eğri Altındaki Alan, D: Doğruluk



Şekil 2. Cleveland veri kümesinde F-Ölçüsü performans ölçütüne göre ilk 100, ilk 500 ve ilk 1000 en iyi modelde en sık kullanılan (a) sınıflandırma yöntemleri, (b) öznelik seçim yöntemleri.



Şekil 3. Cleveland veri kümesi üzerinde, farklı sayıda öznelikler için, MLP sınıflandırıcısının Öznelik Seçimi Uygulanmadığındaki (ÖSU, Ham) ve topluluk öznelik seçimi kullanıldığında Maksimum (Max), Minimum (Min), Ortalama (Ort) F-Ölçüsü Değerleri

IV. TARTIŞMA VE SONUÇ

KAH tanısı için uluslararası kabul görmüş standart bir makine öğrenimi yaklaşımı yoktur. Bu çalışmanın ana amacı, hastanın fiziksel ve biyokimyasal değerlerini veri madenciliği ve makine öğrenimi teknikleri ile analiz ederek, KAH tanısının daha ekonomik ve verimli bir şekilde koyulmasını, ve KAH kaynaklı ölümlerin azalmasını sağlamak; ve böylece topluma katkıda bulunabilmektir. Daha özel olarak amacımız, önerilen topluluk öznelik seçim yöntemini kullanarak, tutarlı bulgular üretecek sınıflandırıcıyı belirleyip, KAH tanısı için istikrarlı bir model elde edebilmektir. Gelecek çalışmalarımızda, önerilen model ile diğer halka açık KAH veri kümelerini incelemeyi planlıyoruz. Böylece, farklı KAH veri kümelerinde iyi performans gösterecek tek bir model bulmayı amaçlıyoruz. Ayrıca, Türk toplumu bireylerinin verilerini içeren kamuya açık bir KAH verisi yoktur. Türk kardiyologlarla iş birliği yaparak, topluma özgü bir veri kümesinin geliştirilmesine katkıda bulunmak, modelimizi bu veri kümesinde de uygulamak ve KAH teşhisi için Türk nüfusuna özgü özelliklerin olup olmadığını gözlemlemek istiyoruz.

KAYNAKLAR

- [1] Cardiovascular diseases (CVDs)," *World Health Organization*. [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). [Accessed: 17-Jan-2020].
- [2] "Mortality due to Cardiovascular Diseases and Diabetes," *ChartsBin*. [Online]. Available: <http://chartsbin.com/view/2621>. [Accessed: 17-Jan-2020].
- [3] Cardiovascular disease diagnosis, <https://www.news-medical.net/health/Cardiovascular-Disease-Diagnosis.aspx> (13.11.2019).
- [4] Kolukisa, Burak, et al. "Evaluation of Classification Algorithms, Linear Discriminant Analysis and a New Hybrid Feature Selection Methodology for the Diagnosis of Coronary Artery Disease." 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018.
- [5] Polat, Kemal, Seral Şahan, and Salih Güneş. "Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing." *Expert Systems with Applications* 32.2 (2007): 625-631.
- [6] Palaniappan, Sellappan, and Rafiah Awang. "Intelligent heart disease prediction system using data mining techniques." 2008 IEEE/ACS international conference on computer systems and applications. IEEE, 2008.
- [7] Das, Resul, Ibrahim Turkoglu, and Abdulkadir Sengur. "Effective diagnosis of heart disease through neural networks ensembles." *Expert systems with applications* 36.4 (2009): 7675-7680.
- [8] Anbarasi, M., E. Anupriya, and N. C. S. N. Iyengar. "Enhanced prediction of heart disease with feature subset selection using genetic algorithm." *International Journal of Engineering Science and Technology* 2.10 (2010): 5370-5376.
- [9] Shouman, Mai, Tim Turner, and Rob Stocker. "Using decision tree for diagnosing heart disease patients." *Proceedings of the Ninth Australasian Data Mining Conference-Volume 121*. Australian Computer Society, Inc., 2011.
- [10] Subanya, B., and R. R. Rajalaxmi. "Artificial bee colony based feature selection for effective cardiovascular disease diagnosis." *International Journal of Scientific & Engineering Research* 5.5 (2014): 606-612.
- [11] El-Bialy, Randa, et al. "Feature analysis of coronary artery heart disease data sets." *Procedia Computer Science* 65 (2015): 459-468.
- [12] Verma, Luxmi, Sangeet Srivastava, and P. C. Negi. "A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data." *Journal of medical systems* 40.7 (2016): 178.
- [13] Reddy, N. Satish Chandra, et al. "Classification and Feature Selection Approaches by Machine Learning Techniques: Heart Disease Prediction." *International Journal of Innovative Computing* 9.1 (2019).
- [14] Guyon, Isabelle, et al., eds. *Feature extraction: foundations and applications*. Vol. 207. Springer, 2008.
- [15] Jensen, Richard, and Qiang Shen. *Computational intelligence and feature selection: rough and fuzzy approaches*. Vol. 8. John Wiley & Sons, 2008.
- [16] Kolukisa, Burak, et al. "Diagnosis of Coronary Heart Disease via Classification Algorithms and a New Feature Selection Methodology." *International Journal of Data Mining Science* 1.1 (2019): 8-15.