Makine Öğrenimi Algoritmaları ile Kardiyovasküler Hastalıkların Tespiti

Rifat KURBAN
Kayseri Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, Kayseri
rifatkurban@kayseri.edu.tr

Özet: Kalp hastalıkları ölüm nedenleri arasında en üst sıradadır ve Dünya Sağlık Örgütü'ne göre her yıl yaklaşık 17,9 milyon insan kardiyovasküler hastalıklardan dolayı vefat etmektedir [1]. Hastalıkların erken teşhisi ile birçok ölüm engellenebilir durumdadır. Ayrıca, erken teşhis ile hasta bakım maliyetleri de düşmektedir [2]. Makine öğrenimi yöntemleri ile veriler arasındaki ilişkiler belirlenerek genelleştirilmiş hastalık sınıflandırma modelleri oluşturulabilmektedir [3]. Pek çok ülkede kalp hekimi ve uzmanı eksikliği bulunmaktadır ve hasta kayıtları bilgisayar ortamına doğru kaydedilememektedir [4]. Hasta verileri ile klinik karar verme süreçlerini yapay zekâ yardımı ile modellenerek erken-aşama kalp hastalıkları tahmin edilebilir. Bu çalışmada, Kaggle platformu üzerinden elde edilen ve çeşitli hastanelerden elde edilmiş ve düzenlenmiş durumda olan 1025 hastaya ait veriler kullanılarak kalp rahatsızlığının makine öğrenimi yöntemleri ile tespiti gerçekleştirilmiştir. Veri setinde yaş, cinsiyet, ağrı tipi, tansiyon, kolesterol, kan şekeri vb. toplam 13 parametre bulunmaktadır [5]. Tüm veriler için uzman hekimlerce kalp hastalığının varlığı veya yokluğu veri seti içinde belirlenmiş haldedir. Danışmanlı makine öğrenimi yöntemlerinden karar ağaçları, diskriminant analizi, lojistik regresyon (LR), Naive Bayes (NB), karar destek makineleri (KDM), k-enyakın komşu, kollektif öğrenme ve yapay sinir ağları (YSA) yöntemleri ile veri seti kullanılarak sınıflandırma modelleri oluşturulmuştur. Veriler 5-kat çapraz doğrulama yöntemi ile her bir seferde verinin %80'i eğitim ve %20'si test verisi olarak kullanılmak suretiyle deneyler gerçekleştirilmiştir. Deneyler sonucunda test verilerinde %99.7 sınıflandırma doğruluğu ile YSA yöntemi en iyi sonucu vermiştir. Kolektif öğrenme ile %99.5, k-enyakın komşu ve KDM %99.4, karar ağacı yöntemi %96.4, LR yöntemi %85.1, diskriminant analizi %84.5 ve NB yöntemi ise %82.6 sınıflandırma doğruluğunu elde etmişlerdir. Deneyler sonucunda neredeyse hatasız bir şekilde eldeki verilerden genelleştirilmiş bir kalp hastalığı teşhis yöntemi geliştirilmiştir. Geliştirilen model bu haliyle kardiyovasküler rahatsızlıkların tespitinde uzman hekimlere yardımcı olacak bir yazılım aracı konumundadır.

Anahtar Kelimeler: kardiyovasküler hastalıklar, makine öğrenimi, sınıflandırma

Prediction of Cardiovascular Diseases with Machine Learning Algorithms

Abstract: Heart diseases are the top causes of death, and according to the World Health Organization, approximately 17.9 million people die from cardiovascular diseases each year [1]. Many deaths are preventable with early diagnosis of diseases. In addition, patient care costs decrease with early diagnosis [2]. Generalized disease classification models can be created by determining the relationships among data by using machine learning methods [3]. There is a shortage of heart doctors and specialists in many countries, and patient records cannot be accurately recorded in the computer environment [4]. Early-stage heart diseases can be predicted by modeling patient data and clinical decision-making processes with the help of artificial intelligence. In this study, the detection of heart disease with machine learning methods is carried out using the data of 1025 patients obtained from various hospitals from the Kaggle platform. In the data set there are 13 parameters in total such as age, gender, pain type, blood pressure, cholesterol, blood sugar, etc. [5]. For all data, the presence or

absence of heart disease is determined in the data set by specialist physicians. Classification models using the data set with decision trees, discriminant analysis, logistic regression (LR), Naive Bayes (NB), support vector machines (KDM), k-nearest neighbor methods, ensemble learning, and artificial neural networks (ANN) methods of supervised machine learning methods are created. Experiments are carried out using the 5-fold cross-validation method using 80% of the data as training data and 20% as test data. As a result of the experiments, the ANN method performs the best with 99.7% classification accuracy in the test data. accuracy of ensemble learning method is 99.5%, k-nearest neighbor and KDM are 99.4%, decision tree method is 96.4%, LR method is 85.1%, discriminant analysis is 84.5% and NB method is 82.6%. As a result of the experiments, a generalized method of diagnosing heart disease is developed from the available data, almost without error. The developed model is a software tool that may help specialist physicians in the detection of cardiovascular diseases.

Keywords: cardiovascular diseases, machine learning, classification

Kaynakça:

[1] https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 [Erişim tarihi: 2021].

[2] Lesyuk, W., Kriza, C., & Kolominsky-Rabas, P. (2018). Cost-of-illness studies in heart failure: a systematic review 2004–2016. BMC cardiovascular disorders, 18(1), 1-11.

[3] Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. Computers in Biology and Medicine, 136, 104672.

[4] Tuli, S., Basumatary, N., Gill, S. S., Kahani, M., Arya, R. C., Wander, G. S., & Buyya, R. (2020). HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments. Future Generation Computer Systems, 104, 187-200.

[5] https://www.kaggle.com/johnsmith88/heart-disease-dataset [Erişim: 2021].