



Düzce Üniversitesi Bilim ve Teknoloji Dergisi

Araştırma Makalesi

Parçacık Sürü Optimizasyonu Kullanılarak Boyutu Azaltılmış Mikrodizi Verileri Üzerinde Makine Öğrenmesi Yöntemleri İle Prostat Kanseri Teşhisi

Serhat KILIÇARSLAN ^{a,*}, Kemal ADEM ^a, Onur CÖMERT ^b

^a Gaziosmanpaşa Üniversitesi Enformatik Bölümü Taşlıçiftlik Yerleşkesi, 60250 Tokat / Türkiye

^b Gaziosmanpaşa Üniversitesi Tokat Teknik Bilimler MYO Taşlıçiftlik Yerleşkesi, 60250 Tokat / Türkiye

*Sorumlu yazar, e-posta: serhat.kilicarслан@gop.edu.tr

ÖZET

Mikrodizi verilerine dayanan veri madenciliği analizi, hastalık teşhisi ve farmakoloji alanlarında kullanılmaktadır. Analiz aşamasında yaşanan en önemli zorluk, mikrodizilerin yüksek boyutlu olması ve çok sayıda gereksiz öznitelik içermesidir. Bu nedenle çalışmada kullandığımız prostat kanseri mikrodizi veri kümesi üzerinde öznitelik boyut azaltılması amacıyla Temel bileşenler analizi (TBA) ve Parçacık sürü optimizasyonu (PSO) kullanılmıştır. Bu sayede hastalıkları etkileyen genler tespit edilmektedir. Boyutu azaltılmış veri kümeleri Destek Vektör Makinesi ve k-En Yakın Komşuluk sınıflayıcı yöntemlerine giriş olarak verilmiş ve sınıflandırma başarı sonuçları değerlendirilmiştir. Sonuç olarak PSO boyut azaltma yöntemi ile prostat kanserinde etkin genler belirlenmiş ve 50 öznitelik ile %95.77 başarı elde edilmiştir.

Anahtar Kelimeler: Mikrodizi, Prostat Kanseri, Temel bileşen analizi, Parçacık sürü optimizasyonu.

Prostate Cancer Diagnosis With Machine Learning Methods On Microarray Data Reduced In Dimension Using Particle Swarm Optimization

ABSTRACT

Data mining analysis based on microarray data is used in disease diagnosis and pharmacology. The major challenge in the analysis phase is the high dimension of microarrays and the large number of unnecessary features. For this reason, Principle Component Analysis (PCA) and Particle Swarm Optimization (PSO) were used to reduce the feature dimension on the prostate cancer microarray dataset used in the study. In this way, genes that affect diseases are determined. Dimension reduced data sets are given as input to Support Vector Machine and k-Nearest neighbor classification methods and classification success results are evaluated. Finally, active genes in prostate cancer were identified by PSO dimension reduction method and 95.77% success was achieved with 50 attributes.

Keywords: Microarray, Prostate cancer, Principal Component Analysis, Particle Swarm Optimization.