GRADUATE SCHOOL OF ENGINEERING & SCIENCE
AGÜ

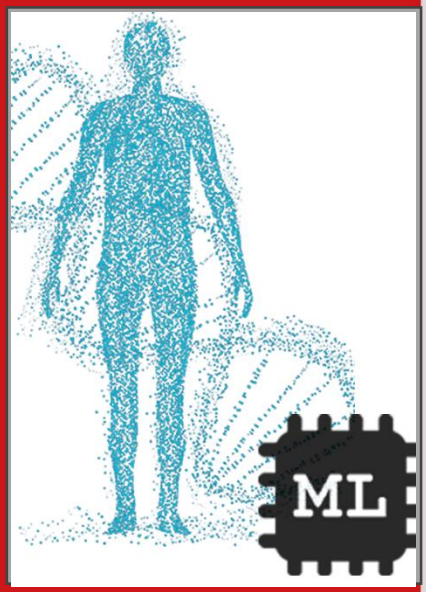
Electrical and Computer Engineering

A Ph.D. Thesis

Yunus Emre IŞIK

yeisik@cumhuriyet.edu.tr

0000-0001-6176-7545



Thesis Advisor

Zafer AYDIN

zafer.aydin@aguu.edu.tr

27.03.2024

Machine Learning Methods For Detecting Genetic And Infectious Diseases

abstract Completion of the whole human genome in the 2003 has led to various advances in many fields, particularly in biology, genetics, health sciences, treatment, and pharmacology. In the following years, spread of faster and cheaper sequencing technologies has enabled us to extract and analyze genetic profiles of individuals digitally. Consequently, individual-specific forecasting and personalized treatment and precision medicine-, what once seemed like science fiction, have become more and more real. In both approaches, one of the crucial steps is identifying the presence of diseases using individual-specific genetic data. This thesis aims to comprehensively and comparatively evaluate the predictive performance of machine learning methods for Behçet's disease and respiratory infections. Additionally, feature selection methods were employed to identify the genetic factors (such as SNPs and genes) associated with disease presence for both diseases. Furthermore, the usability of selected features depending on biological pathway-driven active subnetworks listed in the literature was analyzed for the prediction of Behçet's disease. For the respiratory infection prediction problem, on the other hand, the prediction performance of features calculated by single-sample gene set enrichment analysis (ssGSEA) was evaluated using different machine learning methods. As the data types used in both experiments were different (genome-wide association studies data, gene expression profiles), the performance of machine learning approaches on different data types was also observed. It is hoped that the findings of both experiments will contribute to future machine learning based disease prediction studies.

keywords Disease prediction, Machine Learning, Behçet's Disease Prediction, Respiratory Infection Prediction, Feature Selection and Representation

özet 2003 yılında insan genomunun tamamen dizilenebilmesi, başta biyoloji ve genetik bilimi olmak üzere sağlık bilim, tedavi ve farmakoloji gibi birçok farklı alanda yeni gelişmelerin ortaya çıkmasına neden olmuştur. İlerleyen yıllarda hızlı ve daha ucuz dizileme teknolojilerinin yaygınlaşmasıyla bireylerin genetik profillerinin çıkartılarak dijital ortamda işlenebilmesi mümkün hale gelmiştir. Böylelikle eski zamanlarda bilim-kurgu gibi görünen, bireylere özgü tahmin ve tedavi belirlenmesi, başka bir deyişle kişileştirilmiş ve hassas tıp yaklaşımı hız kazanmıştır. Her iki yaklaşımda da en önemli aşama ise hastalığın bireye özgü genetik veriler kullanılarak belirlenmesidir. Bu tez çalışması Behçet hastalığı ve solunum yolu enfeksiyonu olmak üzere iki farklı türde hastalık için makine öğrenmesi yöntemlerinin tahmin performansını kapsamlı ve karşılaştırmalı olarak değerlendirmeyi amaçlamaktadır. Ayrıca öznitelik seçme yöntemleriyle hastalık tahmininde önemli rol oynayan genetik faktörler (SNP, Gene) her iki hastalık içinde ayrı ayrı belirlenmeye çalışılmıştır. Bunun yanı sıra Behçet hastalığının tahminlenmesinde literatürde yer alan biyolojik yolak temelli aktif-ağlar kullanılarak seçilen özniteliklerin kullanılabilirliği analiz edilmiştir. Öte yandan solunum yolu enfeksiyon tahmin probleminde ise, örneklem bazında uygulanan gen seti zenginleştirme analizi sonrası elde edilen skorların, örneklemlerin temsil edilmesinde ne kadar başarılı olduğu makine öğrenmesi kullanılarak ortaya koyulmuştur. Her iki deneyde kullanılan veri tipleri de farklı olduğu için (genom çapında ilişkilendirme çalışmaları verisi, gen ifadesi profilleri), makine öğrenmesi yaklaşımlarının farklı veri türlerindeki performansları da gözlemlenmiştir. Her iki deney sonucunda elde edilen çıktıların makine öğrenmesi temelli hastalık tahminleme çalışmalarına katkı sağlayacağı umulmaktadır

anahtar kelime Hastalık tespiti, Makine Öğrenmesi, Behçet Hastalığı Tahmini, Solunum Yolu Enfeksiyon Tahmini, Öznitelik Seçimi ve Temsili