

İnfeksiyon Hastalıklarının Prognozunu Öngörmede Yapay Zekâ Uygulamaları (22-Haziran-2025)

Prof. Dr. Gökmen ZARARSIZ

Erciyes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kayseri

HİBRİT

Klimik Derneği Asistan ve Genç
Uzman Hekimler (AGUH) Komisyonu

TEMEL EĞİTİM PROGRAMI

AGUH 2025

21-22 HAZİRAN 2025

ANKARA ÜNİVERSİTESİ TIP FAKÜLTESİ
İBNİ SİNA HASTANESİ HASAN ALİ YÜCEL
KONFERANS SALONU /ANKARA

.....

Biyografi

- 2009-2012, MSc | Erciyes Üniv, Biyoistatistik A.D.

Tıpta yapay zeka, mikroarray verilerinin analizi, yazılım geliştirme

- 2012-2015, PhD | Hacettepe Üniv, Biyoistatistik A.D.

Development and Application of Novel Machine Learning Approaches for RNA-Seq Data Classification

- 2013, Misafir Araştırmacı | Beijing Genomics Institute (BGI)

Yeni-nesil dizileme teknolojisi ve biyoinformatik yöntemleri

- 2019-2020, Biyoistatistik Uzmanı | Roche Diagnostics GmbH, Almanya

Bireysel tıp, klinik ve omiks verilerin entegrasyonu, nörodejeneratif hastalıkların in vitro tanısı

- 2020-.., Hassas Tıp ve İleri Analitik Araştırma Grubu Lideri | Erciyes Üniv

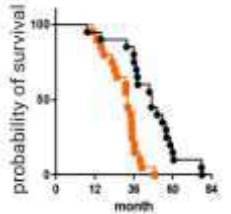
Hassas tıp, tanısal ve prognostik modelleme, metabolomik, multi-omik modeller



pmaa

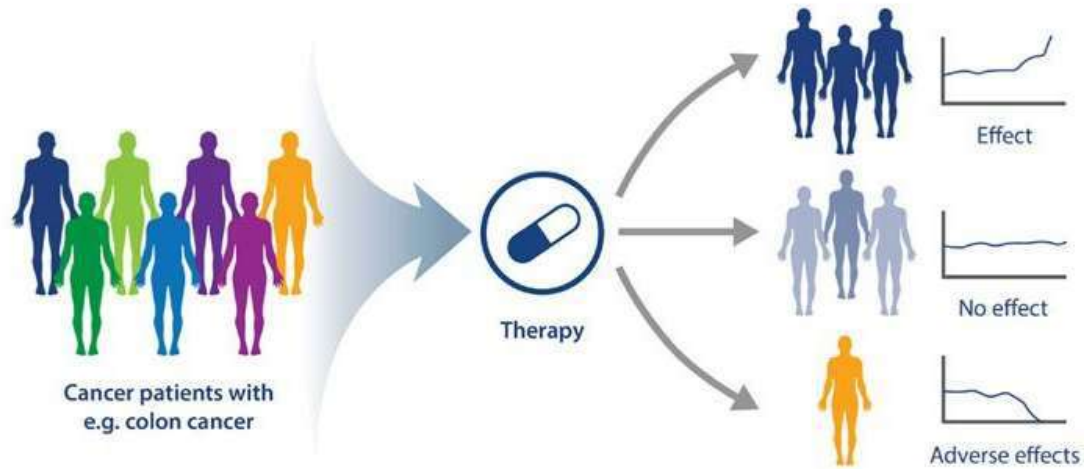
Diyagnostik, Prognostik ve Prediktif Tıp

	AMAÇ	VERİ TİPİ	MODELLEME	ÖRNEK
DİYAGNOSTİK TIP	Kişide bir hastalık var mı yok mu?	Semptomlar, fizik muayene, laboratuvar testleri, görüntüleme	Sınıflandırma modelleri (örn. COVID-19 (+) vs (-)) ROC analizi, duyarlılık, seçicilik, vs.	CRP + akciğer BT ile COVID-19 tanısı
PROGNOSTİK TIP	Hastalık gelişimi, seyri veya sonucu hakkında öngöründe bulunur (hastalık zaten vardır, nasıl ilerleyecek?)	Tanı konmuş hastalara ait klinik/omiks veriler	Sınıflandırma, sağkalım analizi (örn. ICU ihtiyacı, ölüm riski) ROC eğrisi, AUC, Kaplan-Meier, yapay zeka yaklaşımları	Sepsis tanısı konmuş bir hastada 7 gün içinde organ yetmezliği riski
PREDİKTİF TIP	Belli bir tedaviye hastanın nasıl yanıt vereceğini öngörür (kişiye özgü tedavi planlamasında kritik)	Genetik, metabolomik, farmakogenomik, klinik veriler + tedavi bilgisi	Klinik araştırmalar, biyoinformatik ve yapay zeka yaklaşımları	Karbapenem grubu antibiyotiklerden kim fayda görür, kimde direnç gelişme riski yüksektir?

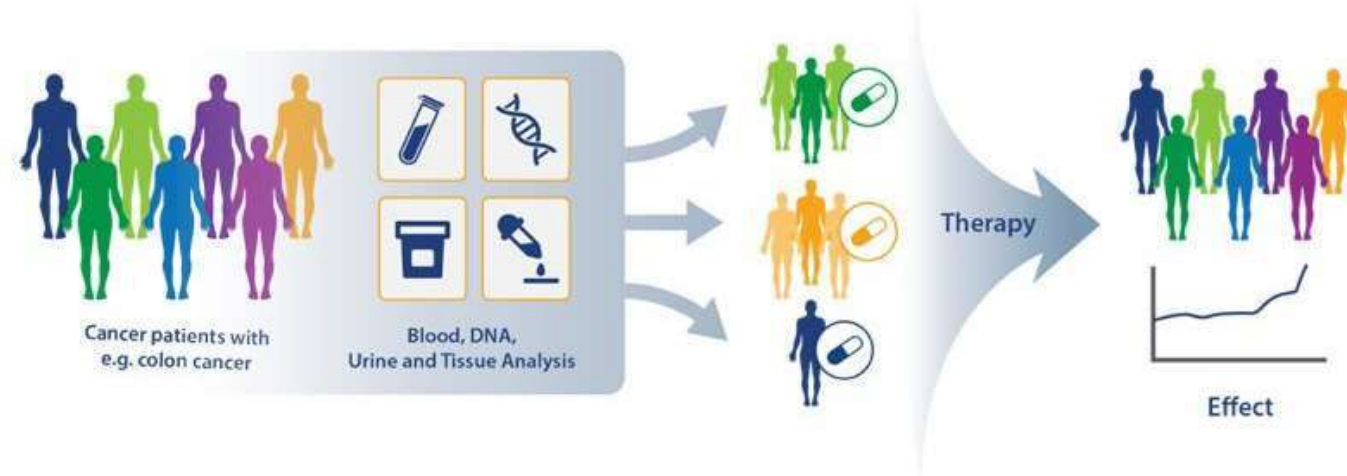


Hassas Tıp

Traditional Medicine One Treatment Fits All



Precision Medicine More Personalized Diagnostics



İnfeksiyon Hastalıklarında Hassas Tıp

- Precision epidemiology: Precision medicine in epidemiology*

Precision medicine

Precision oncology

Precision hematology

Precision nephrology

Precision infectious disease

Precision radiology

PERSPECTIVE

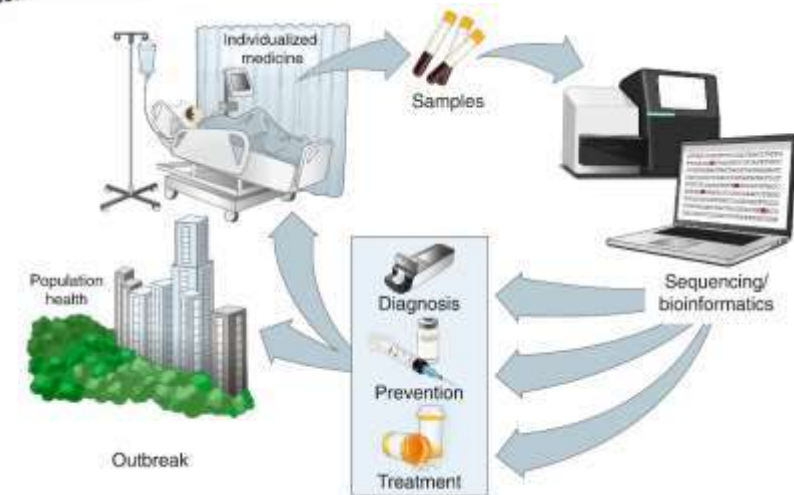
<https://doi.org/10.1038/s41591-019-0345-2>

nature
medicine

Precision epidemiology for infectious disease control

Jason T. Ladner¹, Nathan D. Grubaugh², Oliver G. Pybus³ and Kristian G. Andersen^{4,5*}

Advances in genomics and computing are transforming the capacity for the characterization of biological systems, and researchers are now poised for a precision-focused transformation in the way they prepare for, and respond to, infectious diseases. This includes the use of genome-based approaches to inform molecular diagnosis and individual-level treatment regimens. In addition, advances in the speed and granularity of pathogen genome generation have improved the capability to track and understand pathogen transmission, leading to potential improvements in the design and implementation of population-level public health interventions. In this Perspective, we outline several trends that are driving the development of precision epidemiology of infectious disease and their implications for scientists' ability to respond to outbreaks.



İnfeksiyon Hastalıklarında Prognoz Öngörüsü

- Covid-19 hastalarında deksametazon uygulaması sonrası hastalığın prognostik seyri nasıl değişir? Hangi hastalarda ölüm riskini azaltır?

Girdi: İnflamatuvar belirteçler (IL-6, ferritin, CRP), oksijen ihtiyacı, -omiks veriler

Çıktı: Mortalite, entübasyon ihtiyacı

- Kırım-Kongo Kanamalı Ateşi hastalığında ribavirin hangi hastalarda daha iyi yanıt verebilir?

Girdi: Klinik başvuru bilgileri, laboratuvar verileri, genomik & metabolomik verileri

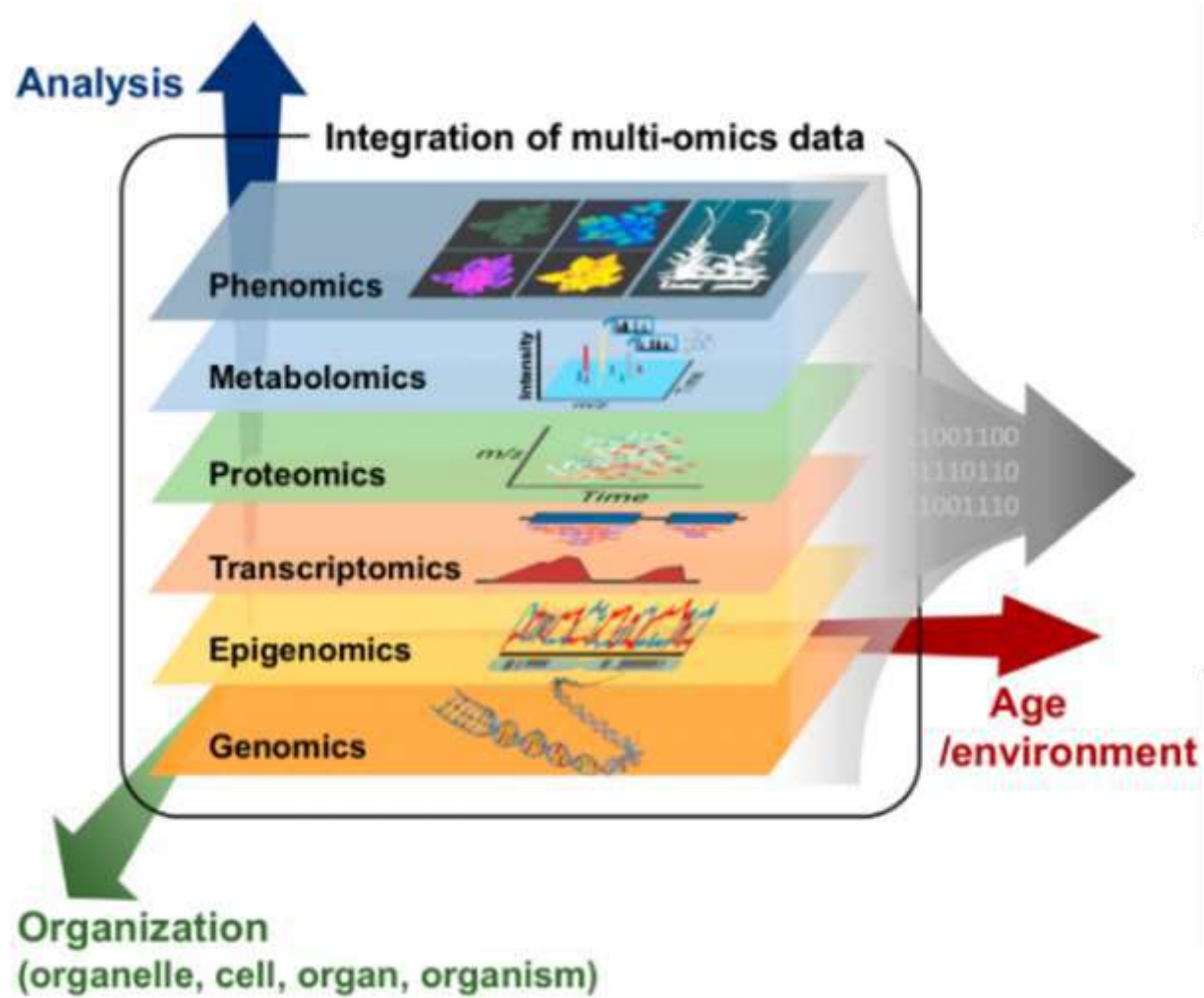
Çıktı: Yoğun bakım yatış, mortalite

- HBV tedavisinde entekavir mi, tenofovir mi? Hangi hastalarda virolojik yanıt daha iyi?

Girdi: HBV DNA seviyesi, karaciğer fonksiyon testleri, genotip

Çıktı: 48 haftada virolojik yanıt

Omiks Bilimleri



Kaynak: Kim ve ark. (2016). Toward Systems Understanding of Leaf Senescence: An Integrated Multi-Omics Perspective on Leaf Senescence Research. Mol Plant; 9(6):813-25.

İnfeksiyon Hastalıklarında Transkriptomik Analizler

- İnfeksiyon hastalıkların moleküler mekanizmasının anlaşılması ve mücadeleye yönelik yeni stratejilerin geliştirilmesi
- Hücre veya dokudan elde edilen örneklerden tüm RNA moleküllerin kümesi olan transkriptom
- Transkriptom teknolojileri: mikroarray, RNA-dizileme ve rtPCR
- Konak-patojen etkileşimleri sırasında oluşan gen ifadesindeki dinamik değişikliklerin tespit edilmesi
- Konak bağışıklık tepkisi ve patojen virölans faktörlerinin araştırılması
- Prognoz ve tedavi yanıtı için biyobelirteçler, ilaç keşfi ve aşı geliştirme için potansiyel hedefler

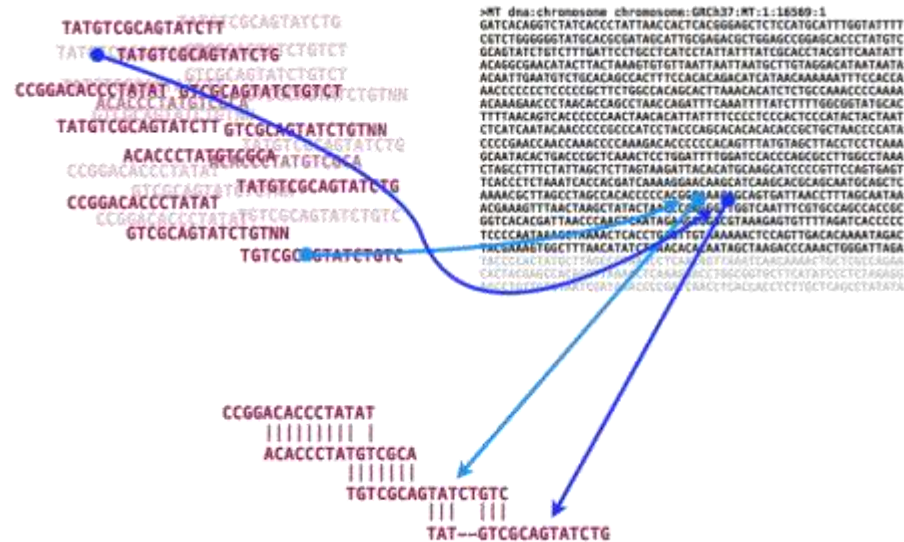
İnfeksiyon Hastalıklarında Metabolomik Analizler

- Metabolomik: Fenotiple en yakından ilişkili –omiks yöntemi
- İnfeksiyon hastalıklarının patofizyolojisinin daha iyi anlaşılması ve yeni tanı/prognoz stratejilerinin geliştirilmesi
- Konak-patojen etkileşimleri sırasında meydana gelen metabolik yanıtların tespiti
- Konak immün yanıtı, enerji metabolizması, oksidatif stres gibi süreçlerin dinamik izlenmesi
- Tanısal ve prognostik biyobelirteçlerin belirlenmesi, kişiselleştirilmiş tedavi hedeflerinin keşfi
- Yeni ilaç hedefleri ve aşı adjuvanları geliştirme süreçlerine katkı

Omiks Verileri ve Analizleri

miRNA	Non-tumor samples					Tumor samples					Total
	Sample-1	Sample-2	Sample-3	...	Sample-29	Sample-30	Sample-31	Sample-32	...	Sample-58	
let-7a	865	810	5,505	...	38	3,343	4,990	5,193	...	1,422	284,257
let-7g	447	173	1,922	...	126	737	4,141	3,760	...	2,081	128,348
miR-125b	1,038	5,007	2,595	...	106	381	1,463	201	...	1	337,394
miR-18a	5	4	10	...	0	10	9	56	...	0	634
miR-29a	320	447	904	...	4	413	4,619	1,398	...	1	134,382
miR-490-5p	0	1	0	...	0	0	0	0	...	0	19
miR-874	2	4	9	...	0	4	0	3	...	0	509
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
miR-93	10	18	126	...	0	36	133	211	...	0	7,028
miR-99b	24	92	97	...	177	36	62	19	...	20	11,718
Total	22,449	39,798	71,717	...	8,034	58,362	431,247	84,850	...	16,338	13,701,148

In this matrix, the rows indicate genes, where the columns indicate samples. The numbers in each cell are the mapped sequence read counts (to a reference genome or transcriptome) of g^i gene for s^j sample. First column of this table describe the names of genes, first row describe the names of classes and second row describe the name of samples. The numbers in last column are the total number of counts for g^i gene. The numbers in last row are the library sizes (i.e. the total number of counts for s^j sample).



INTEGRATIVE MACHINE-LEARNING WORKFLOW

1. DATA SPLITTING

2. OMICS DATA PREPROCESSING

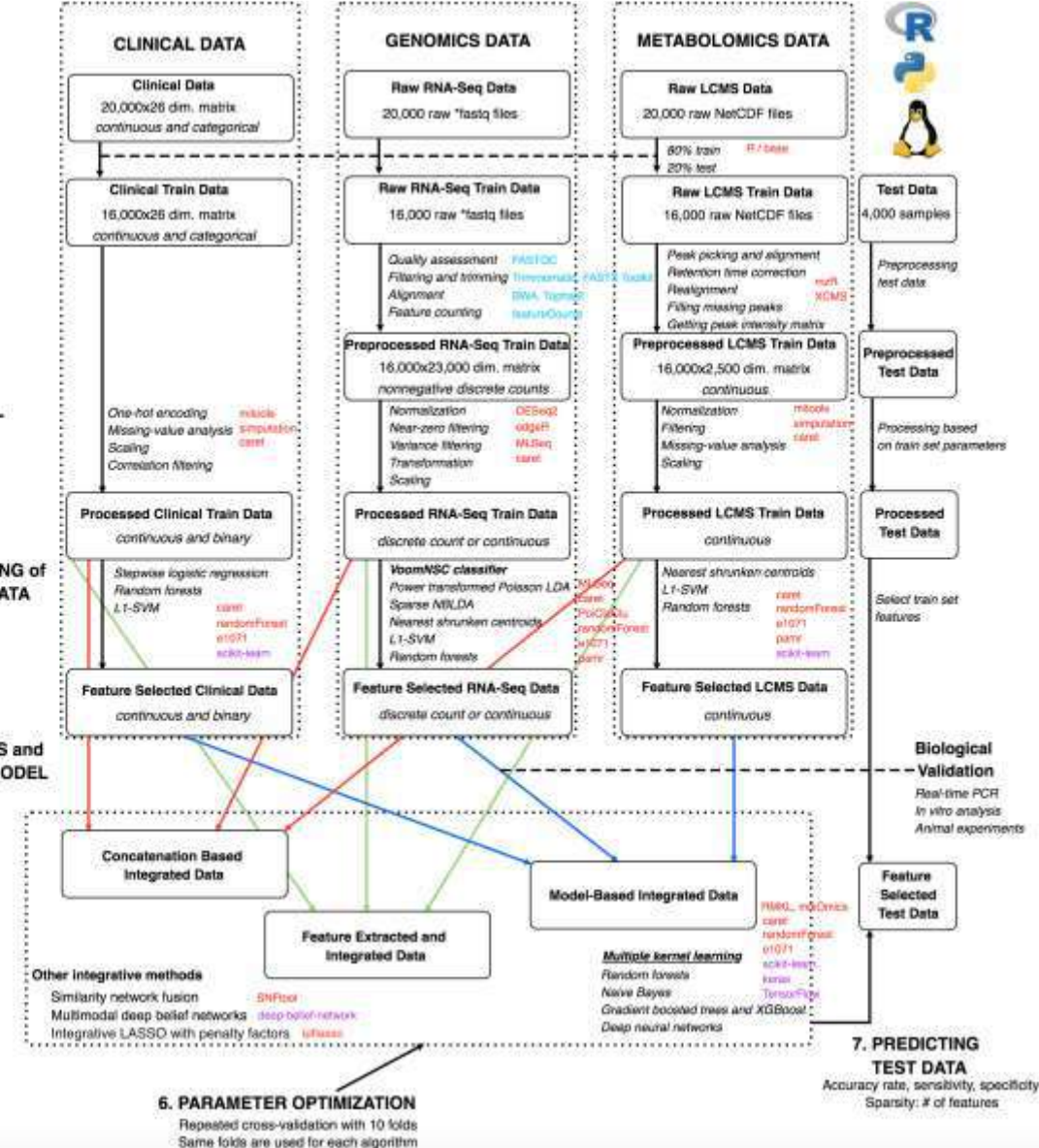
3. PROCESSING CLINICAL and OMICS DATA

4. INDEPENDENT MODELING of CLINICAL and OMICS DATA

5. INTEGRATIVE ANALYSIS and BUILDING THE FINAL MODEL

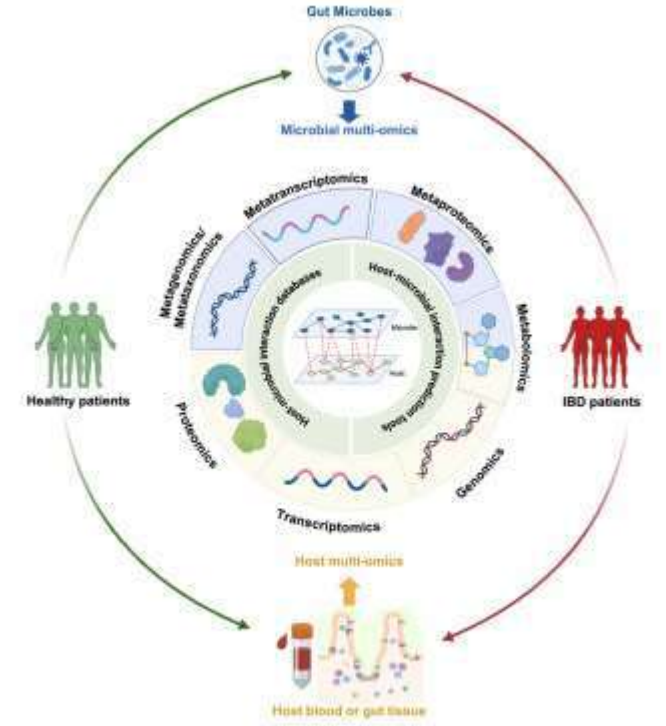
6. PARAMETER OPTIMIZATION

Repeated cross-validation with 10 folds
Same folds are used for each algorithm

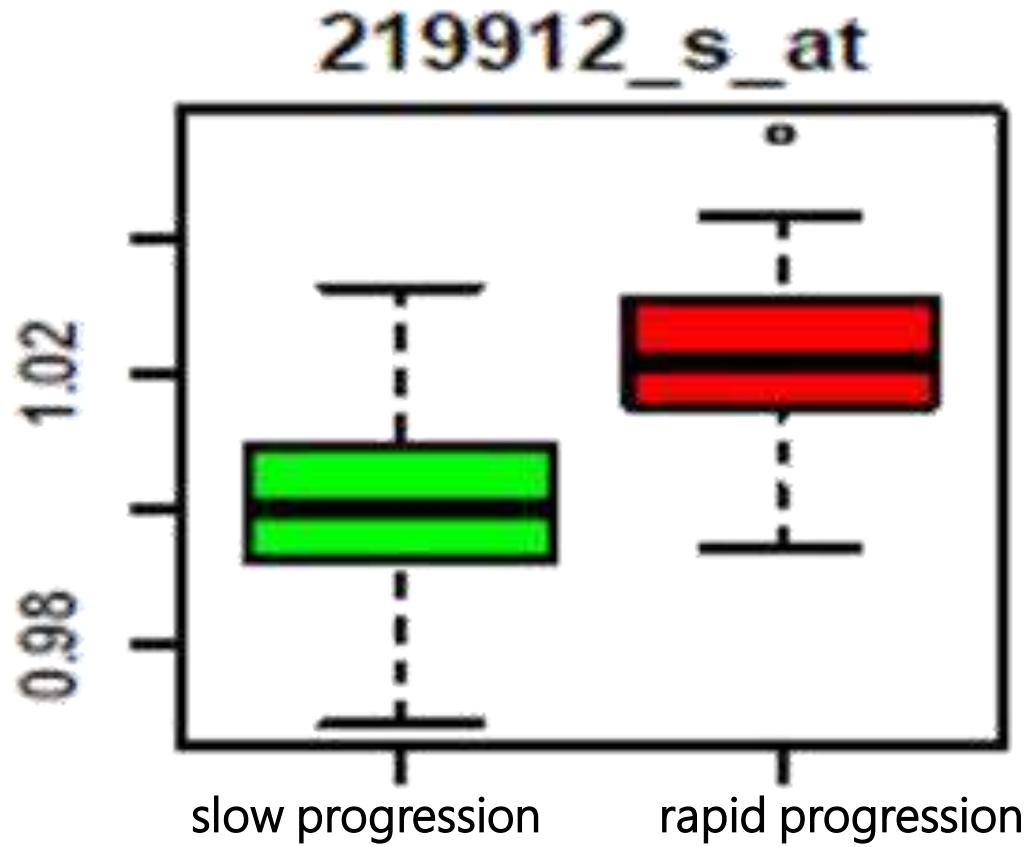


İnfeksiyon Hastalıklarında Prognoz Öngörüsüne Yönelik Veri Analizi Yaklaşımları

- Klinik verilerin biyoistatistik hipotez testleri ile analizlerinin gerçekleştirilmesi ve klinik biyobelirteçlerin belirlenmesi
- Omiks verilerinde anlamlılık analizlerinin yapılması ile prognoz ile ilişkili gen, protein, metabolitlerin ve prognoz üzerinde etkili biyolojik yolların tespit edilmesi
- Kümeleme analizleri ile hastalık alt gruplarının belirlenmesi
- Sağkalım modellemeleri ile hastalığın ilerleyişinin kestirilmesi
- Yapay zeka yaklaşımları ile prognostik omiks biyobelirteçlerinin keşfi
- Multi-omiks yaklaşım ile klinik ve omiks verilerinin entegrasyonu

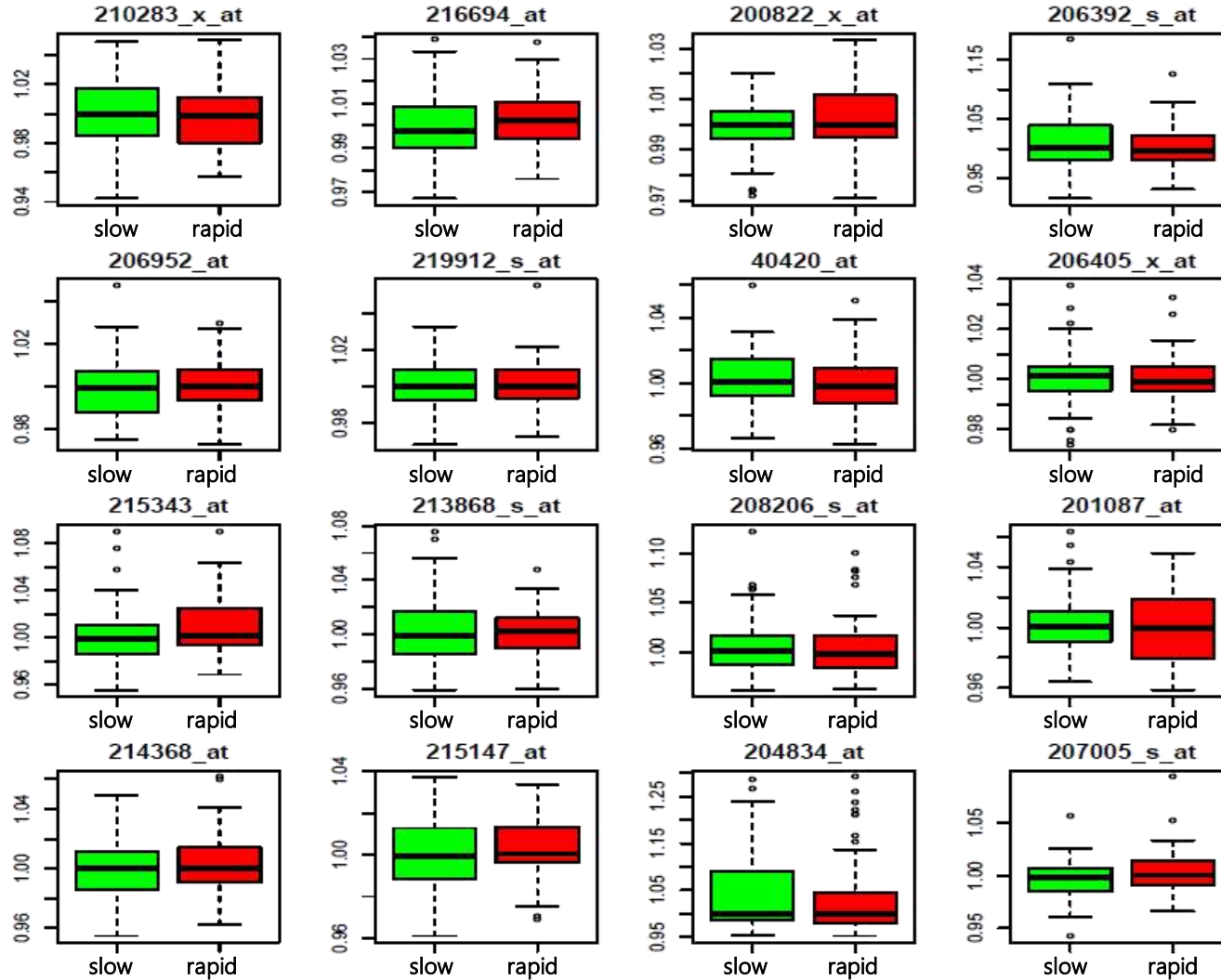


Hastalık Çıktılarının Kestirimi



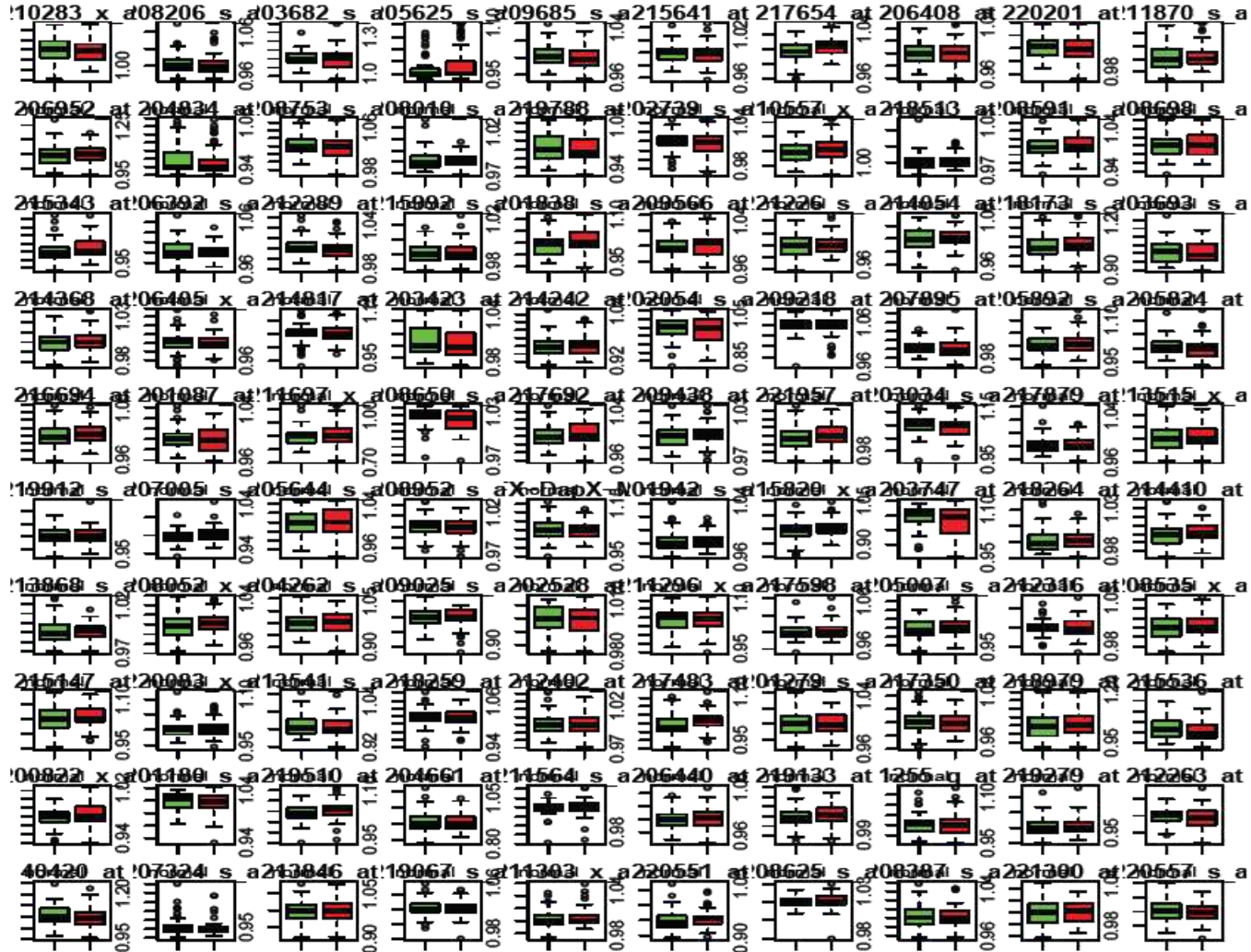
Kaynak: Chen G (2013). Machine Learning for Bioinformatics using R. BGI Bioinformatics Workshop

Hastalık Çıktılarının Kestirimi



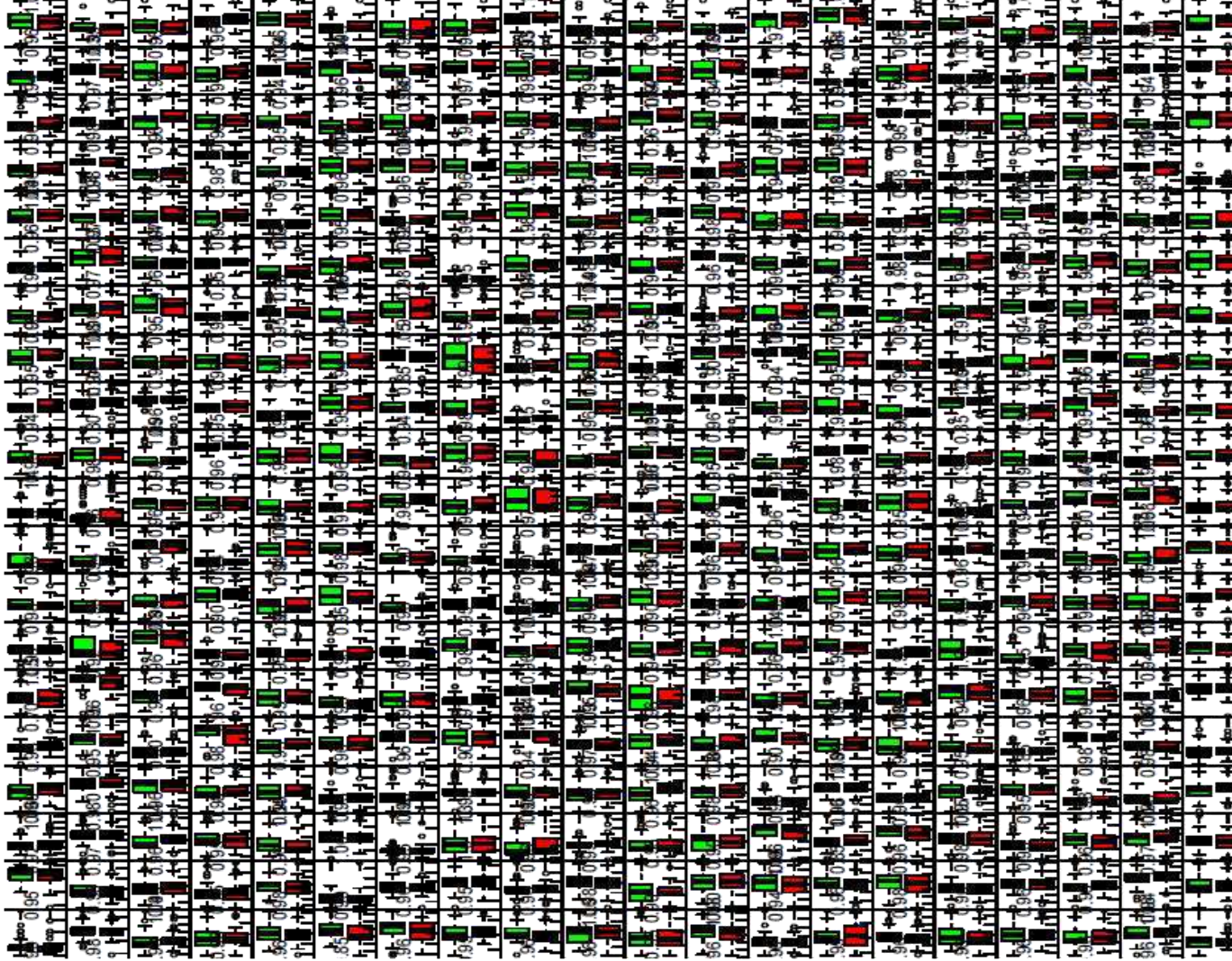
Kaynak: Chen G (2013). Machine Learning for Bioinformatics using R. BGI Bioinformatics Workshop

Hastalık Çıktılarının Kestirimi



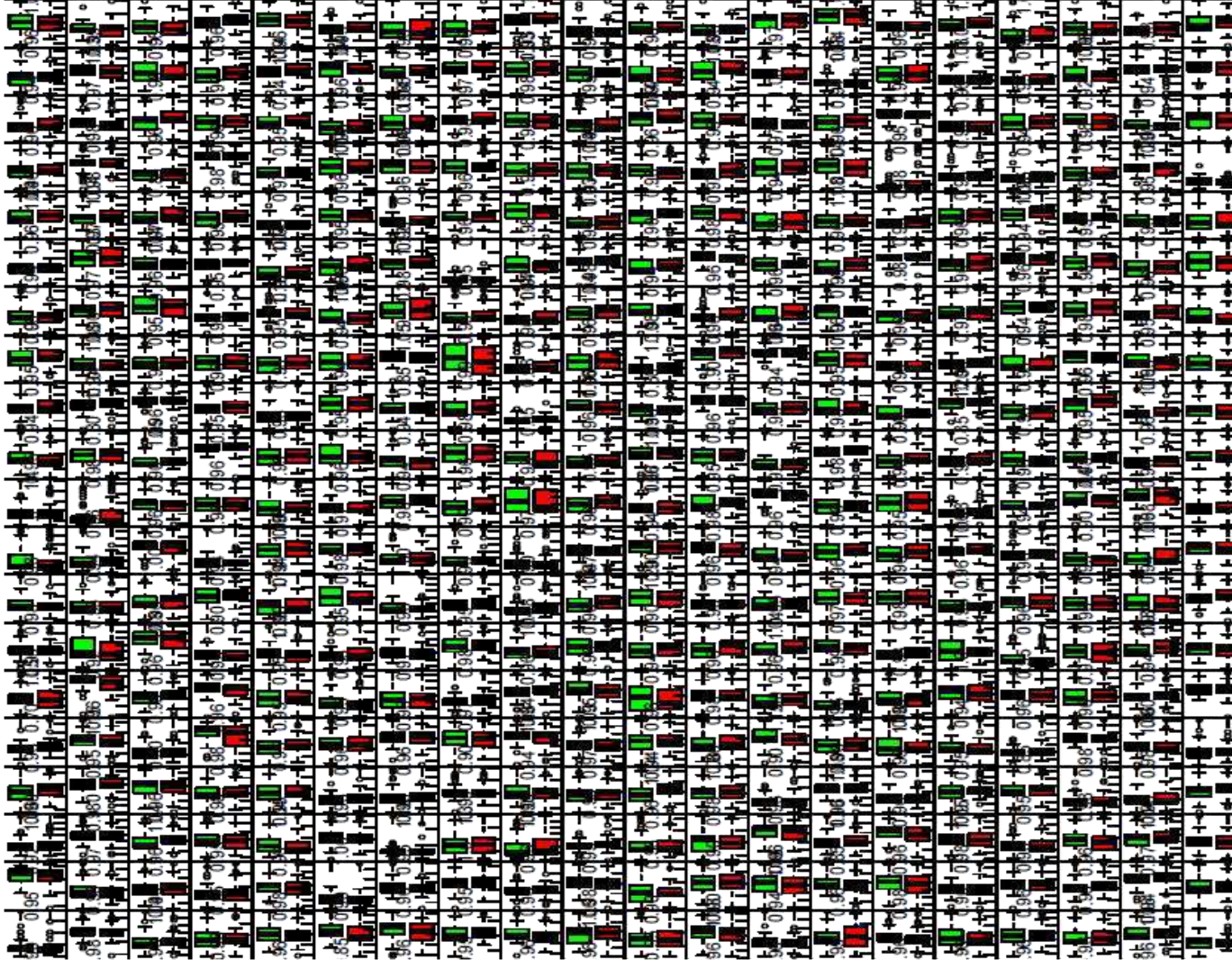
Kaynak: Chen G (2013). Machine Learning for Bioinformatics using R. BGI Bioinformatics Workshop

Hastalık Çıktılarının Kestirimi



Kaynak: Chen G (2013). Machine Learning for Bioinformatics using R. BGI Bioinformatics Workshop

Hastalık Çıktılarının Kestirimi



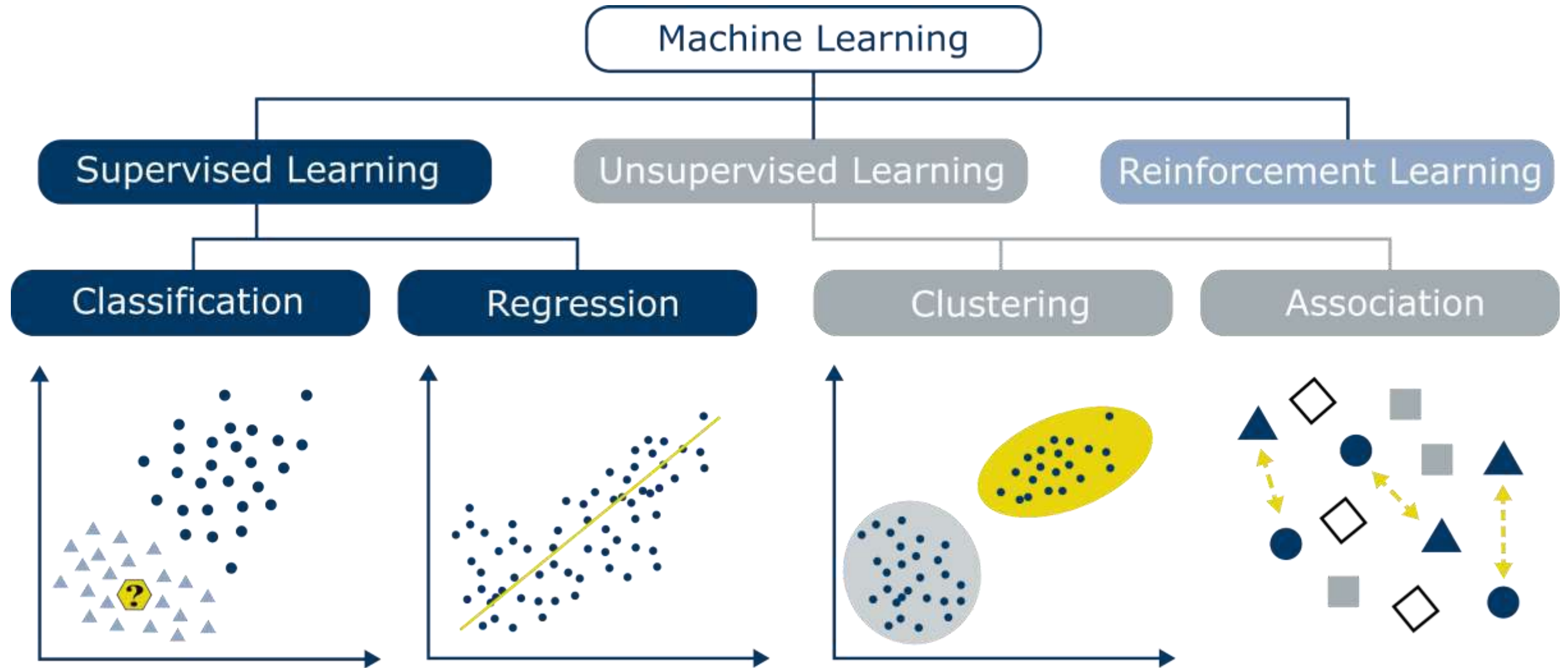
Çözüm

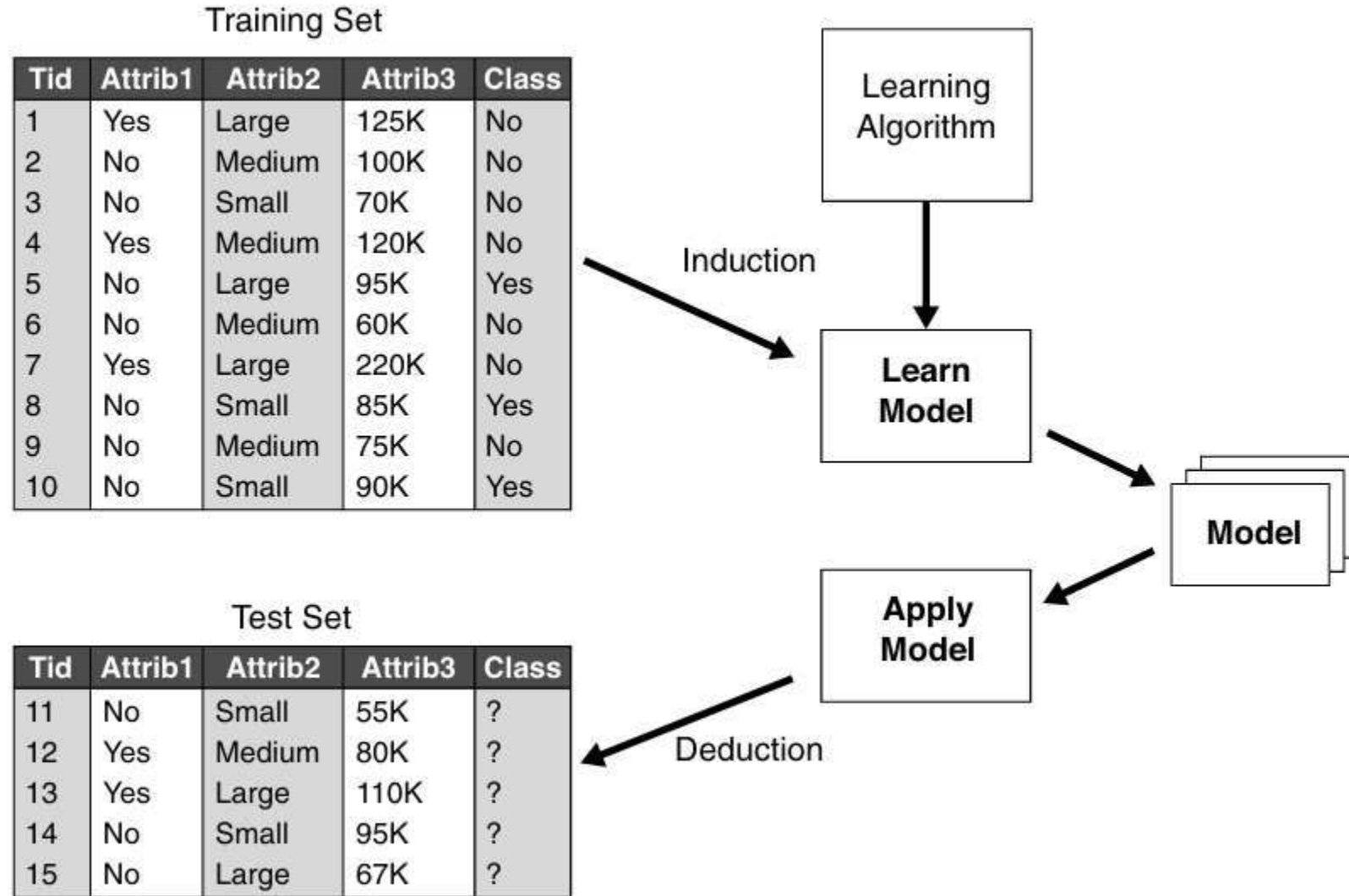


İleri biyoistatistik ve
yapay zeka odaklı
yaklaşımlar

Kaynak: Chen G (2013). Machine Learning for Bioinformatics using R. BGI Bioinformatics Workshop

Yapay Zeka ve Makine Öğrenmesi



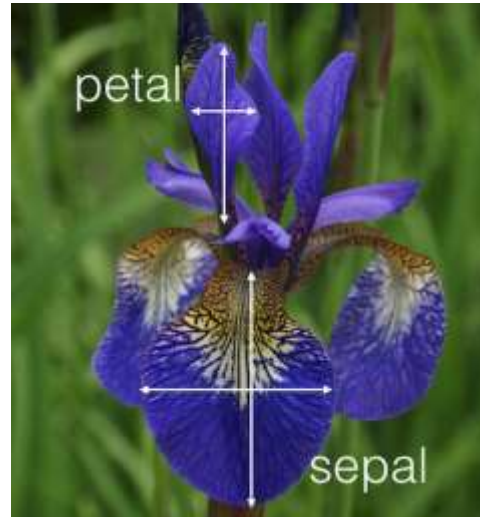




Bu üç zambak türünü ayırt edebilir misiniz?



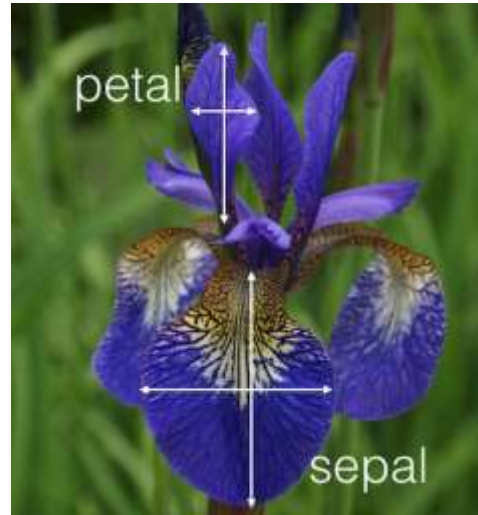
Bu üç zambak türünü ayırt edebilir misiniz?



Yapay Zeka ve Makine Öğrenmesi

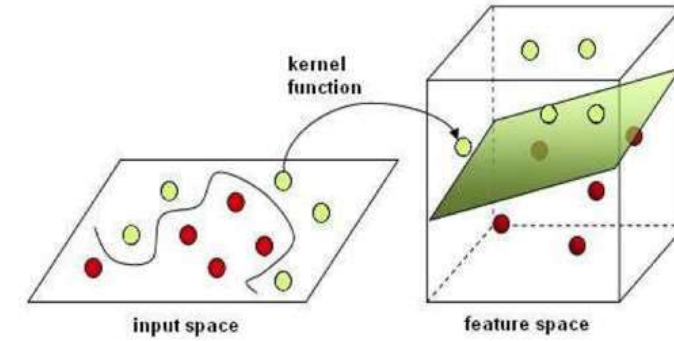
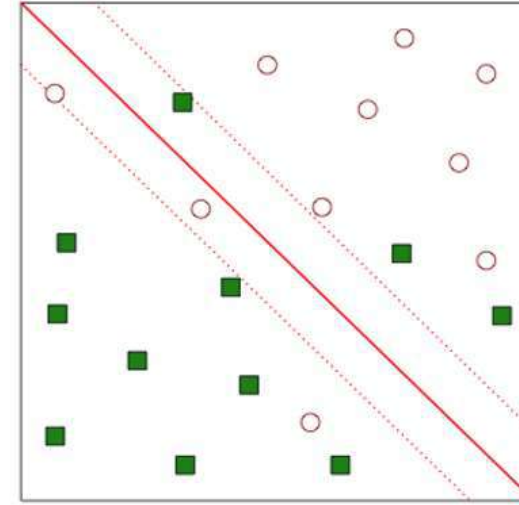
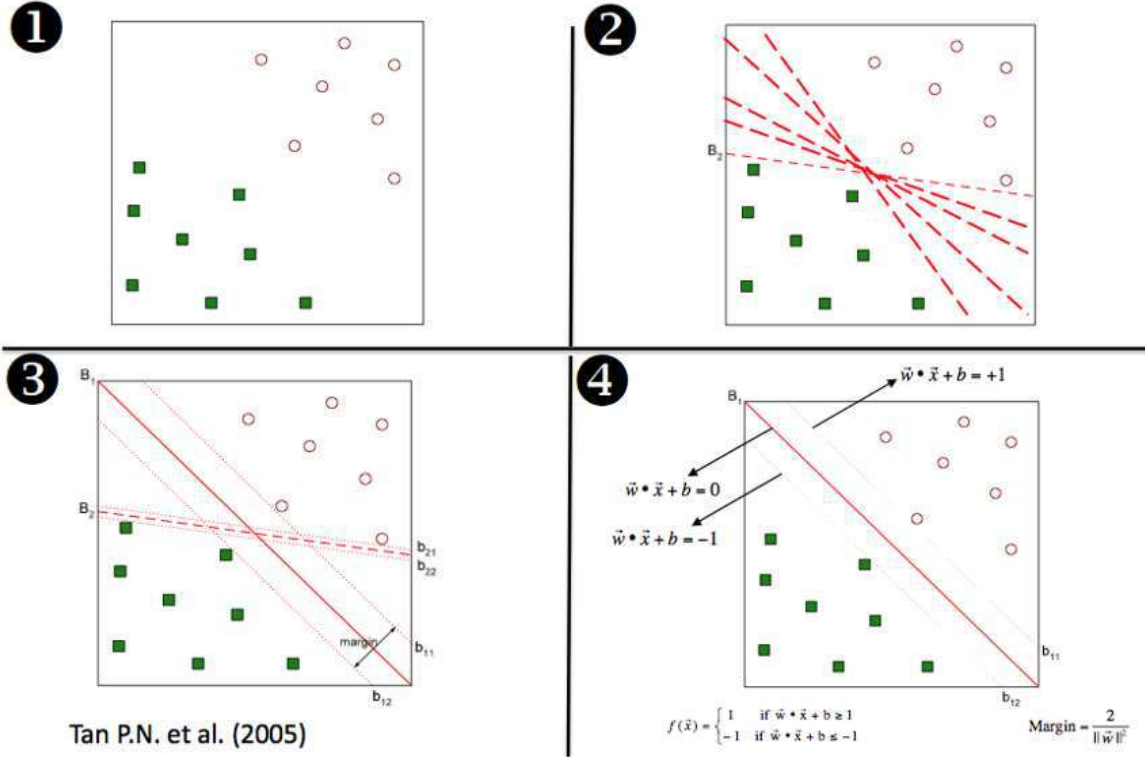


Bu üç zambak türünü ayırt edebilir misiniz?



	sepal.Length	sepal.width	Petal.Length	Petal.width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa

Destek Vektör Makineleri



- For nonlinear problems, slack variables or kernel functions are used

Veri Ön İşleme Süreçlerindeki Zorluklar

- Deneysel süreçlerden kaynaklı oluşan sistematik yanlışlıklar

Omiks verilerine özgü normalleştirme yaklaşımlarının uygulanması

- Veride bilgi içermeyen çok sayıda –omiks değişkeninin yer alması

Bilgi içermeyen bu değişkenlerin kalite kontrol sürecinden geçirilerek filtrelenmesi

- Veride yer alan eksik değerler

Eksik değerlerin rasgeleliğinin araştırılması. MAR & MCAR yaklaşımları. kNN eksik değer ataması

- Yüksek düzeyde korelasyona sahip değişkenlerin varlığı

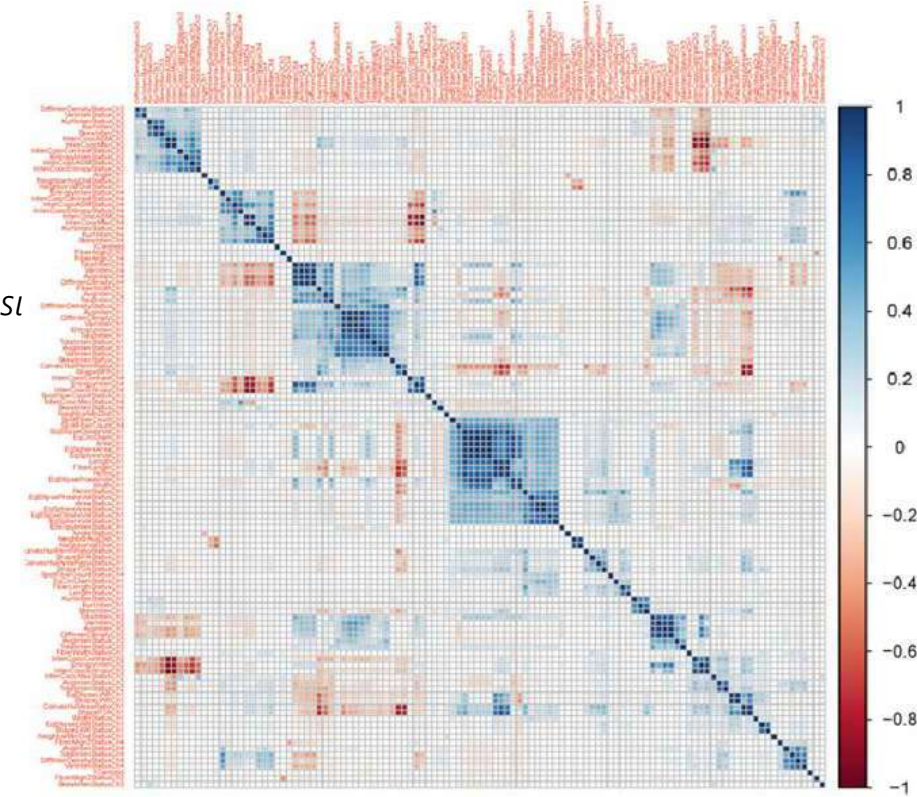
Yüksek korelasyona sahip ve aynı biyolojik anlama sahip değişkenlerde filtreleme

- Veri setindeki aşırı çarpıklıklar

Veri dönüşümü ile verilerin uygun bir dağılıma benzer yapıya getirilmesi

- Prognostik çıktılarda sınıf dengesizliği

Yukarı örnekleme, aşağı örnekleme, hibrit örnekleme (SMOTE, ROSE, vb.) yaklaşımlarının uygulanması



Modelleme Süreçlerindeki Zorluklar - Veri Türü ve Dağılımındaki Farklılıklar

Gee	G1iNS1	G144	G166	G179	CB541	CB660
13CDNA73	3.158	2.659	1.621	1.564	4.269	3.441
A2BP1	2.015	2.049	3.867	2.006	1.698	1.988
A2M	3.655	2.699	1.966	4.526	4.658	5.124
A4GALT	2.598	5.689	0.498	2.557	3.478	2.619
AAAS	4.129	5.254	3.698	2.598	3.647	1.991
AACS	2.549	4.523	4.116	3.987	2.386	3.447
AADACL1	6.251	4.256	3.698	4.251	2.500	2.824
[...]						

Gen	G1iNS1	G144	G166	G179	CB541	CB660
13CDNA73	4	0	6	1	0	5
A2BP1	19	18	20	7	1	5
A2M	2724	2209	13	49	193	548
A4GALT	0	0	48	0	0	0
AAAS	57	29	224	49	202	92
AACS	1904	1294	5073	5364	3737	3511
AADACL1	3	13	239	683	158	40
[...]						

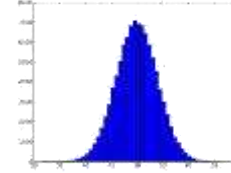
Mikroarray verisi



Sürekli dağılımlar



Normal vb.



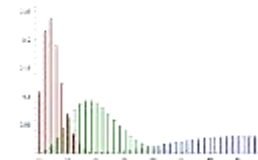
RNA-dizileme sayma verisi



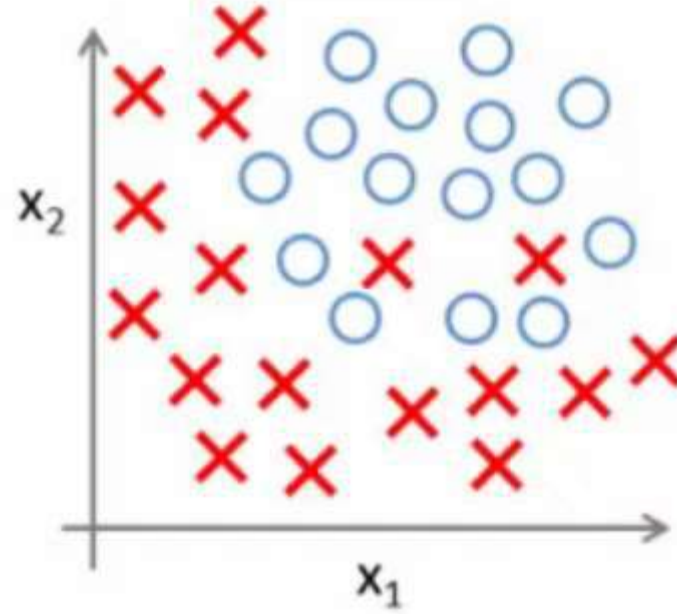
Kesikli dağılımlar



Poisson, negatif binom vb.

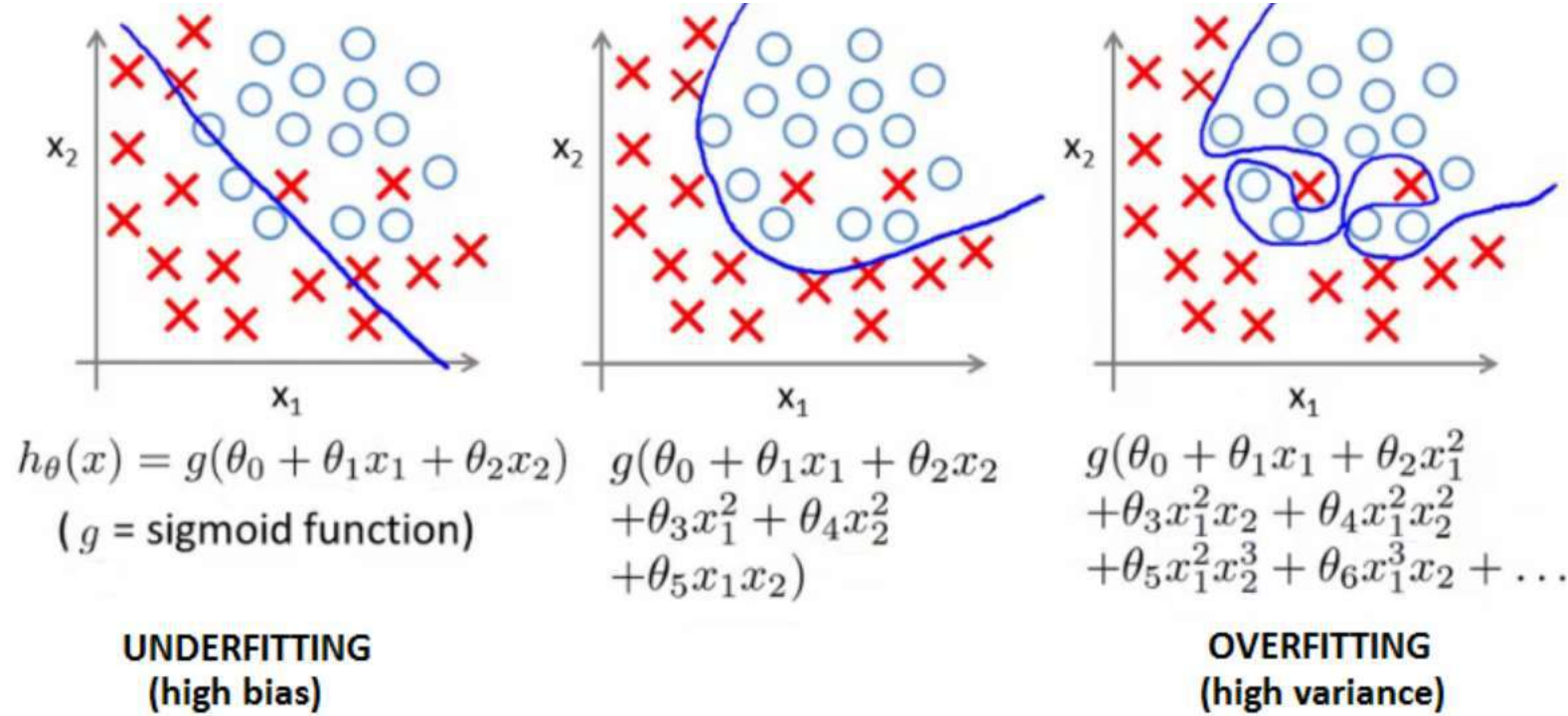


Prognostik Hastalık Gruplarının Sınıflandırılması



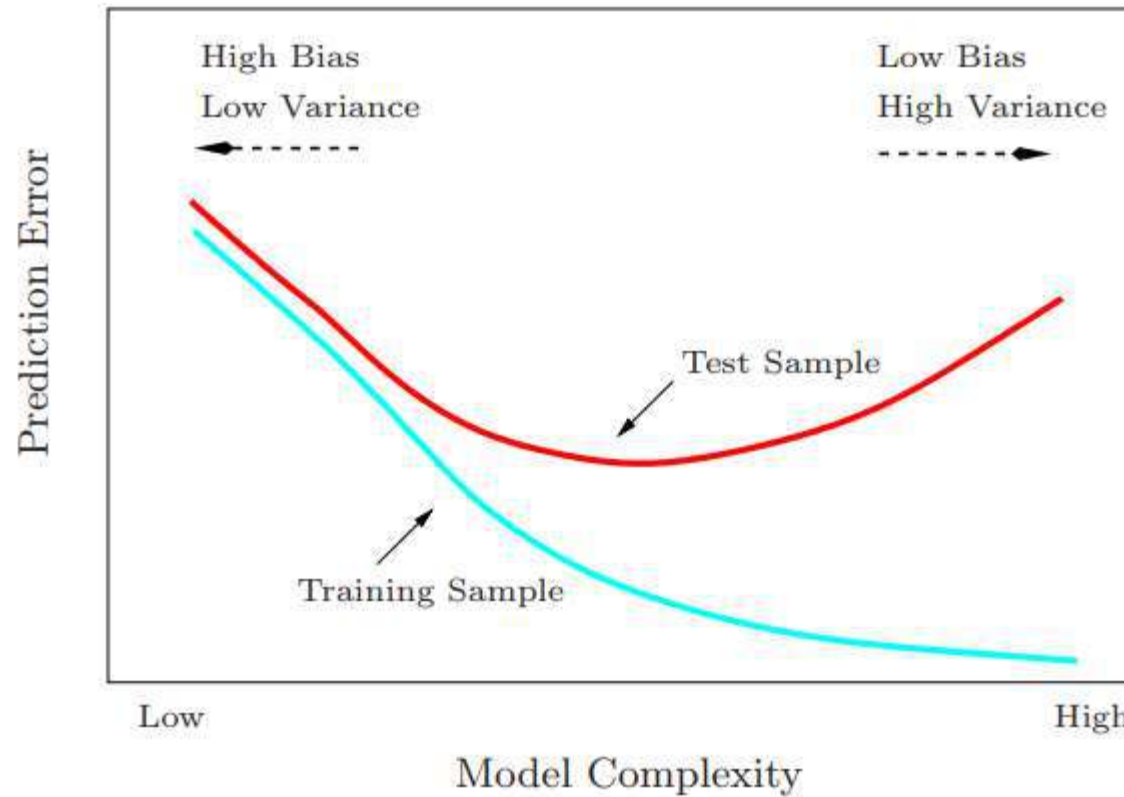
Kaynak: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. *An Introduction to Statistical Learning : with Applications in R*. New York :Springer, 2013

Geliştirilen Modellerde Zayıf Uyum ve Aşırı Uyum Probleminin Giderilmesi



Kaynak: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. *An Introduction to Statistical Learning : with Applications in R*. New York :Springer, 2013

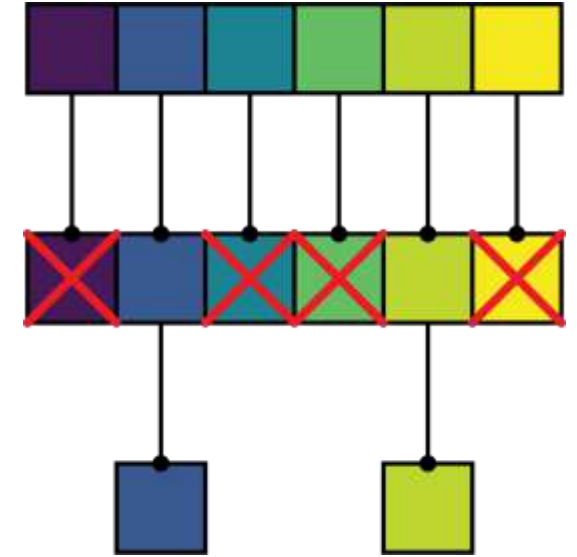
Yanlılık ve Varyans Arasındaki Denge



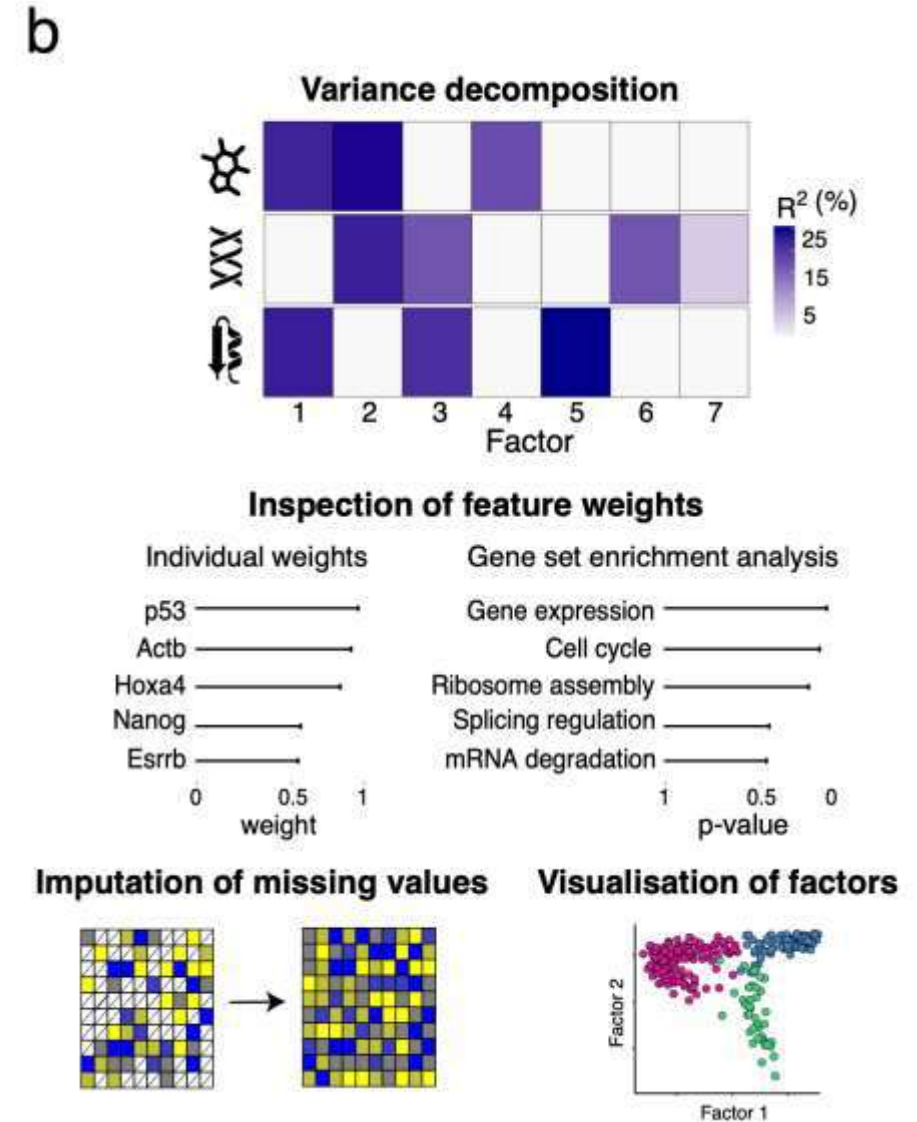
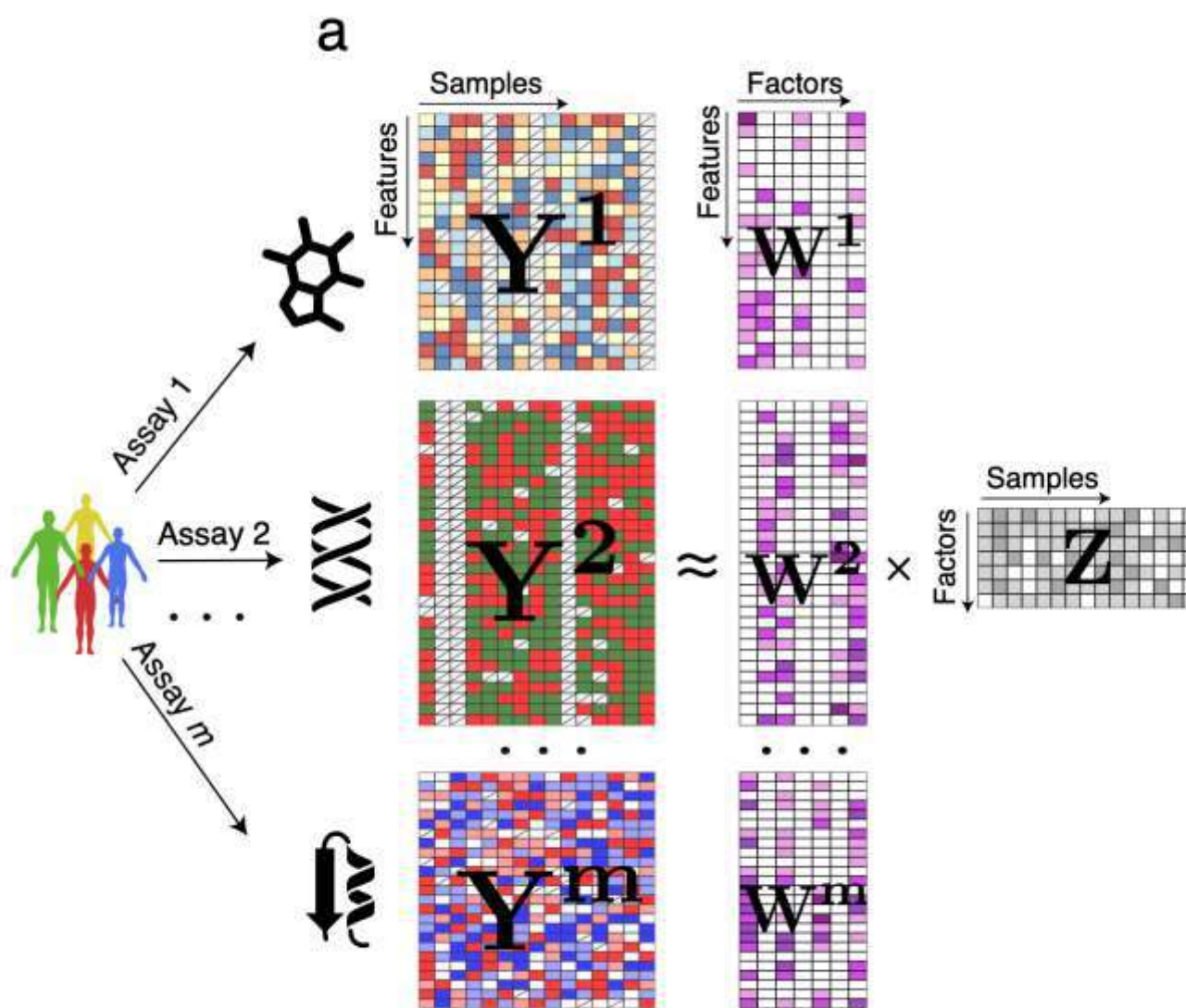
Kaynak: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. *An Introduction to Statistical Learning : with Applications in R*. New York :Springer, 2013

Prognostik Biyobelirteçlerin Tespit Edilmesi

- Differential-expression analizleri ile anlamlı –omiks belirteçlerinin belirlenmesi
- Forward, backward ve stepwise yaklaşımları
- Boruta, recursive feature elimination
- Lasso ve nearest-shrunken centroids gibi spars yaklaşımlar
- Boyut indirgeme yöntemleri
- Biyobelirteçlerin önem katsayılarının elde edilmesi



MOFA2 ile Çoklu Omiks Verilerinin Entegre Edilmesi ile Prognostik Biyobelirteç Tespiti



Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, Stegle O (2020). "MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data." *Genome Biology*, 21.

Sepsis Transkriptom Temelli Prognostik Modeli

Identification of a novel sepsis prognosis model and analysis of possible drug application prospects: Based on scRNA-seq and RNA-seq data

Haihong He^{1†}, Tingting Huang^{2†}, Shixing Guo², Fan Yu¹, Hongwei Shen¹, Haibin Shao², Keyan Chen¹, Lijun Zhang¹, Yunfeng Wu¹, Xi Tang¹, Xinhua Yuan¹, Jiao Liu¹ and Yiwen Zhou^{1*}

¹Department of Emergency Laboratory, Clinical Laboratory Medical Center, Shenzhen Hospital, Southern Medical University, Shenzhen, China, ²Department of General Surgery, Shenzhen Hospital, Southern Medical University, Shenzhen, China

Sepsis is a disease with a high morbidity and mortality rate. At present, there is a lack of ideal biomarker prognostic models for sepsis and promising studies using prognostic models to predict and guide the clinical use of medications. In this study, 71 differentially expressed genes (DEGs) were obtained by analyzing

- Total-RNA-Seq ($n=802$) ile sepsis ve sağlıklı bireyler arasında 1,711 anlamlı gen belirlendi.
- scRNA-Seq ($n=12$) ile sepsis ve sağlıklı bireyler arasında 256 anlamlı gen belirlendi.
- Her iki veride ortak 71 gen tespit edildi.
- Lasso Cox regresyon modeli ile OS üzerinde önemli 11 gen (çoklu modelde 5 gen)

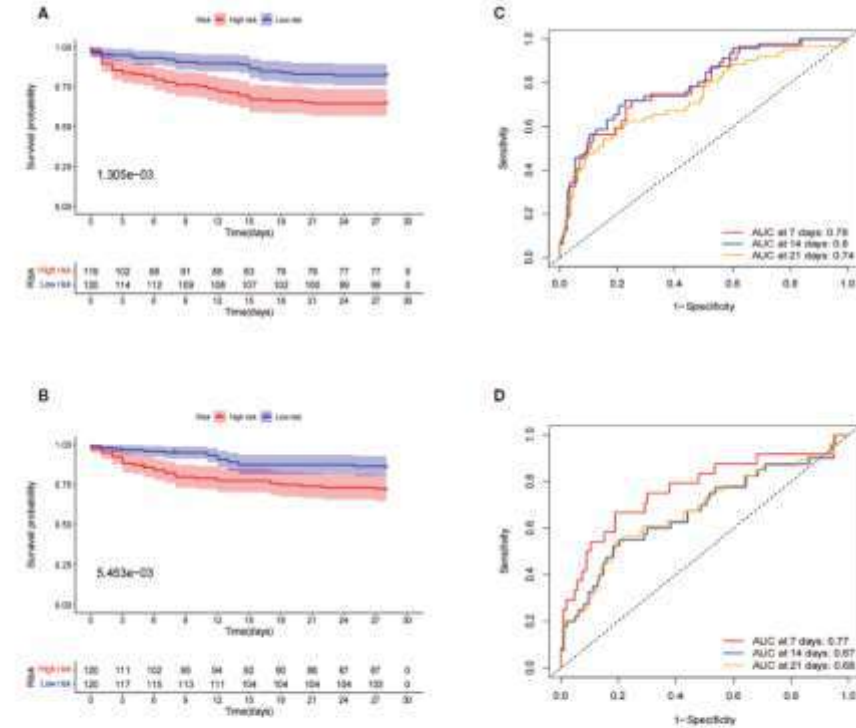


FIGURE 6
Internal validation of prognostic signatures for the entire sepsis dataset. (A, B) Kaplan-Meier survival curve in the train and test cohorts. Patients in the high-risk group had a significantly lower number of days of survival than those in the low-risk group. The red and blue lines represent the high-risk and low-risk groups, respectively. (C, D) ROC curve and AUCs at 7-, 14-, and 21-day survival in the train and test cohorts. Both the train and test cohorts showed better predictive performance, especially at 7 days, with the highest AUC values of 0.78 and 0.77, respectively.

$$\begin{aligned} \text{risk score} = & (-0.292 \times \text{CCL5 exp}) \\ & + (0.231 \times \text{HBD exp}) + (0.442 \times \text{IRF2BP2 exp}) \\ & + (-0.484 \times \text{LTB exp}) + (0.177 \times \text{WFDC1 exp}) \end{aligned}$$

NMR ve yapay zeka ile lipidomik temelli olarak bakteriyel enfeksiyon ve Covid-19 ayrımı



Research article

Differential analysis of lipoprotein and glycoprotein profiles in bacterial infections and COVID-19 using proton nuclear magnetic resonance and machine learning

Simona Iftimie^{a,b}, Núria Amigó^{b,c,d}, Neus Martínez-Micaelo^{b,c,d}, Ana F. López-Azcona^{a,b}, Cristian Martínez-Navidad^e, Helena Castañe^e, Andrea Jiménez-Franco^e, Josep Ribalta^{b,c,f}, Sandra Parra^{a,b}, Antoni Castro^{a,b,g,h}, Jordi Camps^{b,e,i}, Jorge Joven^{b,e}

^a Department of Internal Medicine, Hospital Universitari de Sant Joan, Institut d'Investigació Sanitària Pere Virgili, Universitat Rovira i Virgili, Reus, Spain

^b Department of Medicine and Surgery, Universitat Rovira i Virgili, Reus, Spain

^c CIBER of Diabetes and Associated Metabolic Disease (CIBERDEM), Instituto de Salud Carlos III, Madrid, Spain

^d Biofer Testlab, Reus, Spain

^e Unitat de Recerca Biomèdica, Hospital Universitari de Sant Joan, Institut d'Investigació Sanitària Pere Virgili, Universitat Rovira i Virgili, Reus, Spain

^f Unitat de Recerca de Lipids i Arterioesclerosi, Hospital Universitari de Sant Joan, Institut d'Investigació Sanitària Pere Virgili, Universitat Rovira i Virgili, Reus, Spain

- COVID-19'lu 124 hasta, kateterle ilişkili bakteriyel enfeksiyonu olan 50 hasta ve 50 sağlıklı gönüllüden serum örnekleri alındı.
- Proton NMR ile lipoprotein ve glikolipoprotein profili çıkarıldı.
- Random forests algoritması ile en önemli değişkenler belirlendi.
- PCA analizi ile hastalık grupları sınıflandırıldı.

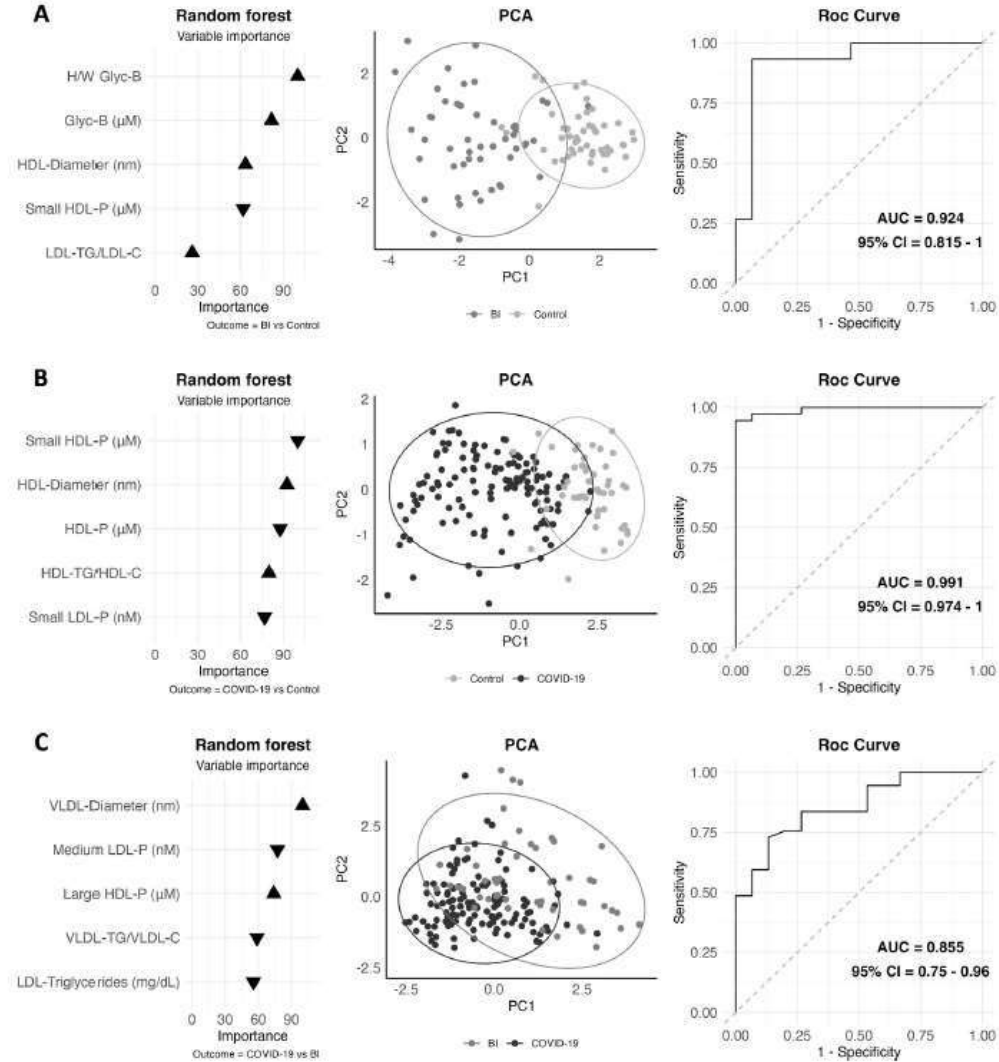


Fig. 5. From left to right, random forest plots, principal component analysis (PCA) and receiver component characteristics (ROC) curves when comparing patients with bacterial infections (BI) and the control group (A), COVID-19 patients with the control group (B) and COVID-19 with BI patients (C). In the random forest plots, unturned triangles indicate a higher value and inverted triangles indicate a lower value in the comparison group. AUC: Area under the curve; C: Cholesterol; CI: Confidence interval; HDL: High-density lipoprotein; H/W: Height to width ratio; LDL: Low-density lipoprotein; P: Particles; TG: Triglycerides; VLDL: Very low-density lipoprotein.

Multi-omiks yaklaşımı ile HIV-enfeksiyonlu hastaların metabolik risk profilinin belirlenmesi



RESEARCH ARTICLE

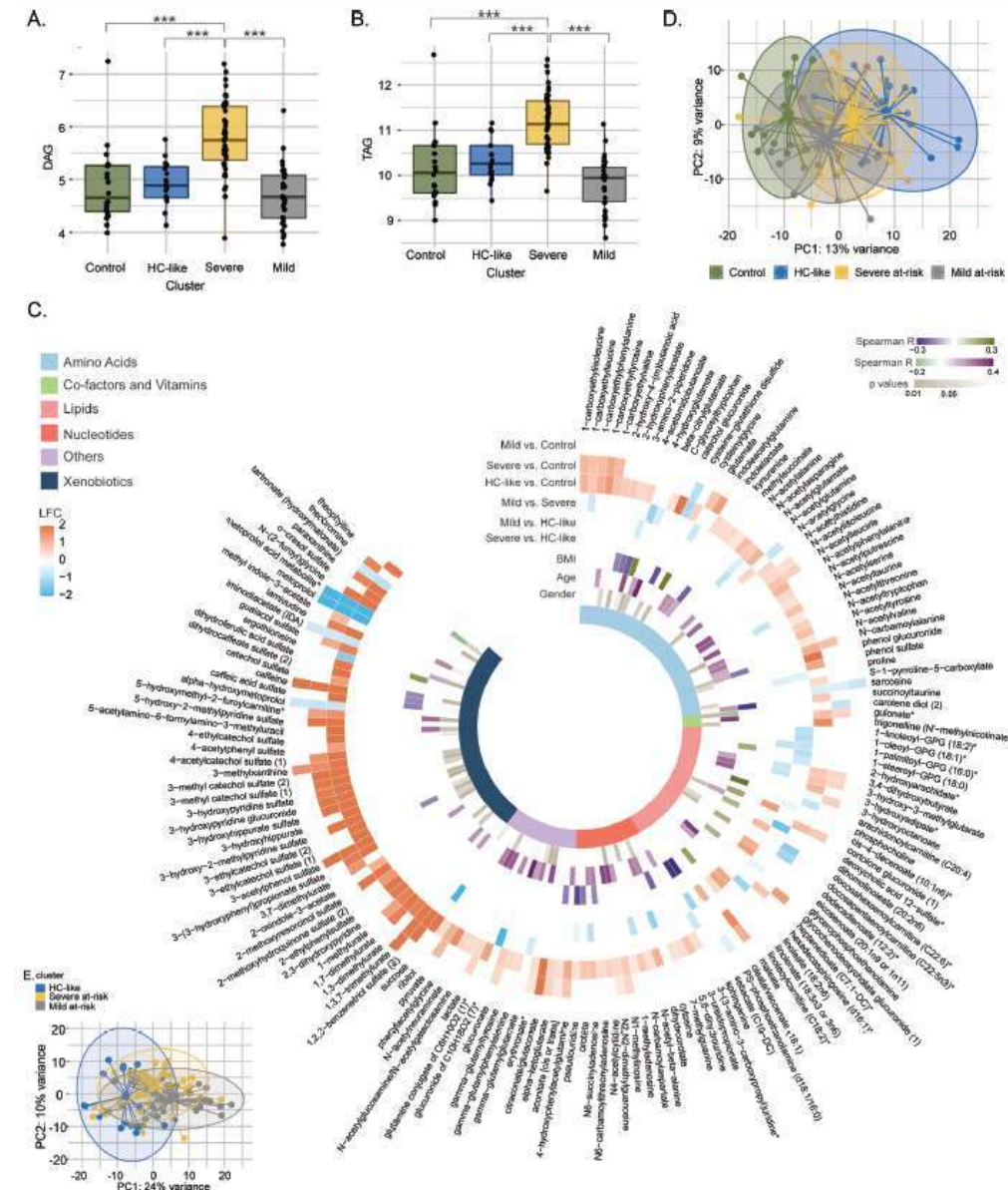


Network-based multi-omics integration reveals metabolic at-risk profile within treated HIV-infection

Flora Mikaeloff^{1*}, Marco Gelpi², Rui Benfeitas³, Andreas D Knudsen², Beate Vestad^{4,5}, Julie Høgh², Johannes R Hov^{4,5,6}, Thomas Benfield⁷, Daniel Murray⁸, Christian G Giske⁹, Adil Mardinoglu^{10,11}, Marius Trøseid^{4,12}, Susanne D Nielsen^{2†}, Ujjwal Neogi^{1††}

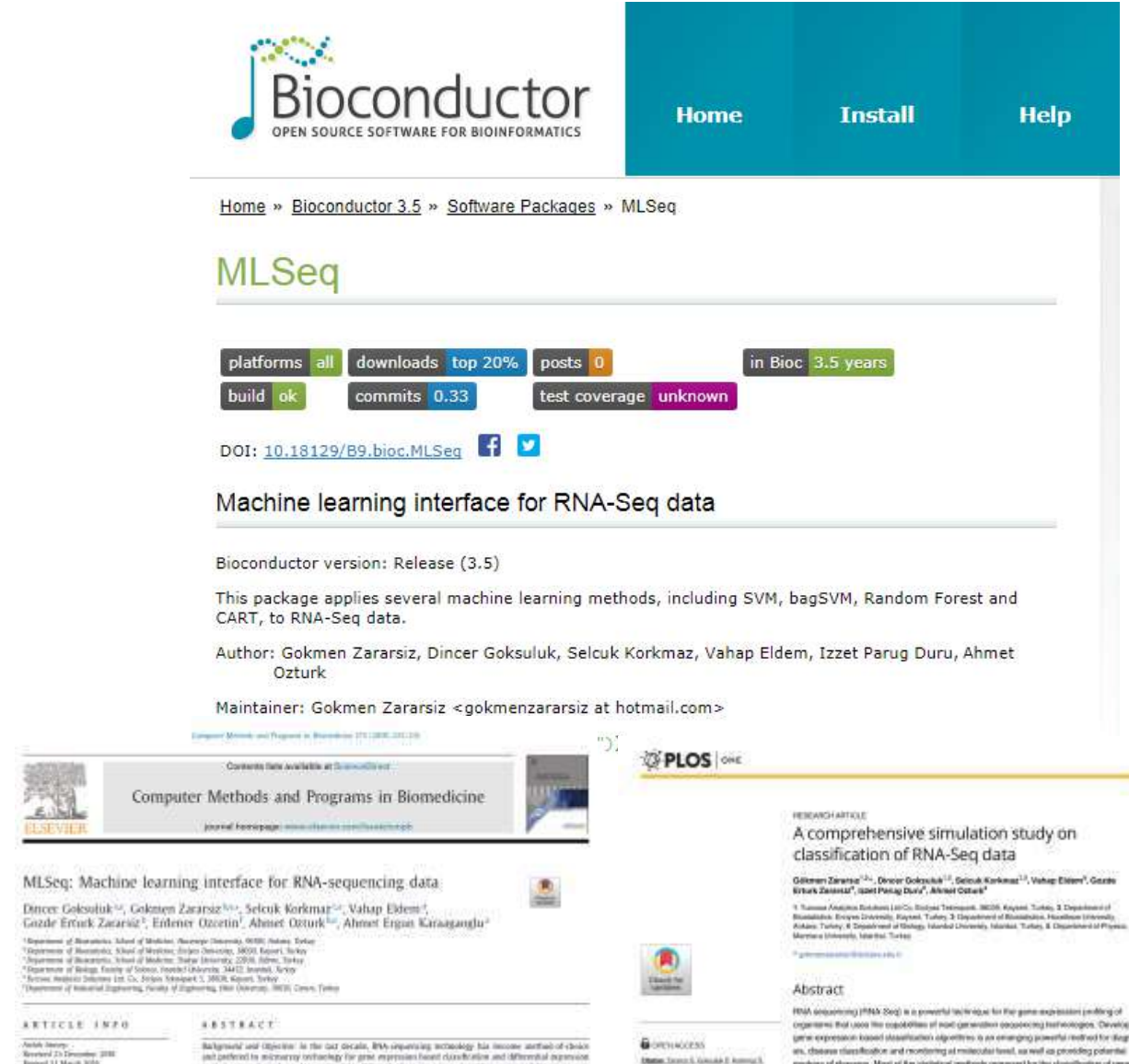
¹The Systems Virology Lab, Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, Stockholm, Sweden; ²Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark; ³National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden; ⁴Research Institute of Internal Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway; ⁵Norwegian PSC Research Center, Oslo University Hospital Rikshospitalet, Oslo, Norway; ⁶Institute of Clinical Medicine, University of Oslo, Oslo, Norway; ⁷Department of Infectious Diseases, Copenhagen University Hospital – Amager and Hvidovre, Hvidovre, Denmark; ⁸Centre of Excellence for Health, Immunity and Infections (CHIP), Rigshospitalet, University of Copenhagen, Copenhagen, Denmark; ⁹Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden; ¹⁰Science for Life Laboratory, KTH - Royal Institute of

- Antiretroviral tedavi gören HIV ile yaşayan kişilerin sistematik risk kategorizasyonunun eksikliği
- Plazma lipidomik, metabolomik ve 16S mikrobiyom analizleri
- Ağ analizi ve similarity network fusion algoritması ile SNF-1 (sağlıklı), SNF-2 (şiddetli risk) ve SNF-3 (hafif risk)
- Hastalık risk grupları açısından önemli metabolitler klinik parametreler ile ilişkilendirildi.



Prognostik –omiks temelli yapay zeka yazılımları

- MLSeq, şu anda RNA-Seq tabanlı biyobelirteç keşfi ve tahmin analizi için en kapsamlı araçtır.
- 17.764 farklı IP adresinden 34.035 indirme (21 Haziran 2025)
- 93 makine öğrenimi modeli
- PLDA, NBLDA, voomDDA sınıflandırıcıları
- SVM, RF vb. dahil olmak üzere popüler makine öğrenimi sınıflandırıcıları
- Derin öğrenme sınıflandırıcıları



The image shows the Bioconductor MLSeq package page and its corresponding PLOS ONE publication. The Bioconductor page includes the package name, version (3.5), and a description: "Machine learning interface for RNA-Seq data". It also lists the authors: Gokmen Zararsiz, Dincer Goksuluk, Selcuk Korkmaz, Vahap Eldem, Izzet Parug Duru, Ahmet Ozturk, and the maintainer: Gokmen Zararsiz. The PLOS ONE article is titled "A comprehensive simulation study on classification of RNA-Seq data" and lists the same authors. The abstract of the article states: "RNA sequencing (RNA-Seq) is a powerful technique for the gene expression profiling of organisms that uses the capabilities of next generation sequencing technologies. Developing gene expression based classification algorithms is an emerging powerful method for diagnosis, disease classification and monitoring at molecular level, as well as providing potential evidence of disease. Most of the statistical methods proposed for the classification of cancer..."

Prognostik –omiks temelli yapay zeka yazılımları

voomDDA: Discovery of Diagnostic Biomarkers and Classification of RNA-Seq Data (ver. 1.5)



Introduction | Data upload | voomDDA | Heatmap | Network | GO | Manual | Authors & News | Citation

Identifying the relevant genes (or other genomic features such as transcripts, miRNAs, lncRNAs, etc.) across the conditions (e.g. tumor and non-tumor tissue samples) is a common research interest in gene-expression studies. In this gene selection, researchers are often interested in detecting a small set of genes for diagnostic purpose in medicine that involves identification of the minimal subset of genes that achieves maximal predictive performance. Biomarker discovery and classification problem.

VoomDDA is a decision support tool developed for RNA-Sequencing datasets to assist researchers in their decisions for diagnostic biomarker discovery and classification problem. VoomDDA consists both sparse and non-sparse statistical learning classifiers adapted with voom method. Voom is a recent method that estimates the mean and variance relationship of the log-counts of RNA-Seq data (log counts per million, log-cpm) at observational level. It also provides precision weights for each observation that can be incorporated with the log-cpm values for further analysis. Algorithms in our tool incorporate the log-cpm values and the corresponding precision weights into biomarker discovery and classification problem. For this purpose, these algorithms use weighted statistics in estimating the discriminating functions of the used statistical learning algorithms.

VoomNSC is a sparse classifier that is developed to bring together two powerful methods for RNA-Seq classification:

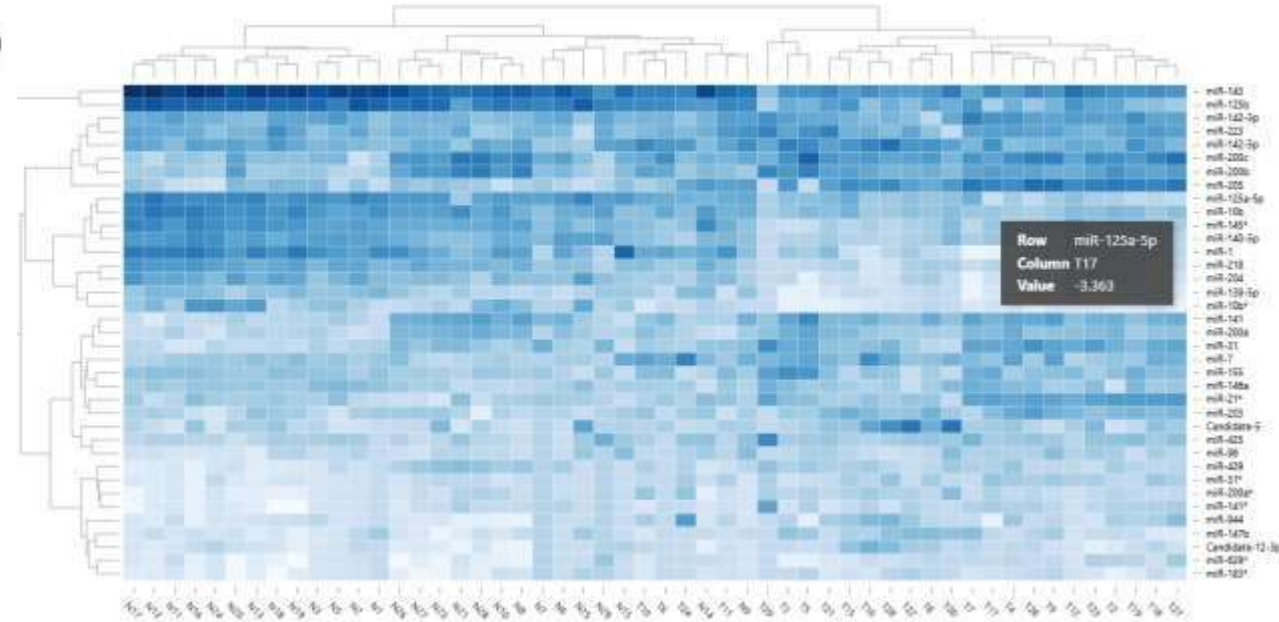
1. to extend voom method for RNA-Seq classification studies.
2. to make nearest shrunken centroids (NSC) algorithm available for RNA-Seq technology.

VoomNSC both provides fast, accurate and sparser classification results for RNA-Seq data. More details can be found in the research paper. This tool also includes RNA-Seq extensions of diagonal linear and diagonal quadratic discriminant classifiers: (i) voomDLDA and (ii) voomDQDA.

References

[1] Zararsiz, G., Goksuluk, G., Korkmaz, S., et al. (2015). VoomDDA: Discovery of Diagnostic Biomarkers and Classification of RNA-Seq Data.

[2] Law, C.W., Chen, Y., Shi, W., et al. (2014). voom: Precision weights unlock linear model analysis tools for RNA-Seq read counts. *Genome Biology* 15:R29.



<http://biosoft.erciyes.edu.tr/app/voomDDA>



voomDDA: discovery of diagnostic biomarkers and classification of RNA-seq data

Gökmen Zararsiz^{1,2}, Dincer Goksuluk^{3,4}, Bernd Klaus⁵, Selcuk Korkmaz^{2,5}, Vahap Eldem⁶, Erdem Karabulut¹ and Ahmet Ozturk^{1,2}

¹Department of Biostatistics, Erciyes University, Kayseri, Turkey

²Tarcusa Analytics Solutions Ltd-Co, Erciyes Teknopark 5, Kayseri, Turkey

³Department of Biostatistics, Hacettepe University, Ankara, Turkey

⁴European Molecular Biology Laboratory, Heidelberg, Germany

⁵Department of Biostatistics, Trakya University, Edirne, Turkey

⁶Department of Biology, Istanbul University, Istanbul, Turkey

ABSTRACT

RNA-Seq is a recent and efficient technique that uses the capabilities of next-

Prognostik –omiks temelli yapay zeka yazılımları

MLMS Webtool

Introduction
Upload Data
Classify
Manual
Authors

Upload Data Options

Train Data:
☒ Use Example Data 1 ☐ Use Example Data 2
☐ Upload Data

Test Data:
☒ Use Example Data 1 ☐ Use Example Data 2
☐ Upload Data

Data Preview

Train Data:
Show 10 entries Search:

	sampleNames	groups	X1.6.Anhydro.beta.D.glucose	X1.Methylnicotinamide	X2.Aminobutyrate	X2.Hyd
1	NETL_003_V2	control	38.47	12.55	15.03	
2	NETL_017_V1	control	22.2	20.7	7.85	
3	NETCR_025_V2	cachexic	16.95	114.43	2.53	
4	PIF_113	cachexic	167.34	19.89	13.46	
5	NETL_004_V1	cachexic	4.71	11.13	43.38	
6	PIF_132	cachexic	214.86	127.74	31.5	
7	PIF_112	control	22.87	10.38	1.28	
8	PIF_178	cachexic	40.85	65.37	18.73	
9	PIF_089	cachexic	123.97	81.45	55.15	
10	NETL_008_V1	cachexic	59.74	50.91	6.82	

Showing 1 to 10 of 53 entries. Previous 1 2 3 4 5 6 Next

Test Data:
Show 10 entries Search:

<https://github.com/gokmenzararsiz/MLMS>

KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü

Received: 29 December 2023 | Accepted: 18 April 2024

DOI: 10.1002/jmv.29637

RESEARCH ARTICLE

JOURNAL OF
MEDICAL VIROLOGY WILEY

Exploring free amino acid profiles in Crimean-Congo hemorrhagic fever patients: Implications for disease progression

Seyit Ali Büyüktuna¹ | Serra İlayda Yerlitaş^{2,3} | Gözde Ertük Zararsız^{2,3} | Kübra Doğan⁴ | Demet Kablan⁵ | Gökhan Bağcı⁶ | Selda Özer⁵ | Cihad Baysal¹ | Yasemin Çakır¹ | Ahu Cephe⁷ | Necla Koçhan⁸ | Gökmen Zararsız^{2,3} | Halef Okan Doğan⁵

¹Department of Infectious Disease and Clinical Microbiology, Cumhuriyet University School of Medicine, Sivas, Türkiye

²Department of Biostatistics, Erciyes University School of Medicine, Kayseri, Türkiye

³Drug Application and Research Center (ERFARMA), Erciyes University, Kayseri, Türkiye

⁴Department of Biochemistry, Minister of Health Sivas Numan Hospital, Sivas, Türkiye

⁵Department of Biochemistry, Cumhuriyet University School of Medicine, Sivas, Türkiye

⁶Faculty of Medicine, Medical Biochemistry, Altınbaş University, İstanbul, Türkiye

⁷Institutional Data Management and Analytics Unit, Erciyes University Rectorate, Kayseri, Türkiye

⁸Department of Mathematics, İzmir University of Economics, İzmir, Türkiye

Correspondence

Halef Okan Doğan, Department of Biochemistry, Cumhuriyet University School of Medicine, 58140, Sivas, Türkiye.
Email: halefokan@gmail.com

Gökmen Zararsız, Department of Biostatistics, Erciyes University School of Medicine, Kayseri, Türkiye.

Email: gokmen.zararsiz@gmail.com

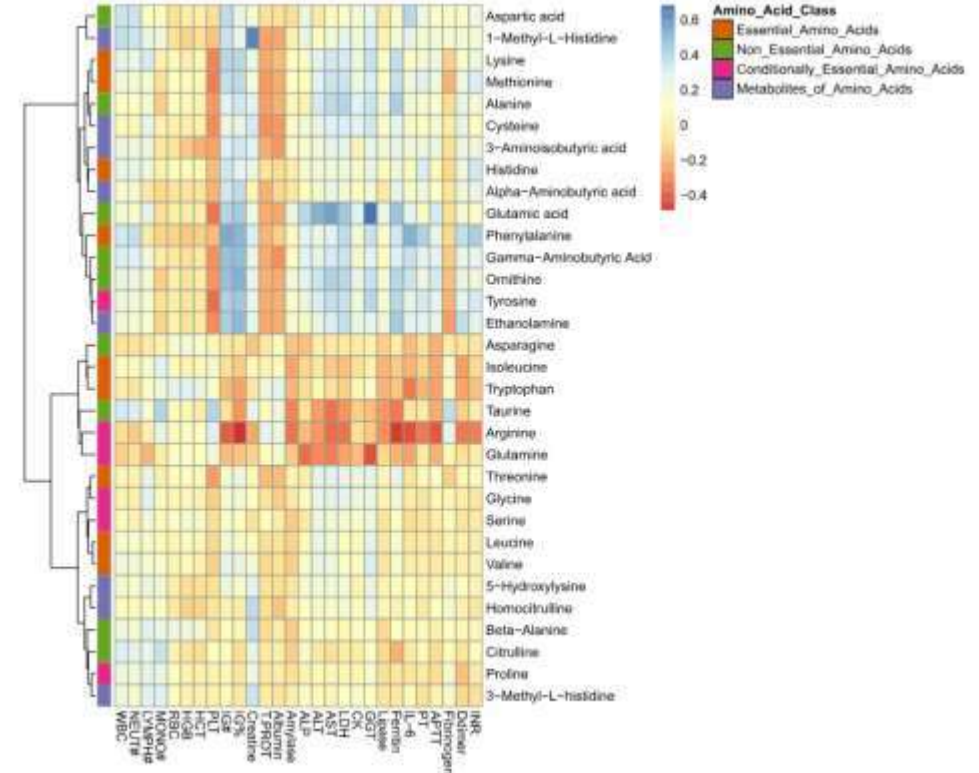
Funding information

Research Fund of Erciyes University,
Grant/Award Number: TSG-2021-10912

Abstract

This study investigated the intricate interplay between Crimean-Congo hemorrhagic fever virus infection and alterations in amino acid metabolism. The primary aim is to elucidate the impact of Crimean-Congo hemorrhagic fever (CCHF) on specific amino acid concentrations and identify potential metabolic markers associated with viral infection. One hundred ninety individuals participated in this study, comprising 115 CCHF patients, 30 CCHF negative patients, and 45 healthy controls. Liquid chromatography-tandem mass spectrometry techniques were employed to quantify amino acid concentrations. The amino acid metabolic profiles in CCHF patients

- Çalışmaya KKKA tanısı almış 115 hasta, 20 KKKA negatif birey ve 45 kontrol bireyi dahil edildi.
- KKKA hastalarının 18'i ICU olmak üzere 16 hasta ex oldu.
- Hedefli metabolomik yaklaşımı ile amino-asit düzeyleri kantitatif olarak elde edildi.
- Arjinin, histidin, taurin, glutamik asit ve glutamin metabolitlerinin prognostik önemi



KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü

- Veri seti %70 eğitim, %30 test seti olarak iki parçaya ayrıldı. Tüm süreç 30 kez tekrar edildi.
- RFE ve Boruta yöntemleri ile hastalığın prognozu ile ilişkili metabolitler seçildi.
- Sınıf dengesizliğinden dolayı yukarı, aşağı ve SMOTE örnekleme yöntemleri uygulandı.
- Random forests, lasso, kNN, NSC, SVM ve XGBoost yöntemleri ile ICU tahmin edildi.
- Rasgele sağkalım ormanı, sağkalım lasso ve sağkalım XGBoost ile OS tahmin edildi.
- Model parametrelerinin optimizasyonu için 10 tekrarlı, 5 parça çapraz geçerlilik.



`cran/mlr3proba`

KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü

Sınıflandırma modelleri:

- Eğri altında kalan alan (AUC)
- Doğruluk oranı (ACC)
- Duyarlılık (SENS)
- Seçicilik (SPEC)
- Pozitif kestirim değeri (PPV)
- Negatif kestirim değeri (NPV)
- Pozitif olasılık oranı (LR+)
- Negatif olasılık oranı (LR-)

Sağkalım modelleri:

- Harrell'in uyum indeksi (C-indeksi)
- Birleştirilmiş Brier skoru (IBS)
- D-kalibreli skor

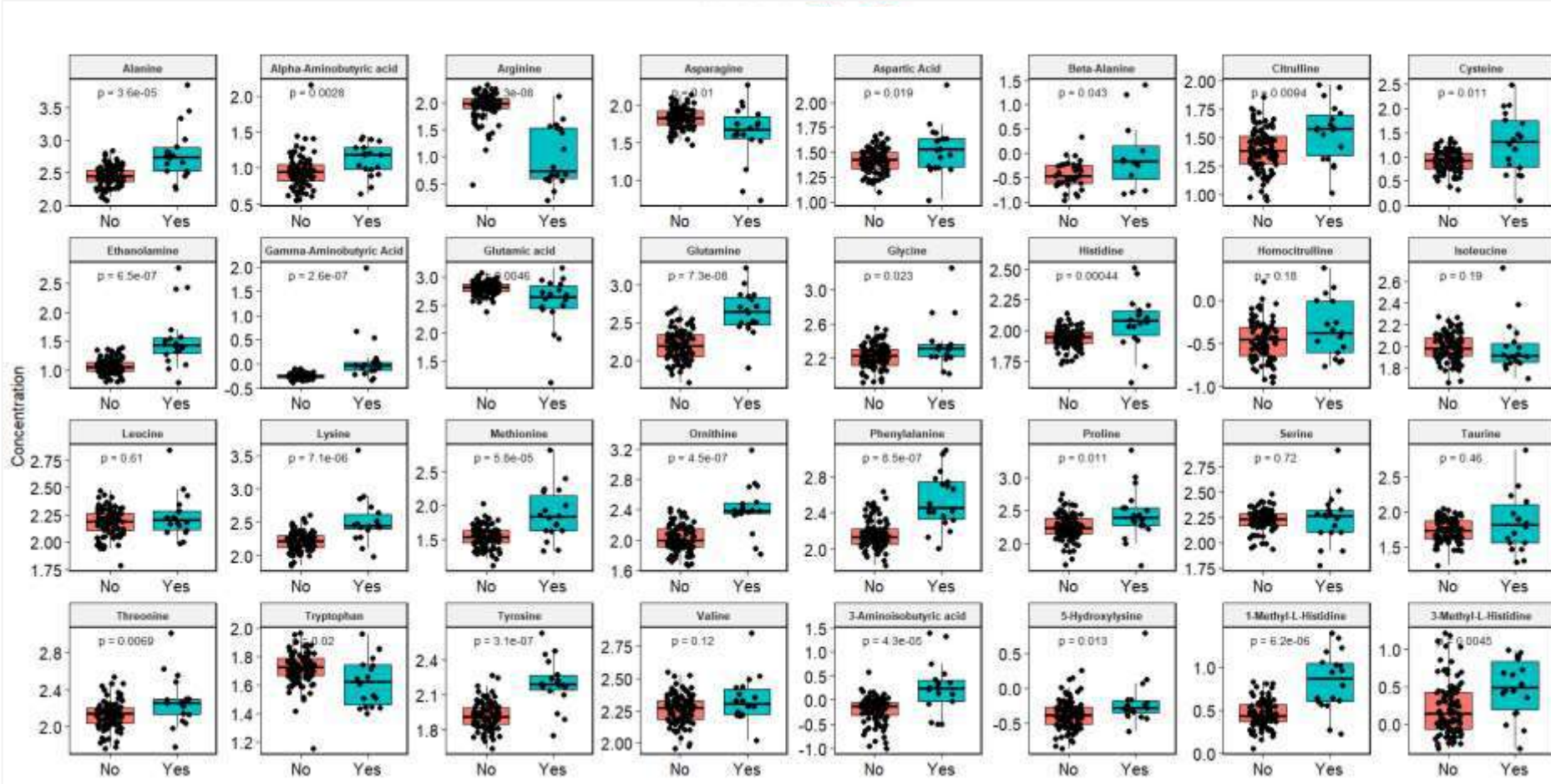
Değişken Önemliliği (Variable Importance):

- Her amino asidin yoğun bakım ihtiyacını belirlemede modelin kestirim gücüne katkısını sıralamak için değişken önem analizi yapıldı.

Kestirim Eğrisi (Predictiveness Curve):

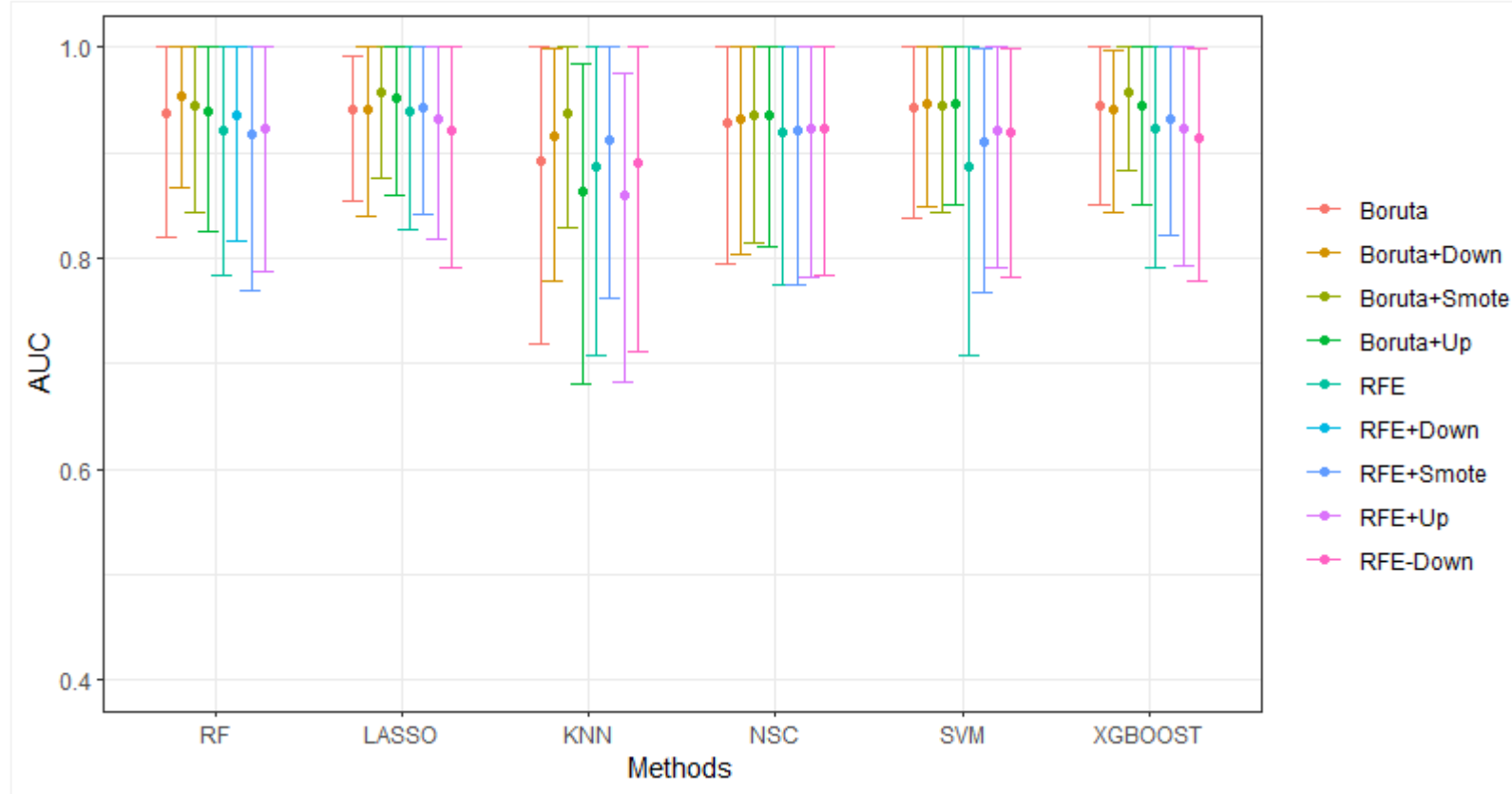
- Oluşturulan modellerin sonucu ne kadar iyi tahmin ettiğini görsel olarak değerlendirmek için kestirim eğrileri çizildi.

KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü



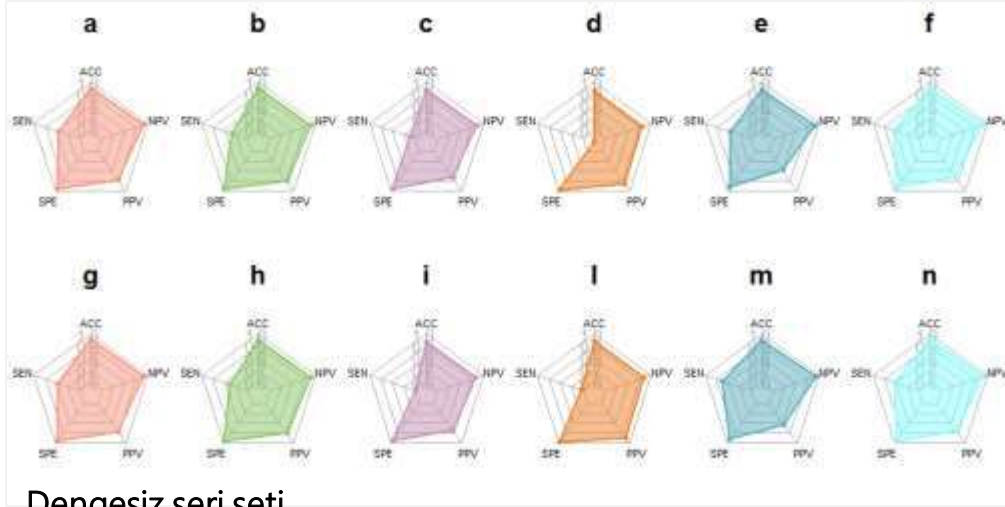
Yoğun bakım yatış durumuna göre aminoasit düzeylerinin karşılaştırılması

KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü

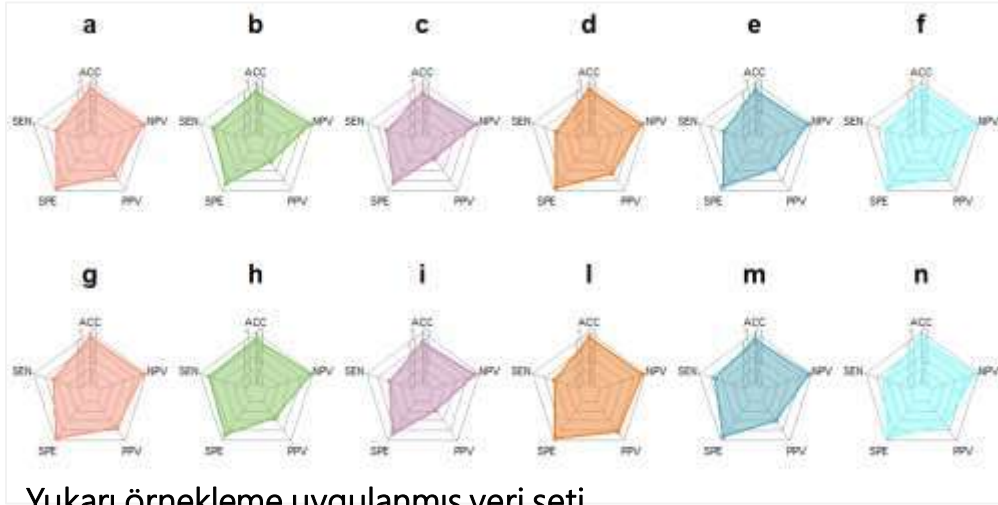


Makine öğrenmesi modellerinin AUC düzeylerinin karşılaştırılması

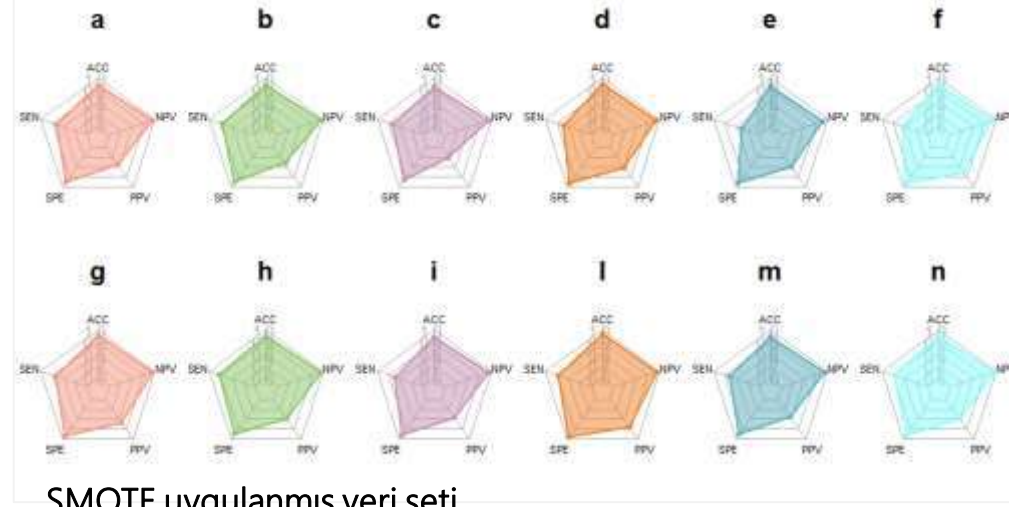
KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü



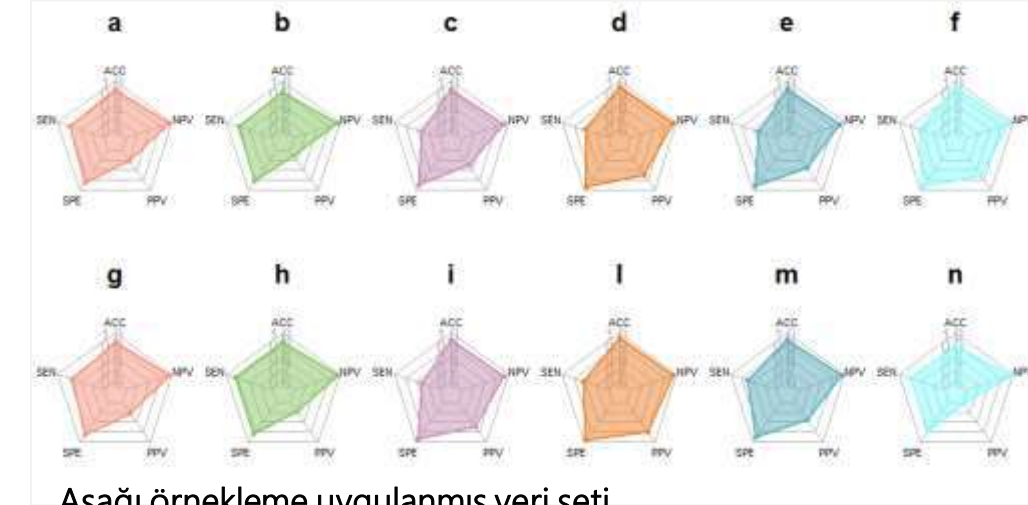
Dengesiz seri seti



Yukarı örnekleme uygulanmış veri seti



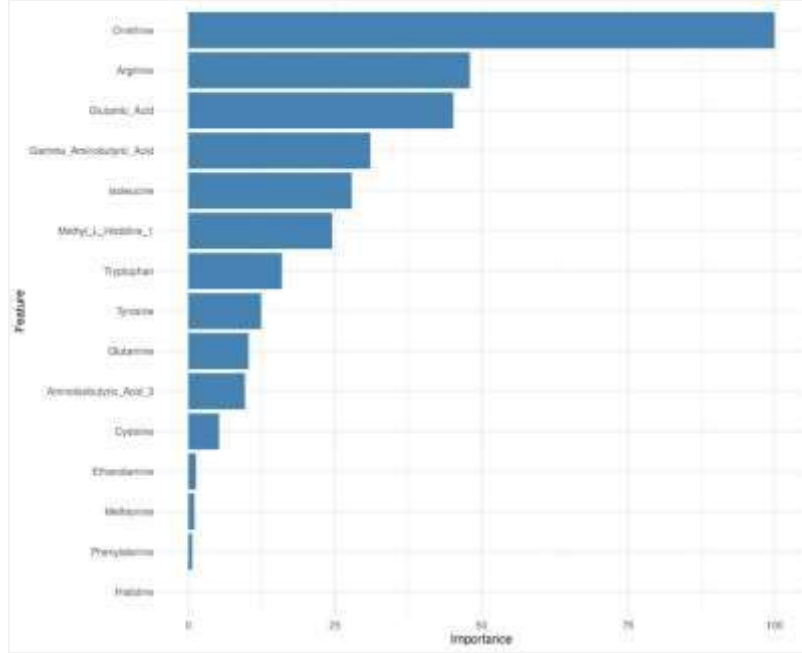
SMOTE uygulanmış veri seti



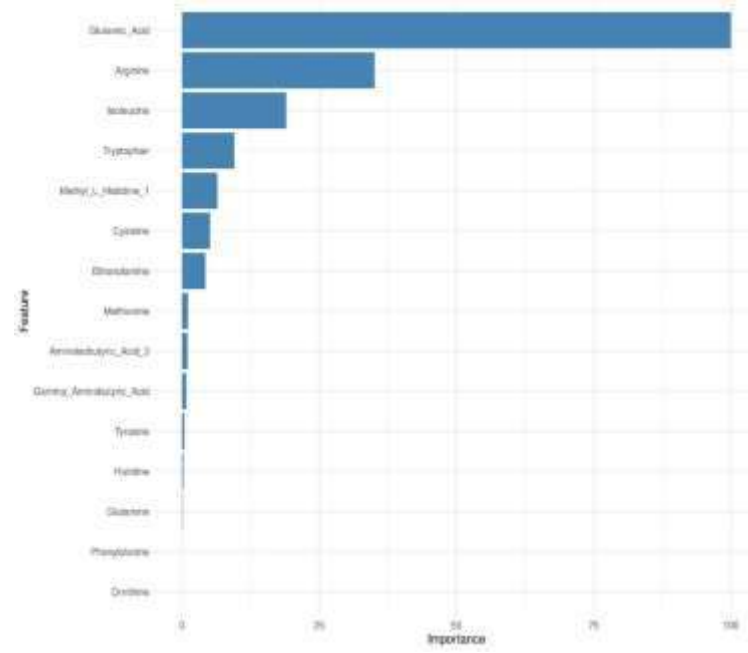
Aşağı örnekleme uygulanmış veri seti

- a. RF - RFE
- b. LASSO - RFE
- c. KNN - RFE
- d. NSC - RFE
- e. SVM - RFE
- f. XGBOOST - RFE
- g. RF - Boruta özellik seçimi
- h. LASSO - Boruta özellik seçimi
- i. KNN - Boruta özellik seçimi
- l. NSC - Boruta özellik seçimi
- m. SVM - Boruta özellik seçimi
- n. XGBOOST - Boruta özellik seçimi

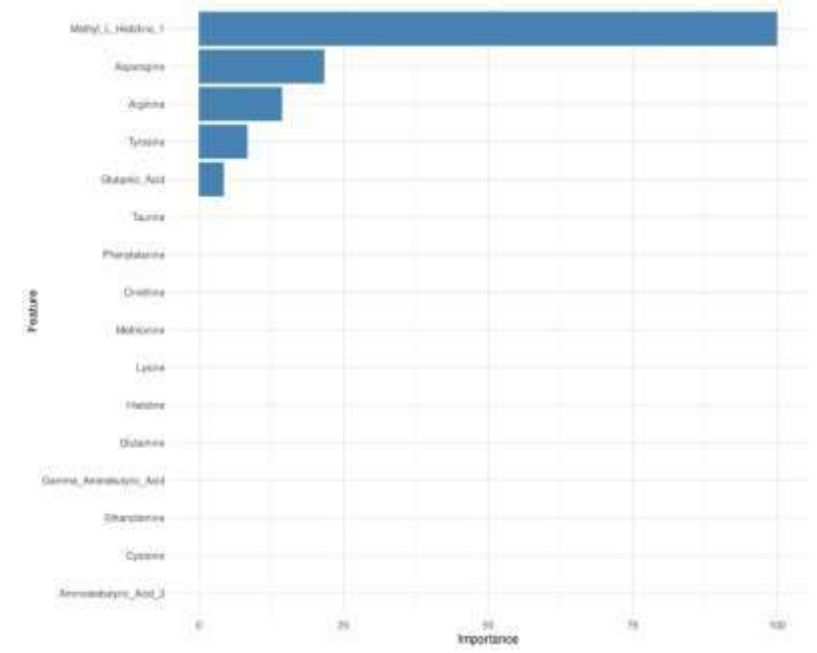
KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü



XGBOOST+BORUTA

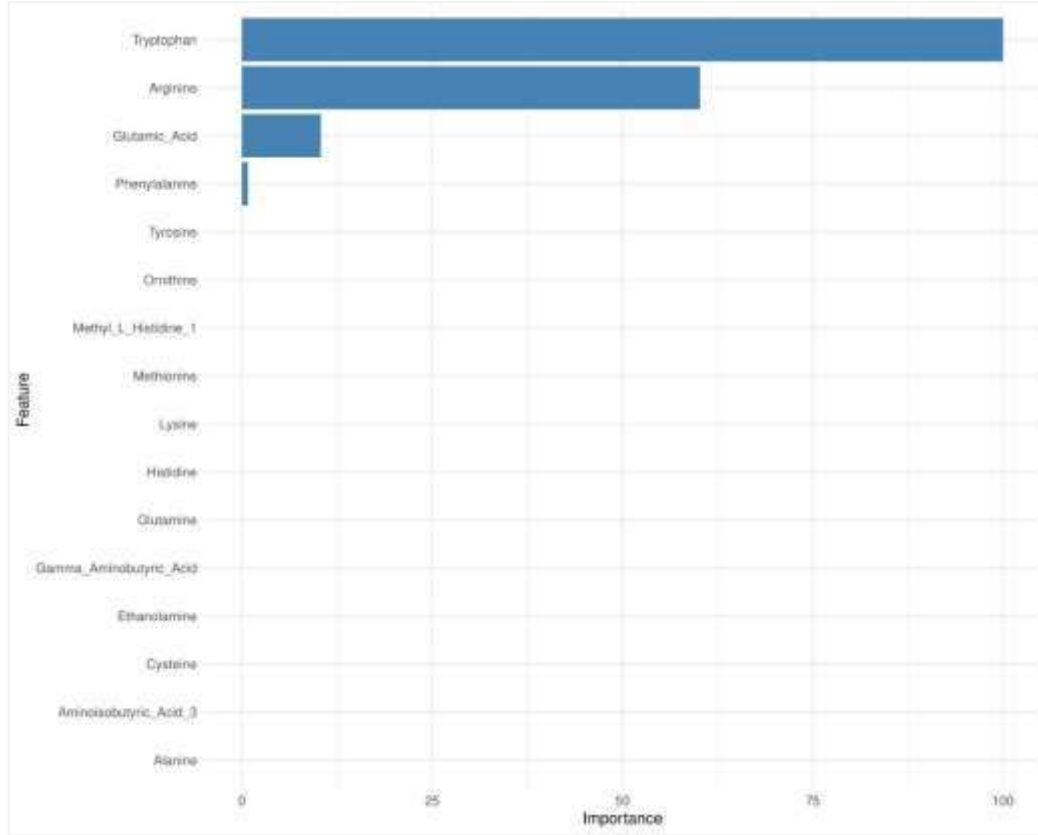


XGBOOST+BORUTA+SMOTE

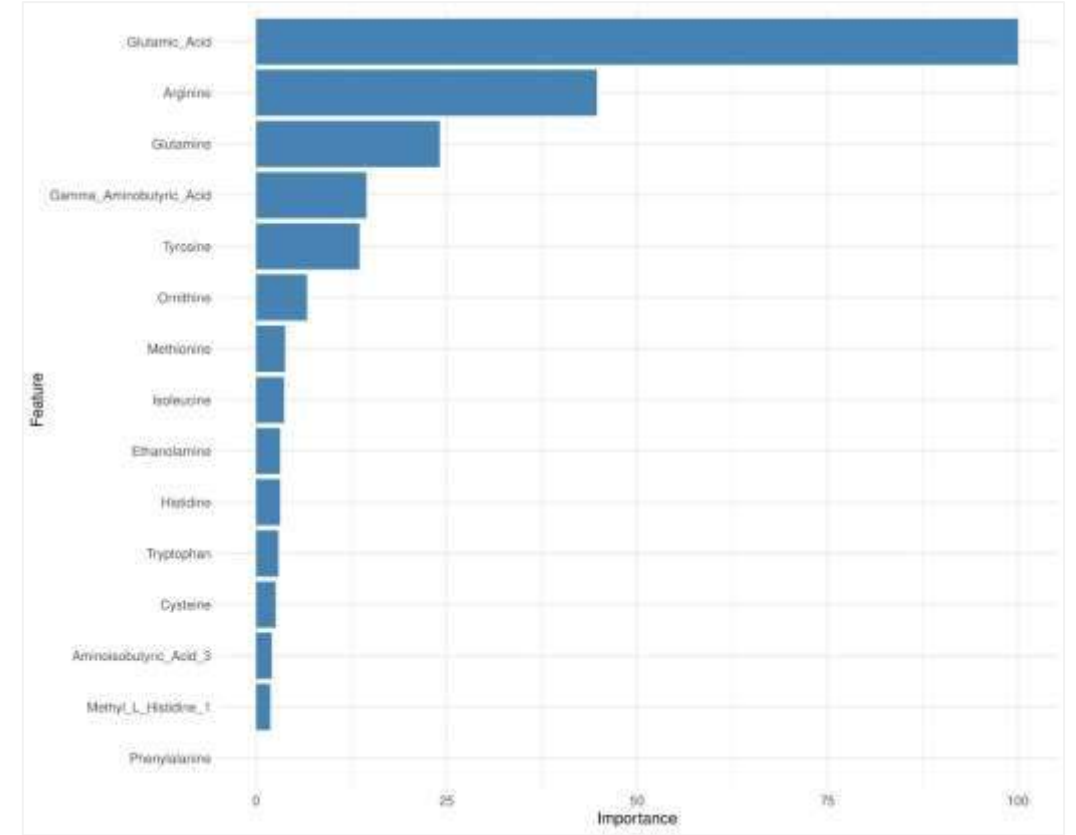


LASSO+BORUTA+SMOTE

KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü

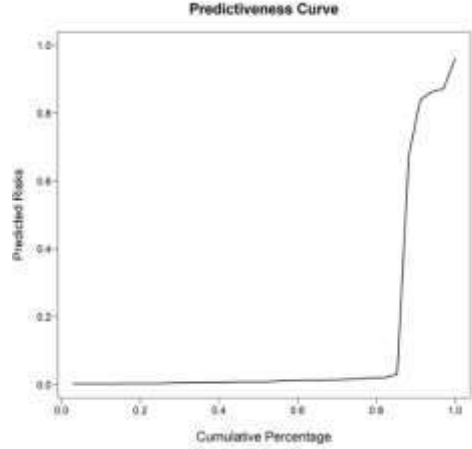


LASSO+BORUTA+UP

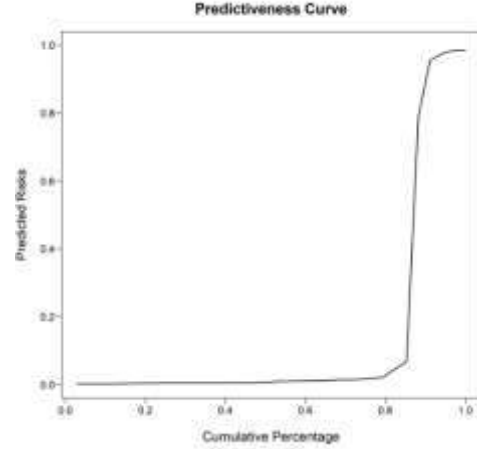


RF+BORUTA+DOWN

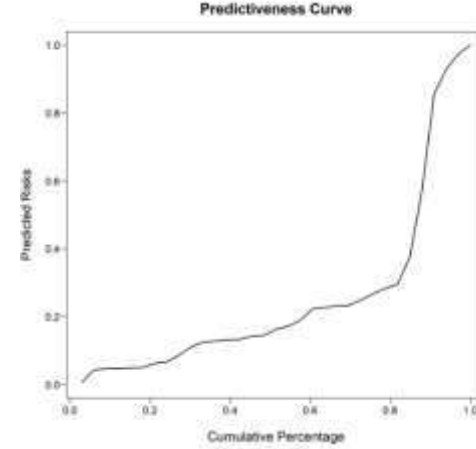
KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü



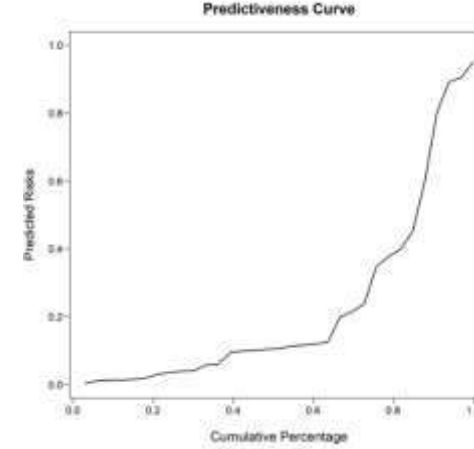
XGBOOST+BORUTA



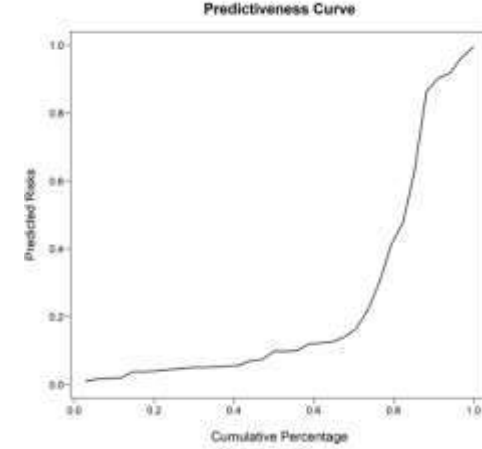
XGBOOST+BORUTA+SMOTE



LASSO+BORUTA+SMOTE

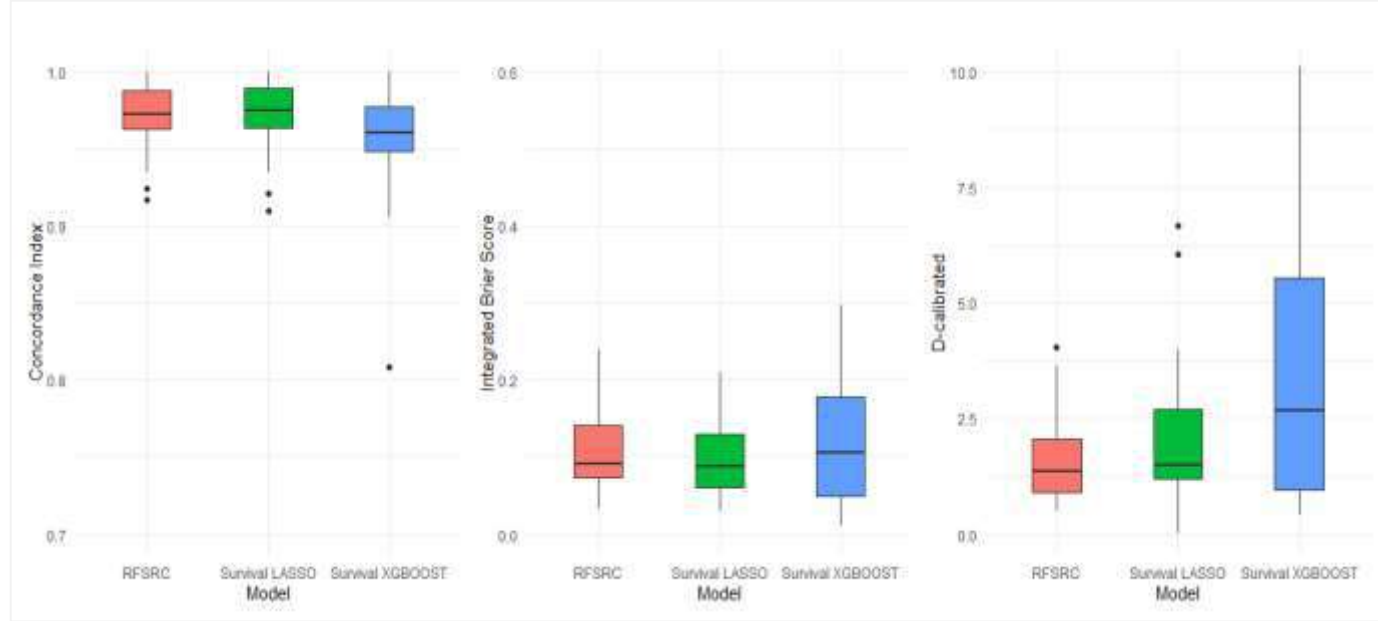


LASSO+BORUTA+UP

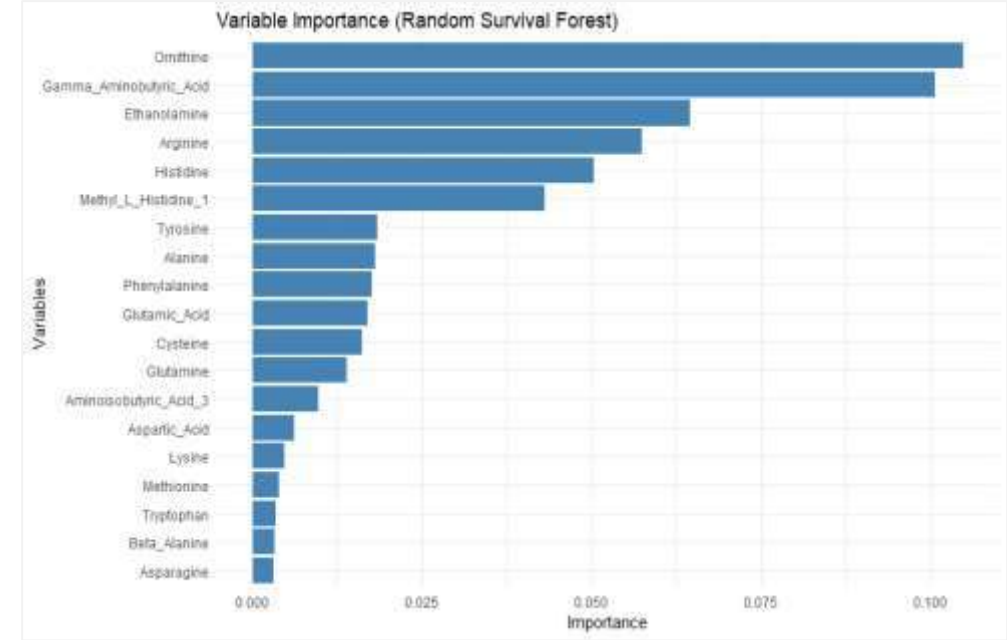


RF+BORUTA+DOWN

KKKAH hastalarında metabolomik temelli ve yapay zeka temelli prognoz öngörüsü



Sağkalım makine öğrenme algoritmalarının karşılaştırılması



RFSRC modeli değişken önemlilikleri

KKKAH hastalığı ile ilgili devam eden projeler

- Klinik laboratuvar verileri ile yapay zeka destekli kapsamlı prognoz öngörüsü
- KKKAH yağ asidi ve organik asit profilinin ortaya konması
- Multi-omiks yaklaşımı ile yapay zeka destekli prognostik kestirimlerin elde edilmesi
- MAS-10® ile KKKAH hedef gözetmeyen metabolomik profilinin çıkarılması
- Kişiye özgü klinik ve metabolomik temelli ve yapay zeka destekli hasta takip raporunun oluşturulması





pmaa

Hassas Tıp ve İleri Analitik Araştırma Grubu



Erciyes Üniversitesi Araştırma Dekanlığı ARGEPARK Binası –
Köşk Mahallesi, Fakülte İçi Küme Evleri, No:4, 38030 Melikgazi/Kayseri



+903522076666 - 12500



<https://avesis.erciyes.edu.tr/arastirma-grubu/pmaa/hakkimizda>



zararsizlab@erciyes.edu.tr



[@zararsizlab](https://twitter.com/zararsizlab)



<https://www.linkedin.com/company/zararsizlab/>



[@zararsizlab](https://www.instagram.com/zararsizlab)