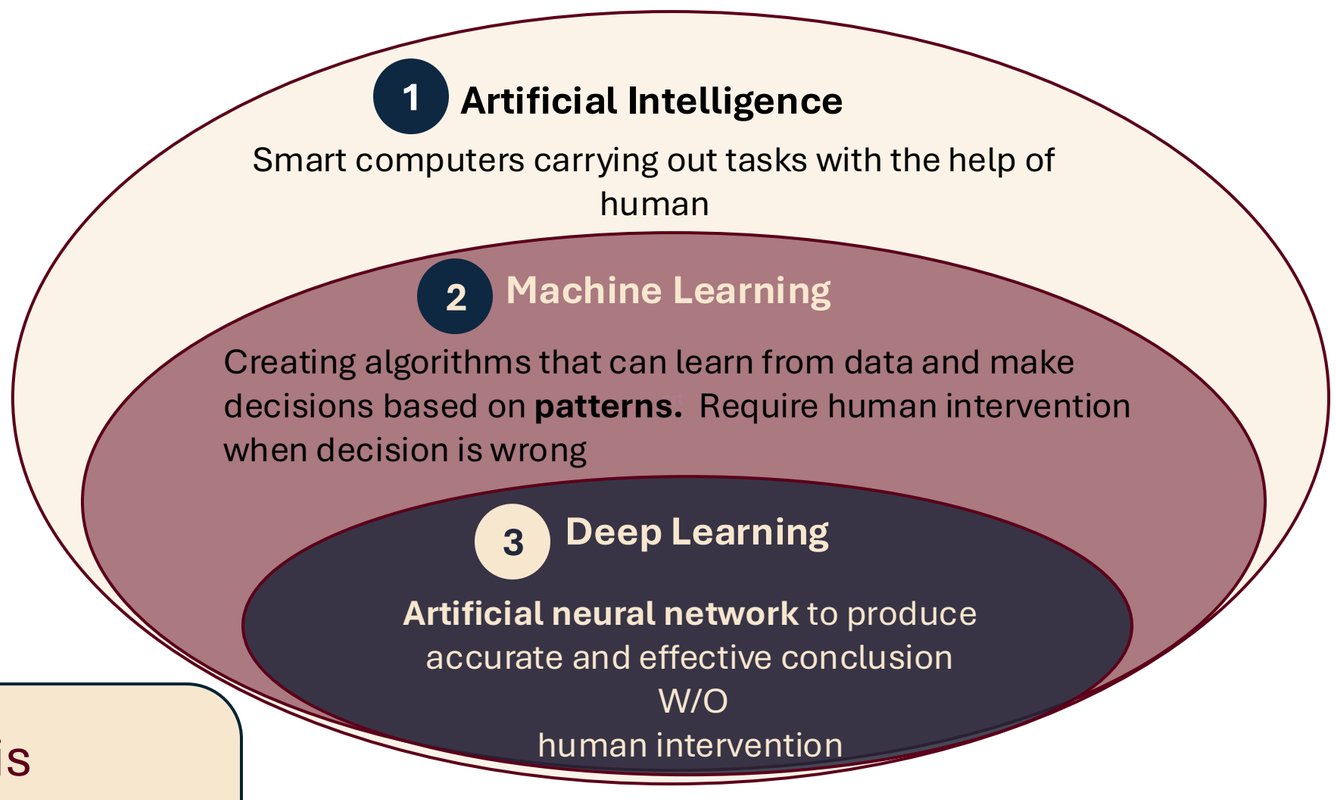
Exploring Artificial Intelligence & Machine Learning for the Diagnosis and Treatment of Interstitial Cystitis / Bladder Pain Syndrome

Inal Gultekin G[1], Kalkan H[2], Agit S[2], Cetin Z[3], Mangir N[4]

1. Istanbul Okan University Faculty of Medicine Department of Physiology, 2. Gebze Technical University Computer Engineering, 3. Technische Hochschule Bingen University of Applied Sciences, 4. Hacettepe University Faculty of Medicine Department of Urology

IC/BPS

- Unknown origin
- Chronic condition
- Adverse impact on patients, Society, Healthcare systems,
- No major advancements in the treatment



AIMS

1- Identifying biomarkers for disease diagnosis

2- Investigate disease pathways

- ✓ ML methods & Bioinformatic tools
- ✓ Publicly available expression datasets

Material & Methods

1. Gene symbols in datasets converted to **Ensemble IDs**.
   a. Ensuring inter-dataset consistency and addressing the absence of gene symbols.
2. Following purification, the samples were **binary labeled:**
   b. Patient or non-patient, and each dataset was normalized internally.
3. Three datasets merged and normalized → **increase consistency**
4. **The LASSO method**
5. **The Shapley explainable AI method** revealed the most trusted genes by the classifier.
6. The common selected 94 (features) genes were re-analysed
   a. String v.12.0 for Gene Enrichment
   b. EnrichR for clinical relevance
   c. Reactome v.87 pathway analysis bioinformatic tools.

RESULTS

1. The classifier was trained and tested by 5-fold cross-validation,
2. Averages of 86.67% accuracy, 90.0% precision,
3. 56.667% recall, 79.333% F1 score,
4. 97.1429% ROC area under the curve (AUC) values (Fig 1).
5. The trained model's feature importance attribute was utilized to identify the 26 most influential genes in the model's predictions.
6. The gene enrichment and pathway analysis screenings revealed **high activity in the immune system.**

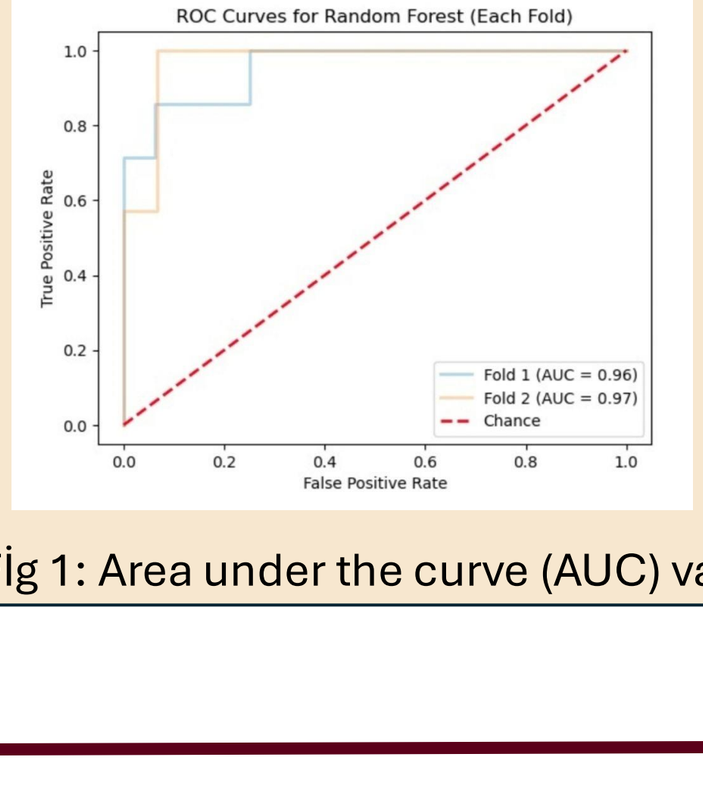


Fİg 1: Area under the curve (AUC) values

INTERPRETATION OF RESULTS

This study, conducted with **data engineering**, ***1- combined, 2- harmonized, and 3- normalized*** three independent human IC/BPS patient and control datasets.

The merging has:

- Combined the features,
- Filled in the missing information of each feature,
  - Resulting in a large pool of more then 7.000 genes (features in AI terminology)

Additional AI methods reduced the feature connections to 94 selected genes, with

- ✓ **high accuracy and precision (**86.67% accuracy, 90.0% precision**).**

7004 Genes → → → 94 genes

1. High AUC values represent that the ML is successful in accuracy and precision for disease and healthy patient identification.
2. Gene enrichment analysis of the selected genes (94 genes) increases the importance of inflammatory defense pathways for IC/BPS.

CONCLUDING MESSAGES

1. The curated dataset is now ready for future ML and deep learning processes.
2. Biomarker identification is crucial for future treatment options in IC/BPS.
3. The identified genes could have crucial roles in biological processes and disease pathophysiology mechanisms, making them potential targets for different settings, including in vitro and in vivo models.
4. Bioinformatic data discloses the immune system as a major actor in the pathophysiology of IC/BPS.