Machine learning uncovers functional microbial signatures in cirrhosis patients with type 2 diabetes mellitus

In Gyu Park[1], Sang Joon Yoon[1] , Sung Min Won[1] , Ki Kwang Oh[1] , Un Joo Lee[2] , Young Lim HAM[3] and Ki Tae Suk[1]

1 Institute for Liver and Digestive Diseases, Hallym University, Chuncheon, Korea, Rep. of South
2 Hallym University, Department of Electronic Engineering, Chuncheon, Korea, Rep. of South
3 Daewon University College, Department of Nursing, Jaecheon, Korea, Rep. of South

Contact information
Ki Tae Suk, MD, PhD
ktsuk@hallym.ac.kr

EASL CONGRESS Barcelona Spain 27–30 May 2026

Introduction

Cirrhotic Diabetes Mellitus (CDM) — the comorbidity of liver cirrhosis and type 2 diabetes mellitus (T2DM) — is associated with significantly worse clinical outcomes than either condition alone. Conventional taxonomic profiling of the gut microbiome has shown limited ability to capture CDM-specific functional microbial alterations. An interpretable machine learning approach integrating both taxonomic and functional metagenomic data may uncover key biomarkers for this complex phenotype.

Aim

To develop an interpretable ML framework that identifies functional microbial biomarkers distinguishing CDM from single-disease groups, and to evaluate the relative contribution of taxonomic and functional features (KEGG Orthology) in stratifying this high-risk comorbid phenotype.

Method

- **Cohort**: 149 participants — NC (n=59), T2DM (n=46), Cirrhosis (n=30), CDM (n=14). All cirrhosis patients were diagnosed based on clinical, imaging, and/or histological criteria.
- **Sequencing**: Shotgun metagenomic sequencing was performed to simultaneously extract taxonomic profiles (species-level) and functional gene abundances (KEGG Orthology, KO).
- **Feature sets**: Three feature configurations were evaluated — Species only, KO gene only, and Combined (Species + KO) — to assess the relative contribution of taxonomic versus functional information.
- **ML Framework**: XGBoost classifier with Boruta feature selection for robust variable selection in high-dimensional data.
- **Explainability**: SHAP (SHapley Additive exPlanations) was applied to quantify individual feature contributions and visualize model decision logic.
- **Validation**: Data was split into discovery (n=119) and validation (n=30) sets to evaluate generalization performance.

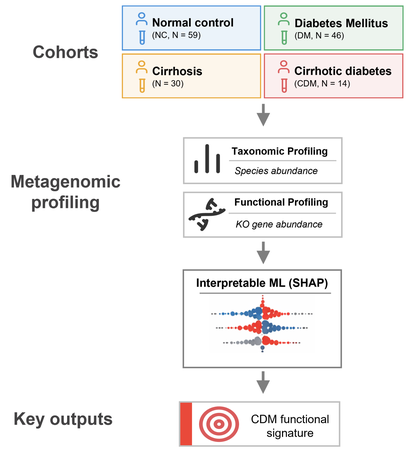

Figure 1. Study design and analytical workflow.

Results

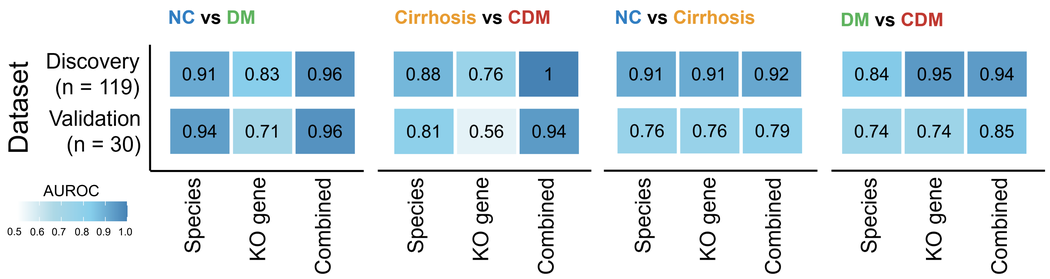

Figure 2. Classification performance of XGBoost across four binary comparisons and three feature sets. Each cell represents the AUROC value. Models were independently trained on Species, KO gene, and Combined (Species + KO) features. Performance was assessed in both discovery (n=119) and validation (n=30) datasets.

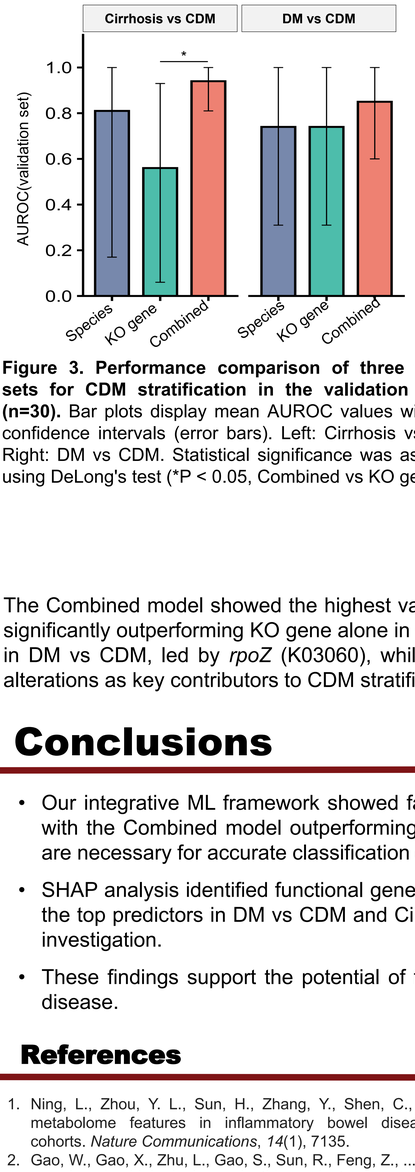

Figure 3. Performance comparison of three feature sets for CDM stratification in the validation cohort (n=30). Bar plots display mean AUROC values with 95% confidence intervals (error bars). Left: Cirrhosis vs CDM. Right: DM vs CDM. Statistical significance was assessed using DeLong's test (*P < 0.05, Combined vs KO gene).

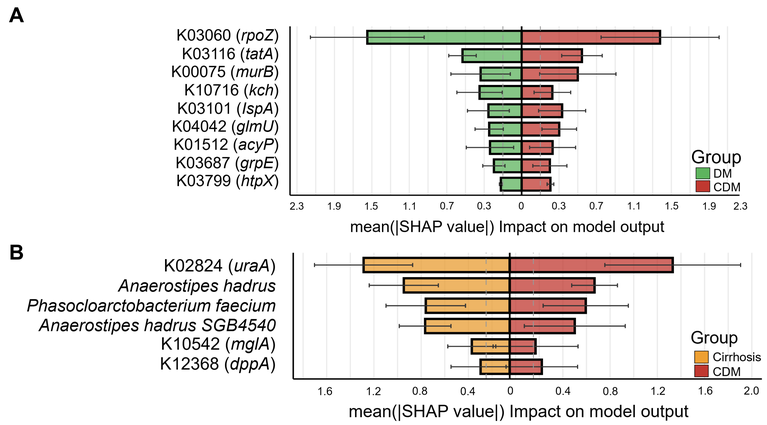

Figure 4. SHAP analysis of top predictive features selected by Boruta in the Combined model across two classification tasks. Bar length represents mean |SHAP value|, indicating each feature's contribution to model output. (A) DM vs CDM. (B) Cirrhosis vs CDM. Italicized labels denote taxonomic features; K-number labels denote functional genes (KO). Colors represent prediction direction toward each group.

The Combined model showed the highest validation AUROC across all comparisons (0.94, Cirrhosis vs CDM; 0.85, DM vs CDM) **(Fig 2)**, significantly outperforming KO gene alone in Cirrhosis vs CDM (*P < 0.05) **(Fig 3)**. SHAP analysis **(Fig 4)** identified exclusively KO features in DM vs CDM, led by *rpoZ* (K03060), while *uraA* (K02824) was the top predictor in Cirrhosis vs CDM — suggesting functional gene alterations as key contributors to CDM stratification.

Conclusions

- Our integrative ML framework showed favorable CDM stratification performance (validation AUROC = 0.94, Cirrhosis vs CDM), with the Combined model outperforming single-feature approaches — indicating that both taxonomic and functional information are necessary for accurate classification of this complex comorbid phenotype.
- SHAP analysis identified functional genes (KO) as key contributors to CDM prediction, with *rpoZ* (K03060) and *uraA* (K02824) as the top predictors in DM vs CDM and Cirrhosis vs CDM, respectively. The biological roles of these genes in CDM warrant further investigation.
- These findings support the potential of function-integrated microbiome profiling for precision risk stratification in advanced liver disease.

References

1. Ning, L., Zhou, Y. L., Sun, H., Zhang, Y., Shen, C., Wang, Z., ... & Hong, J. (2023). Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts. *Nature Communications*, *14*(1), 7135.
2. Gao, W., Gao, X., Zhu, L., Gao, S., Sun, R., Feng, Z., ... & Jiao, N. (2023). Multimodal metagenomic analysis reveals microbial single nucleotide variants as superior biomarkers for early detection of colorectal cancer. *Gut Microbes*, *15*(2), 2245562.

Acknowledgements

The authors thank all participants and staff involved in this study.