



RES-127



Establishment of a MALDI-TOF MS Database for Taiwanese Helicobacter Pylori Isolates and Its Application to Disease Classification and **Antimicrobial Resistance Analysis**

Yun Shan Lu

Departments of Laboratory Medicine, National Taiwan University Hospital, Taiwan

Background

Helicobacter pylori is linked to multiple gastrointestinal diseases. While MALDI-TOF MS is widely used for bacterial identification, existing databases lack sufficient coverage for H. pylori. This study aimed to enhance identification accuracy by building a Taiwan-specific spectral database and applying machine learning to explore associations between strain types and clinical disease categories.

Method

A total of 300 clinical H. pylori isolates were confirmed by 16S rRNA sequencing; 100 were used to build a local spectral database, and 200 for validation. Altogether, 697 MALDI-TOF spectra data from 2001 to 20000 m/z were processed and labeled with its corresponding clinical diagnosis: chronic gastritis, gastric ulcer, duodenal ulcer, gastric cancer, or MALToma, and machine learning models (CatBoost and LightGBM) were developed for disease classification.

Results

Figure 1. Dendrogram comparing 100 clinical *Helicobacter pylori* isolates from NTUH with the Bruker reference library; 27 spectral types absent from the Bruker database were identified.

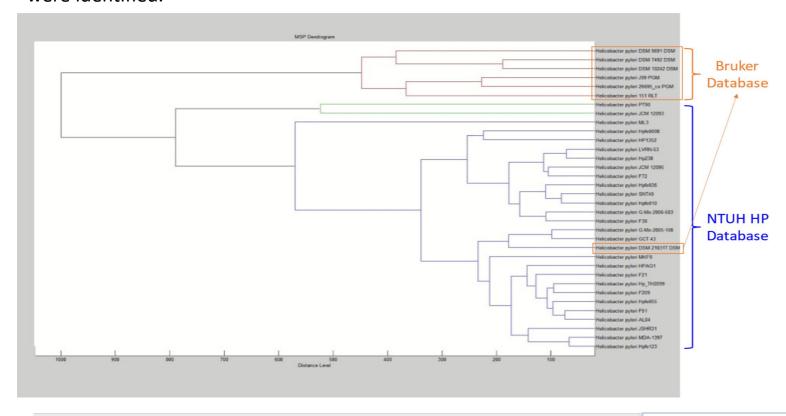


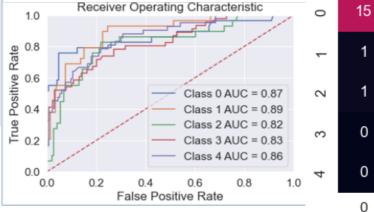
Table 1. When using the Bruker database alone, only 3% of the strains achieved an identification score ≥2.0, while 53% had scores ≥1.7. In contrast, re-analyzing the same 100 isolates with our local database showed a substantial improvement: 93% of identifications scored ≥2.0, and 100% scored ≥1.7.

MALDI-TOF	Bruker database	NTUH HP database		
MS(score)				
1.300~1.399	3	0		
1.400~1.499	9	0		
1.500~1.599	18	0		
1.600~1.699	17	0		
1.700~1.799	22	0		
1.800~1.899	22	2		
1.900~1.999	6	5		
≧2.000	3	93		
≥1.700	53	100		

CatBoost bin_size=30	AUC	accuracy	avg. sensitivity	avg. specificity	avg. precision	avg. F1
Train	0.9265	0.7033	0.7219	0.6879	0.6879	0.6958
Test	0.8549	0.6229	0.6724	0.6048	0.6048	0.6189

Figure 2. Performance of the five-class (gastric cancer, duodenal ulcer, MALToma, chronic gastritis, and gastric ulcer) CatBoost model and ROC curve achieved an AUC of 0.85 with a sensitivity of 67.0% on the test set.

CatBoost bin_size=30	AUC	accuracy	avg. sensitivity	avg. specificity	avg. precision	avg. F1
Train	0.9402	0.8742	0.9508	0.7563	0.8571	0.9016
Test	0.8438	0.8231	0.9494	0.6275	0.7979	0.8671



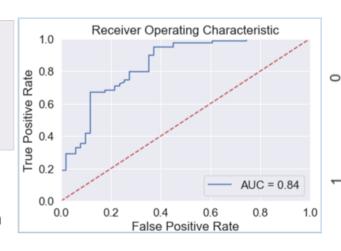




Figure 3. Performance of the binary CatBoost model for classifying chronic gastritis vs peptic ulcer (gastric + duodenal) achieved an AUC of 0.84 and a sensitivity of 94.9%, demonstrating strong classification performance.

Conclusions

Integrating MALDI-TOF MS with machine learning enables accurate classification of H. pyloriassociated diseases and enhances bacterial identification. Establishing a local spectral database provides a foundation for future clinical applications, including antimicrobial resistance prediction and personalized treatment strategies.