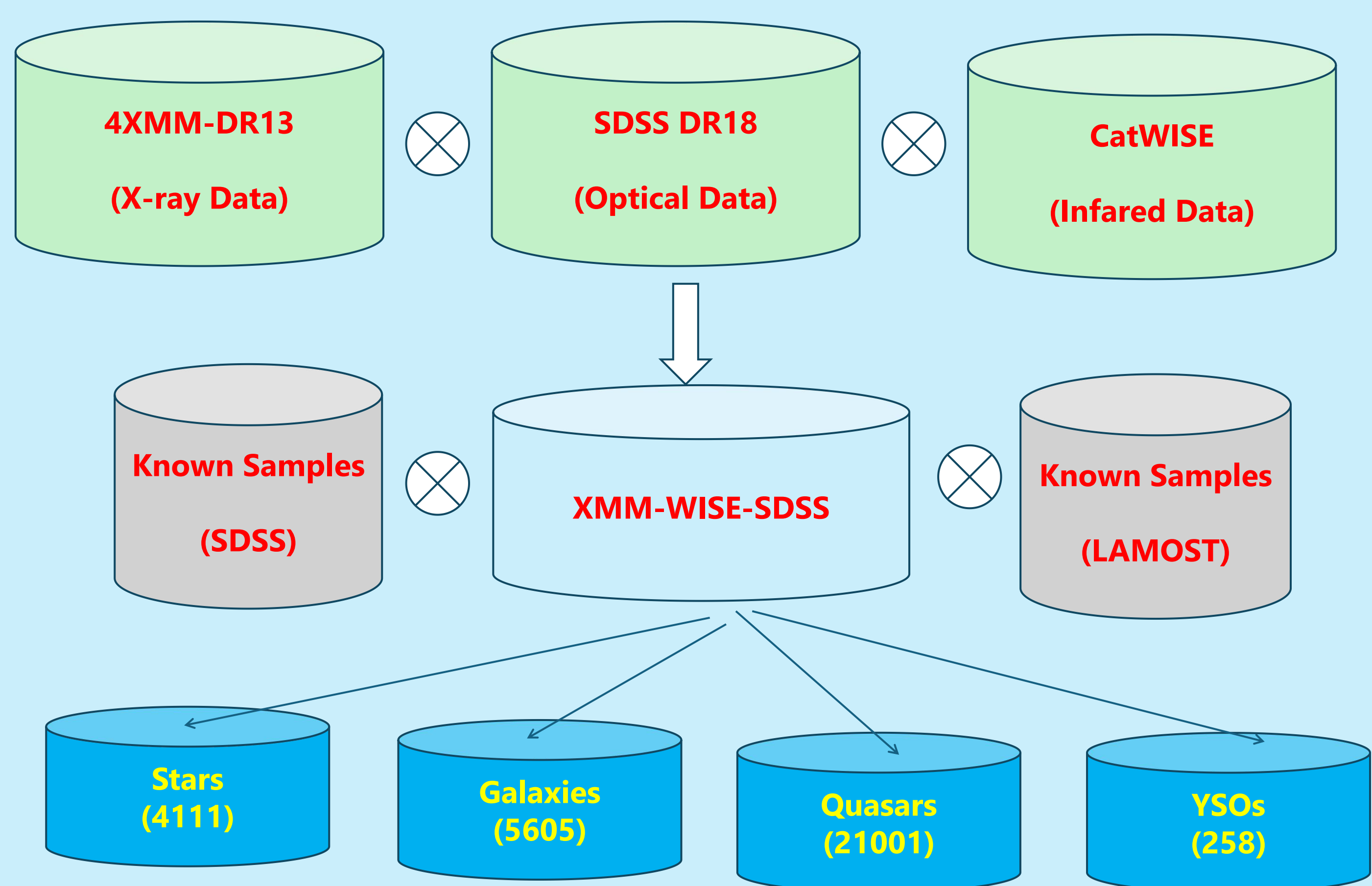# Targeting YSOs from Large Sky Surveys by Machine Learning

Yanxia Zhang, Ma Xiangyao, Jingyi Zhang, Changhua Li

National Astronomical Observatories, CAS, China

Classifying and summarizing extensive datasets from diverse sky surveys is essential for advancing astronomical research. Integrating data from 4XMM-DR13 (X-ray), SDSS DR18 (optical), and CatWISE (IR) surveys, we constructed the XMM-WISE-SDSS sample. Cross-matching with SDSS/LAMOST spectral classifications provided a training set of stars, galaxies, quasars, and young stellar objects (YSOs). We classified the full sample using CatBoost and Self-Paced Ensemble (SPE) machine learning. The SPE classifier excelled in YSO identification, detecting 1,102 YSO candidates—including 258 known YSOs. Verification using LAMOST spectra and SIMBAD/VizieR databases confirmed 412 new YSO candidates. These discoveries substantially expand the known YSO sample, enabling deeper studies of star formation and evolution. A comprehensive classification catalog for the XMM-WISE-SDSS sample is provided (http://paperdata.china-vo.org/mxy/table3.csv).
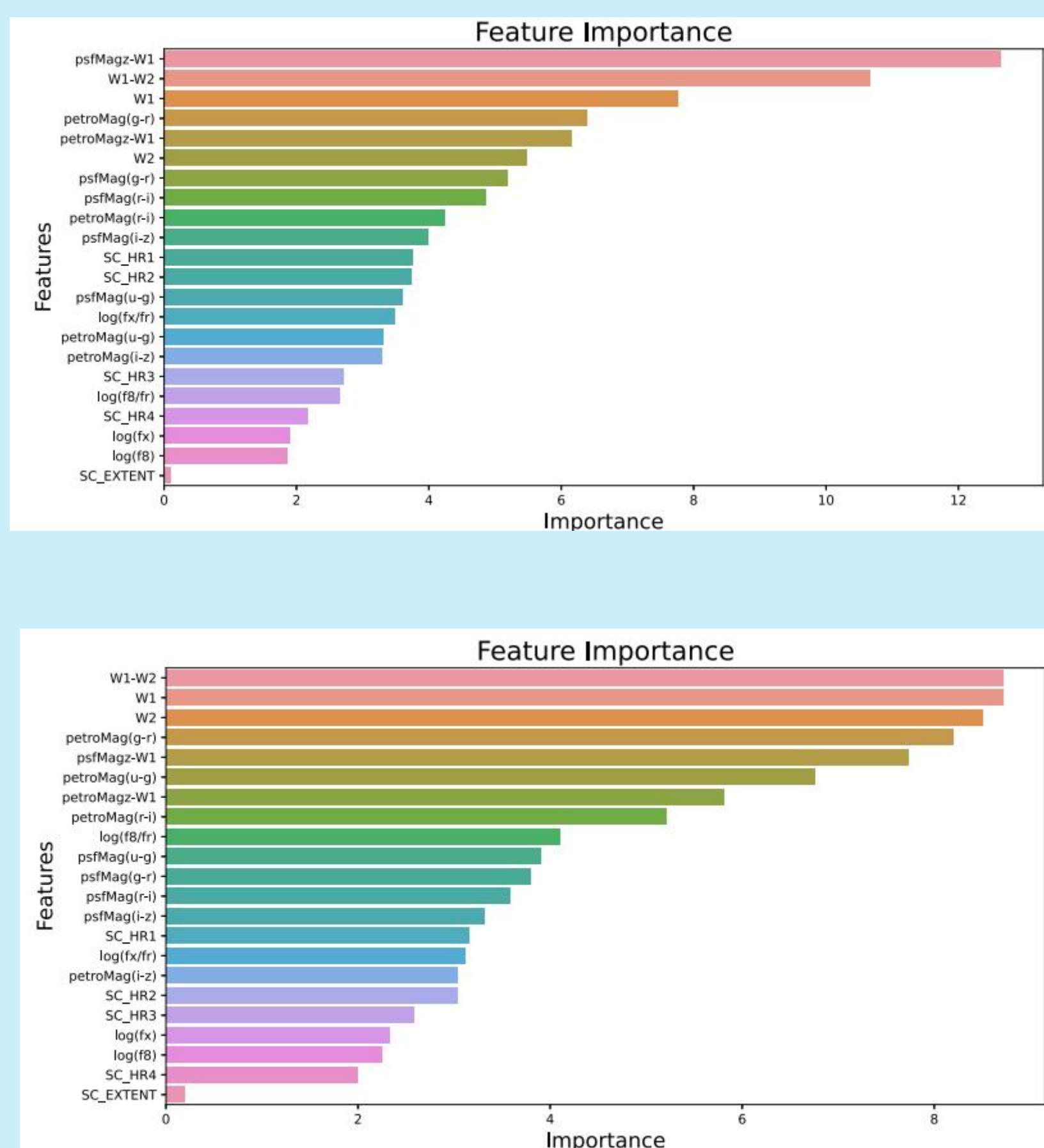
Data

Methods

Catboost

Self-Paced Ensemble (SPE)

Feature Importance

Results

The spectrum of LAMOST LRS spec-56617-VB081S05V2_sp06-073.fits. It shows strong Hα emission at 6562.8 Å.

The spectrum of LAMOST MRS med-59980-TD043724N254338B01_sp14-231.fits. It shows strong Hα emission at 6562.8 Å.
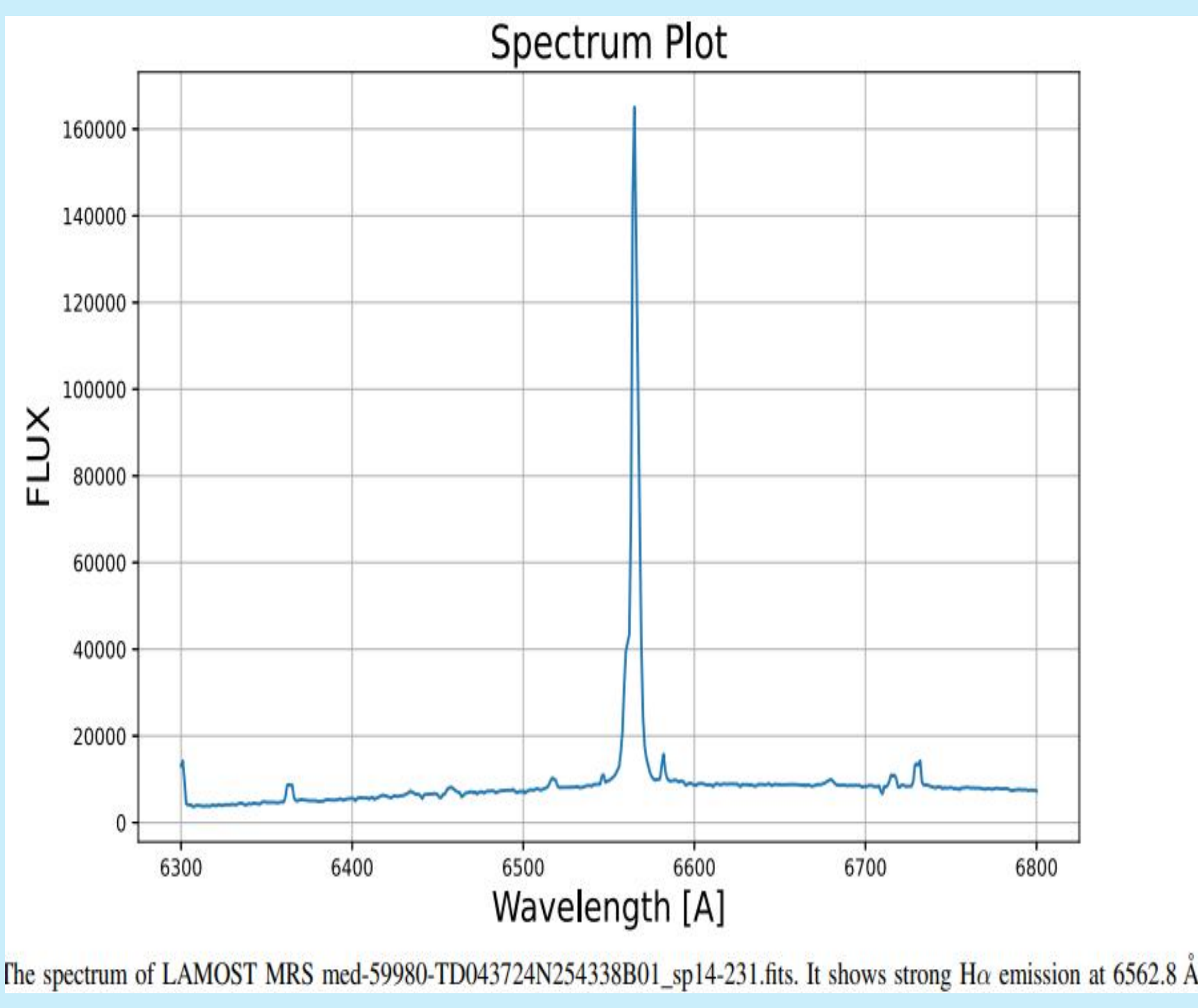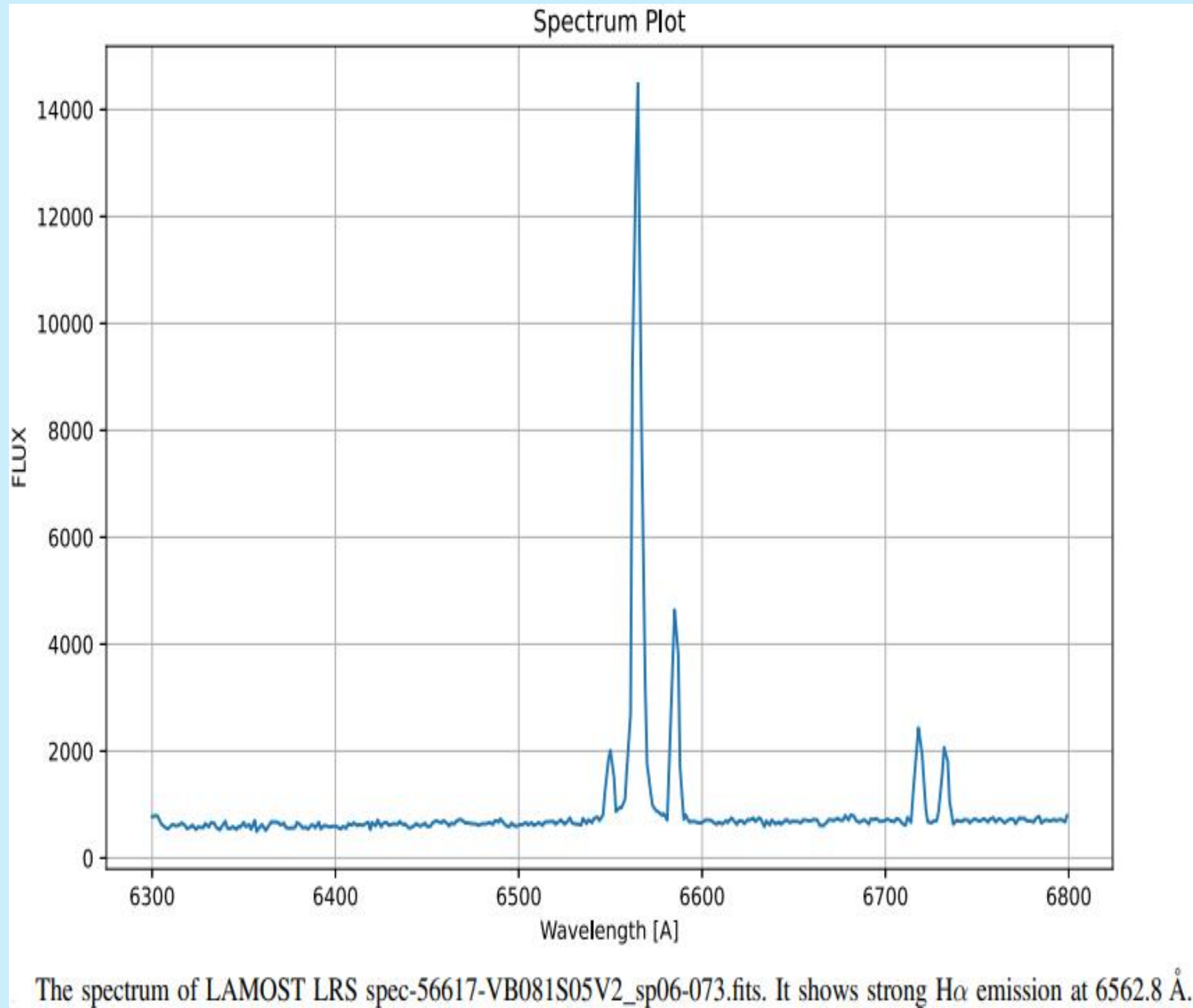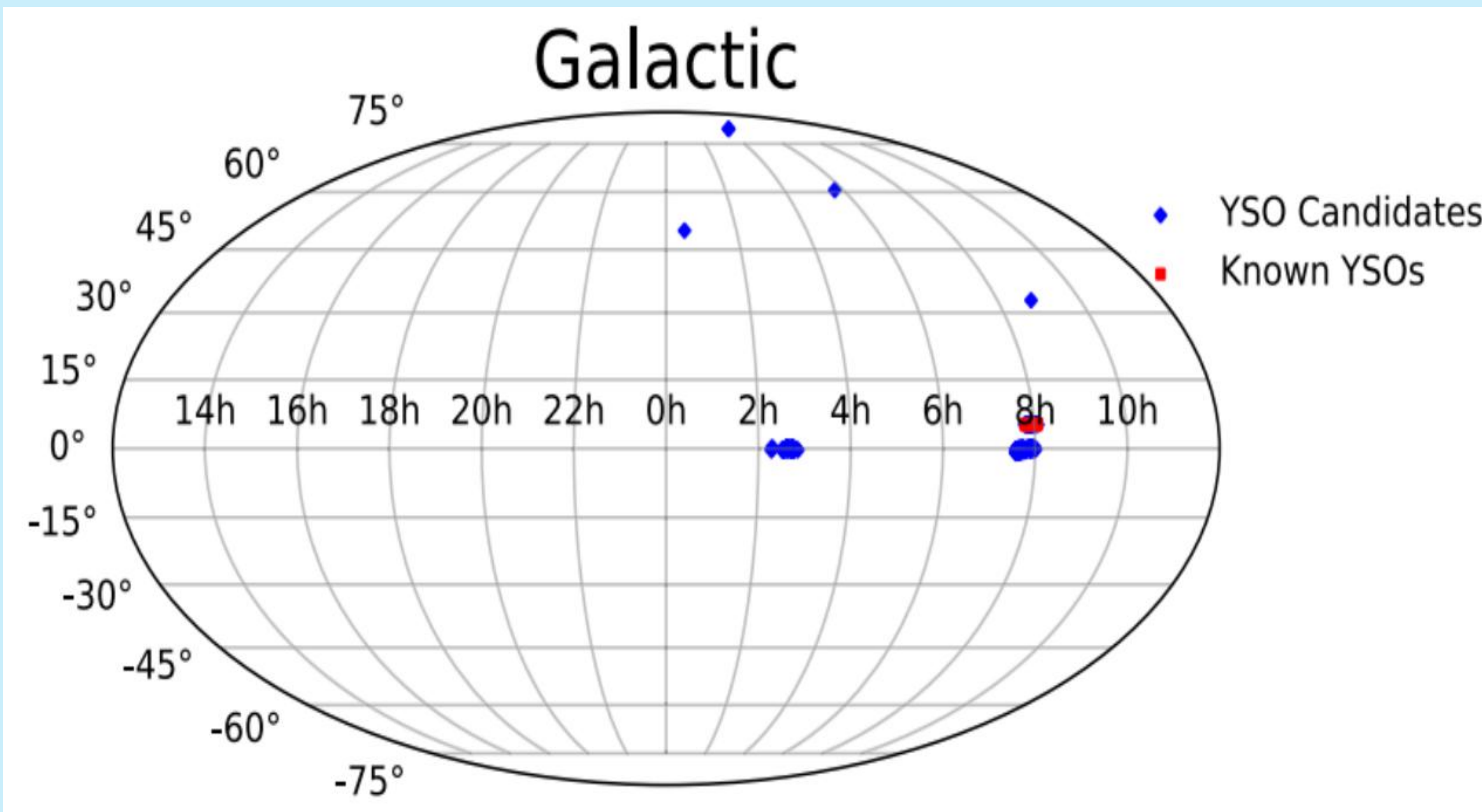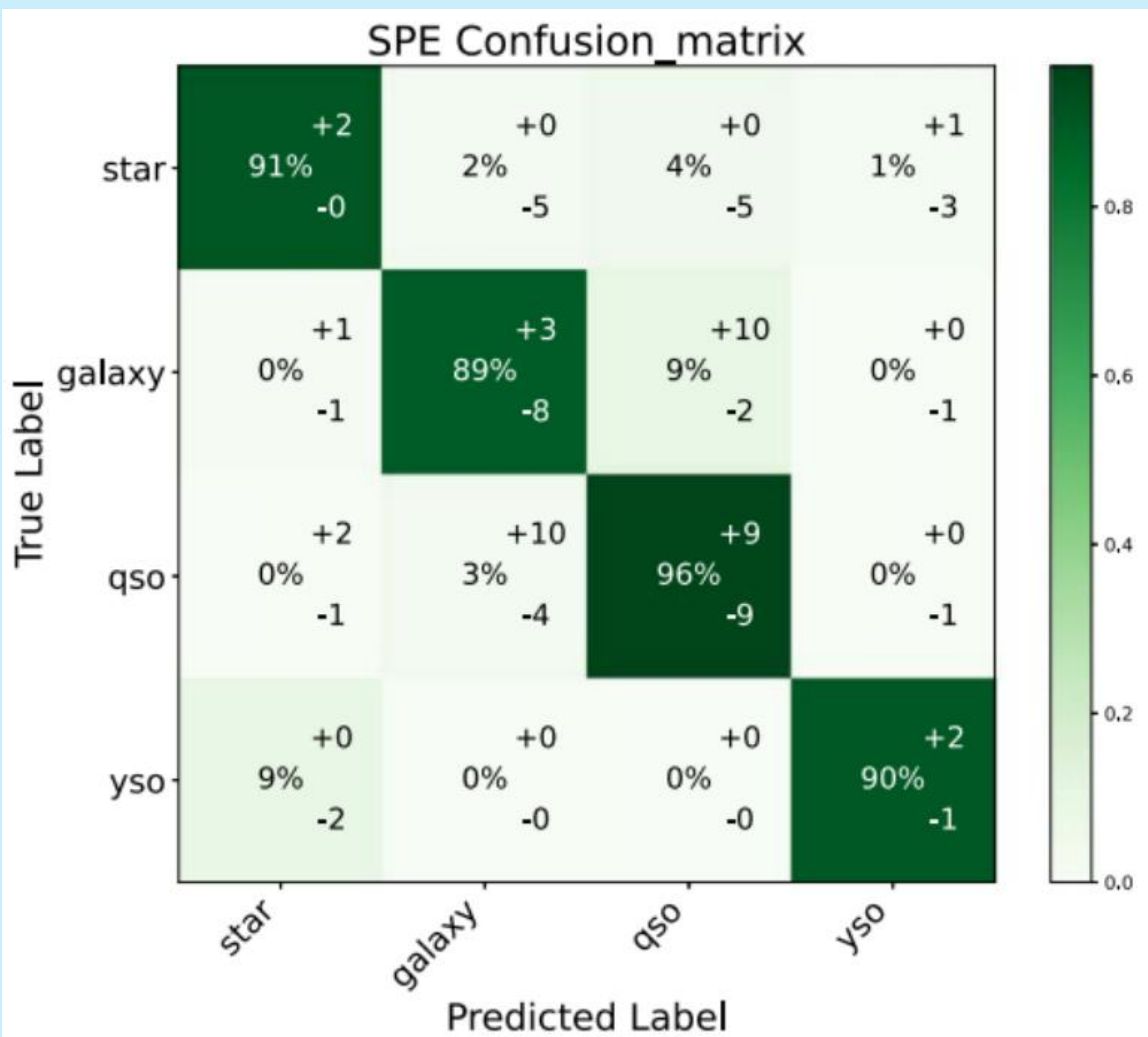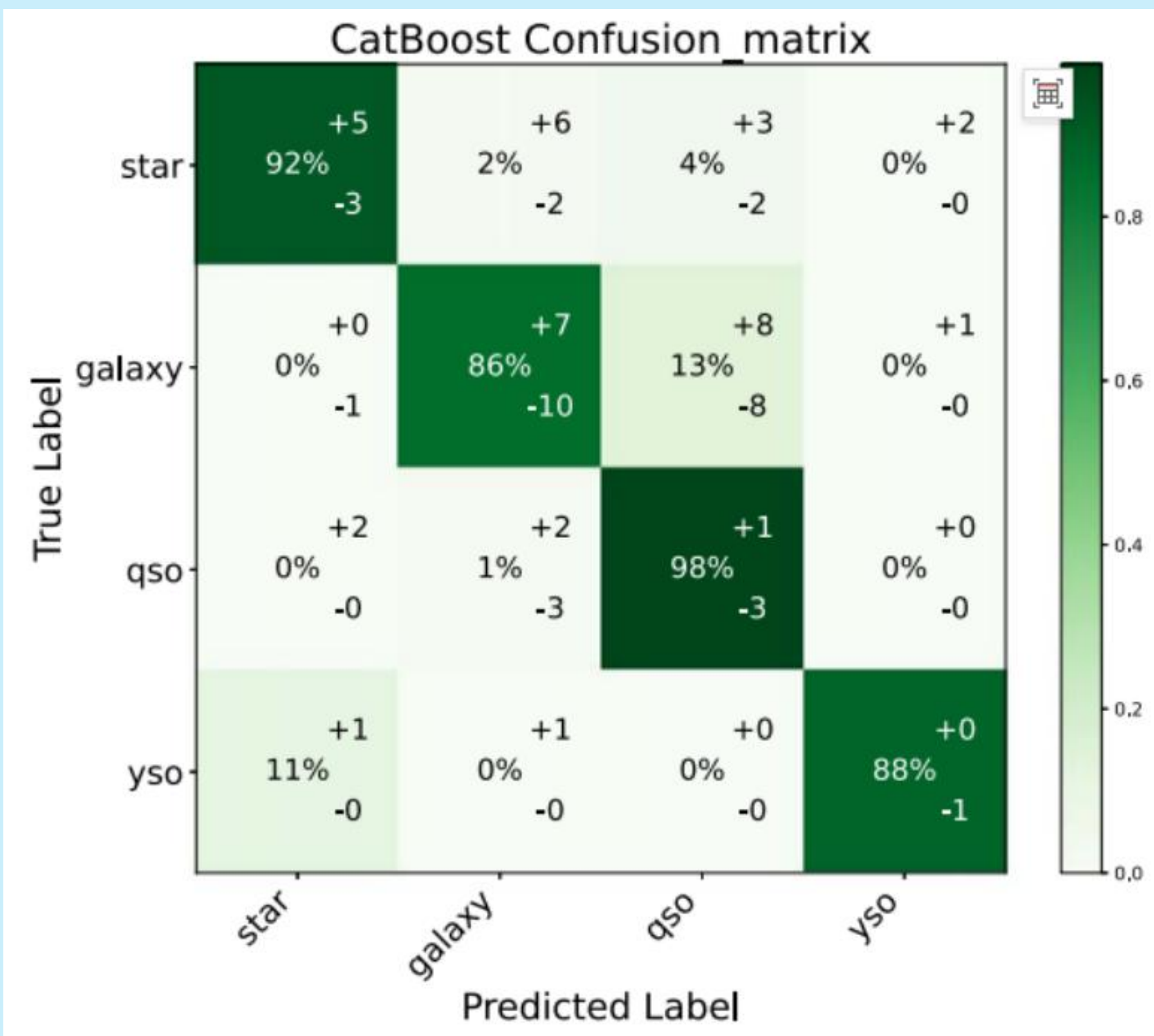
Table 3
The Classification Result of the XMM-WISE-SDSS Sample by CatBoost and SPE

SRCID	SC_RA	SC_DEC	Class_CatBoost	P_CatBoost	Class_SPE	P_SPE
2065168030010002	281.636194	−2.353715	STAR	0.543	YSO	0.784
2016456020100018	86.785682	0.00457154	STAR	0.617	YSO	0.632
2079567010100042	226.654933	3.638037	QSO	0.904	QSO	0.536
2001142010100045	100.278052	9.7909098	YSO	0.638	YSO	0.765
2067272010100003	122.242080	20.812036	STAR	0.998	STAR	0.941
2020339050100028	166.078129	40.7661459	GALAXY	0.967	GALAXY	0.543
2020339060100006	229.405224	42.624143	QSO	0.989	QSO	0.710
2001142020100078	100.4326076	9.6805009	YSO	0.694	YSO	0.812
2001142010100236	100.356461	9.734582	YSO	0.599	YSO	0.775

Contact: zyx@bao.ac.cn