2025 NSRI Workshop

Simulation-Based Machine Learning to Classify Stellar Origins in Galaxies from Observations

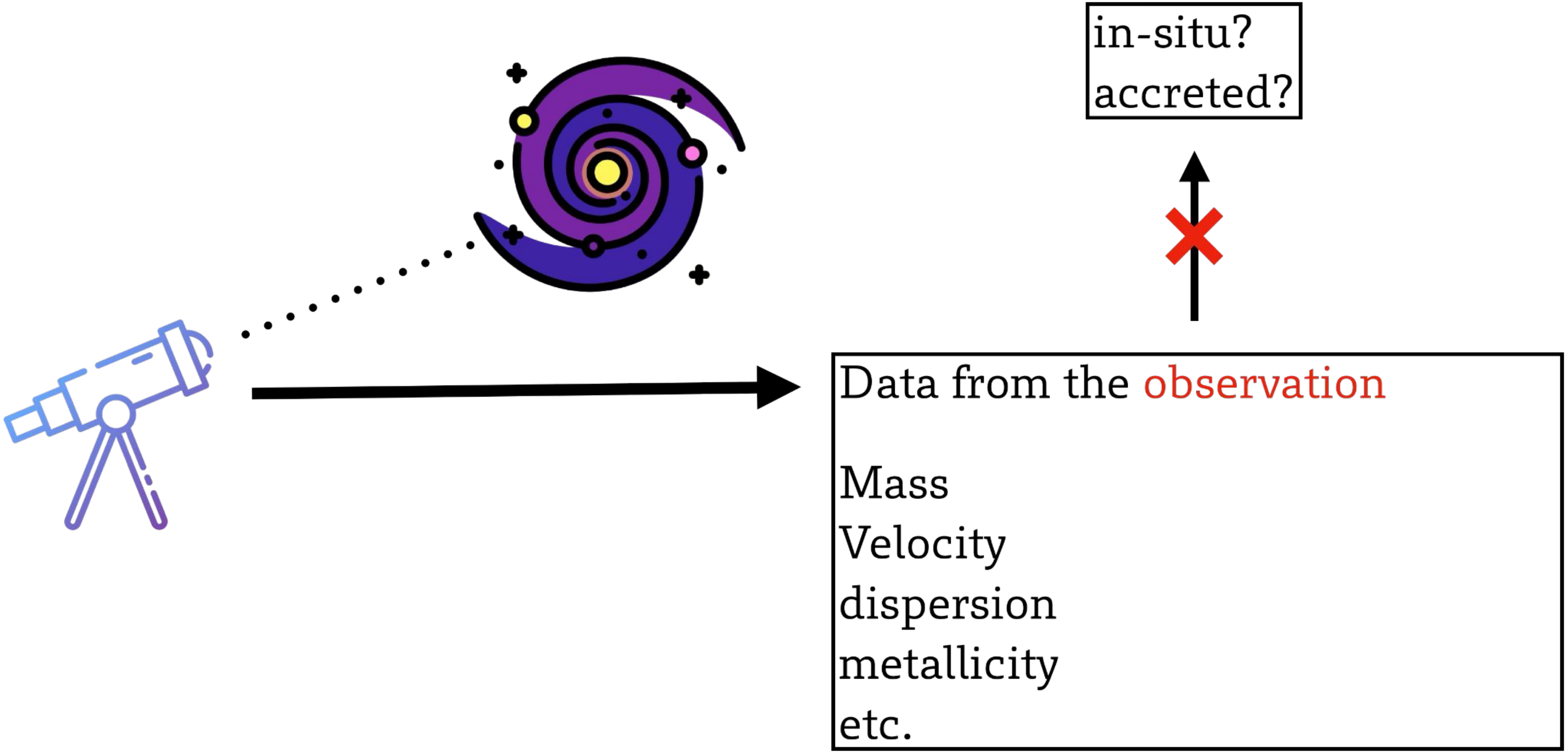
Minki Jeong

minkijeong@uos.ac.kr

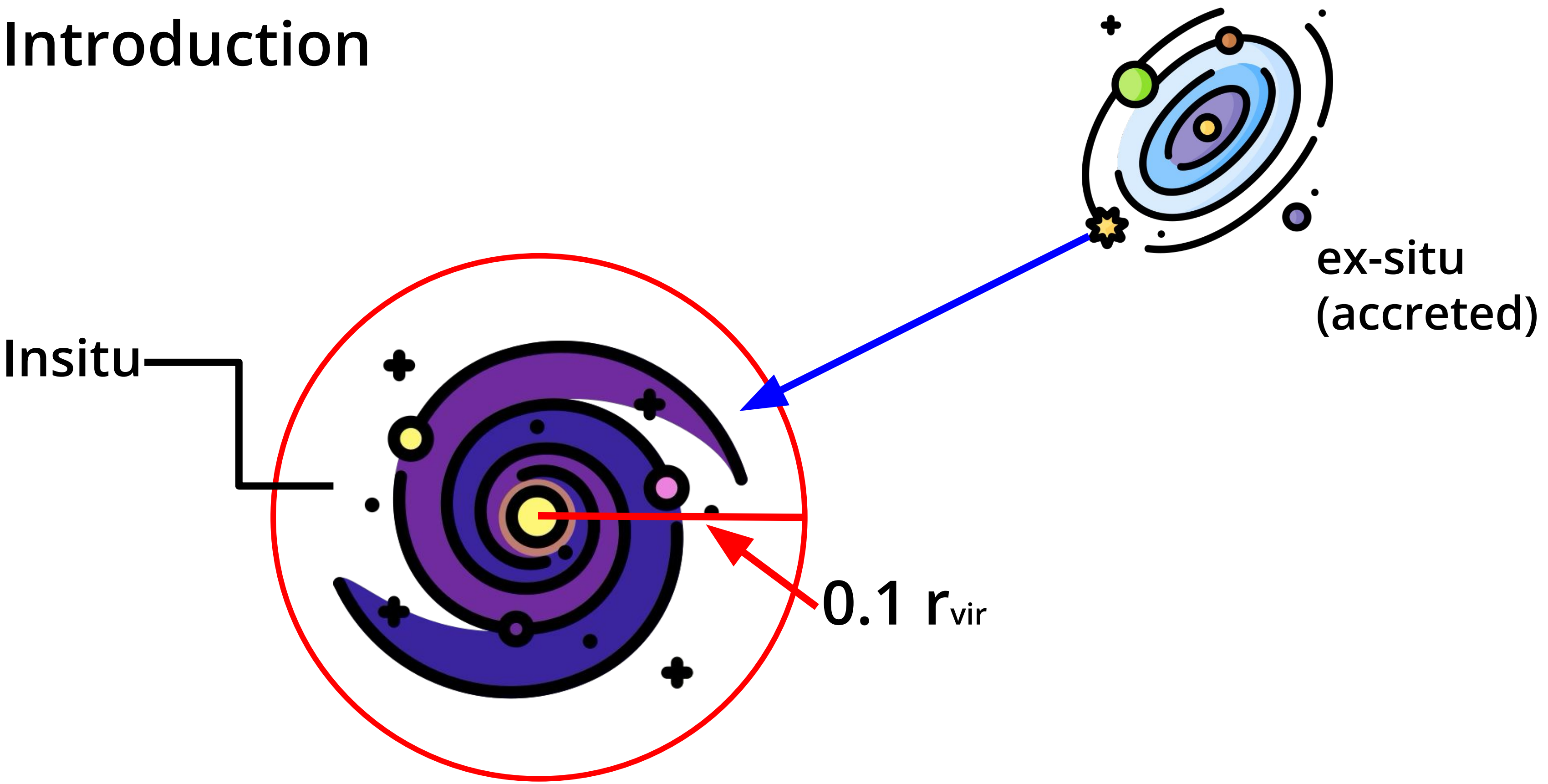
Outline

1. Introduction
2. Machine Learning(using GADGET3)
3. Machine Learning(using IllustrisTNG)
4. Future Work

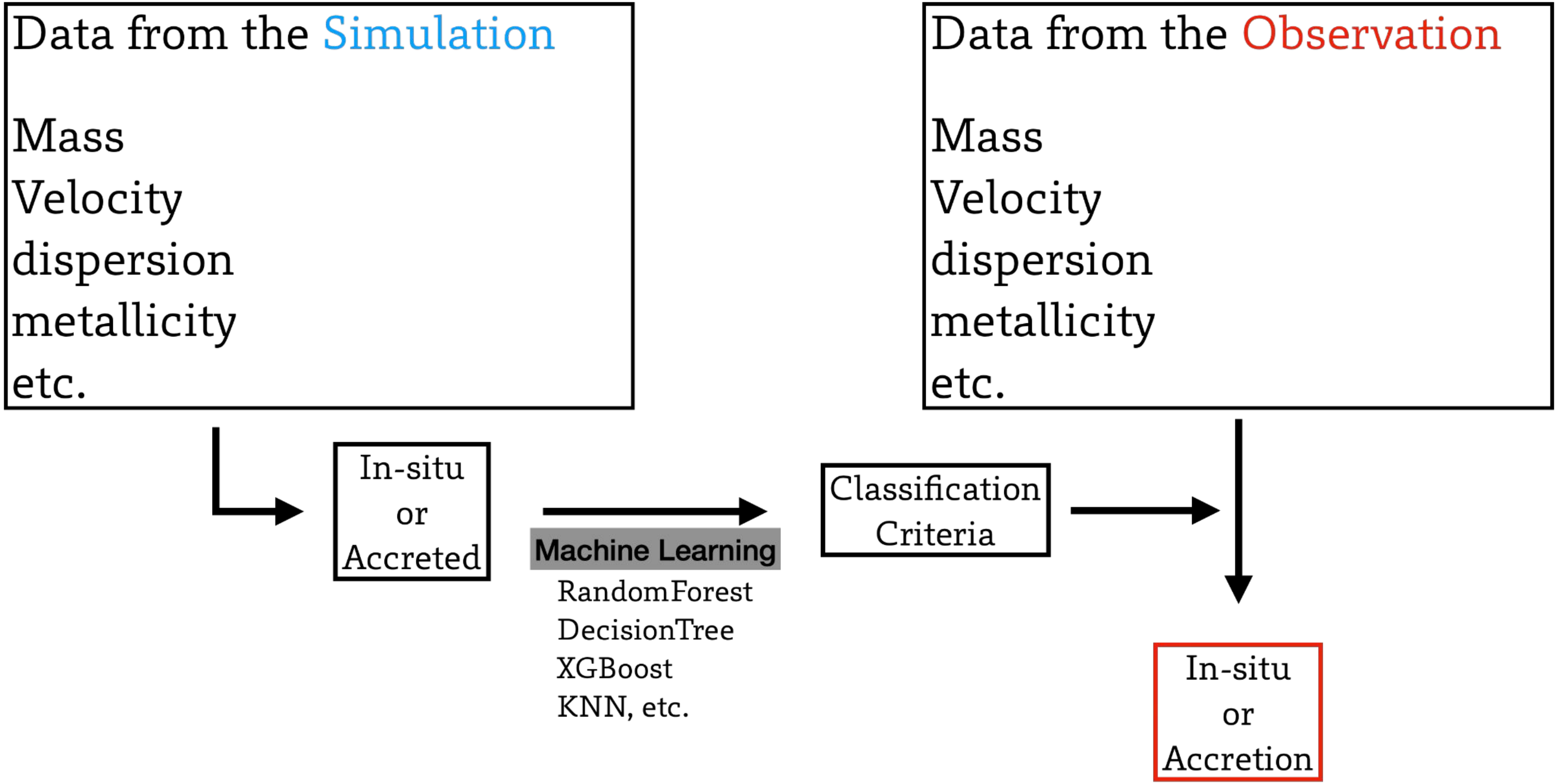
Introduction



Introduction



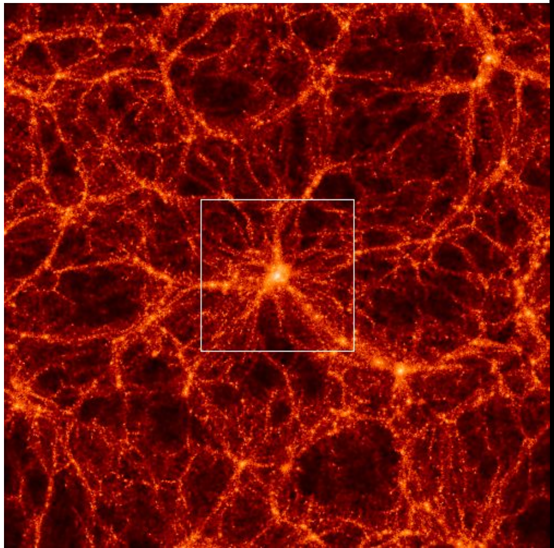
Introduction

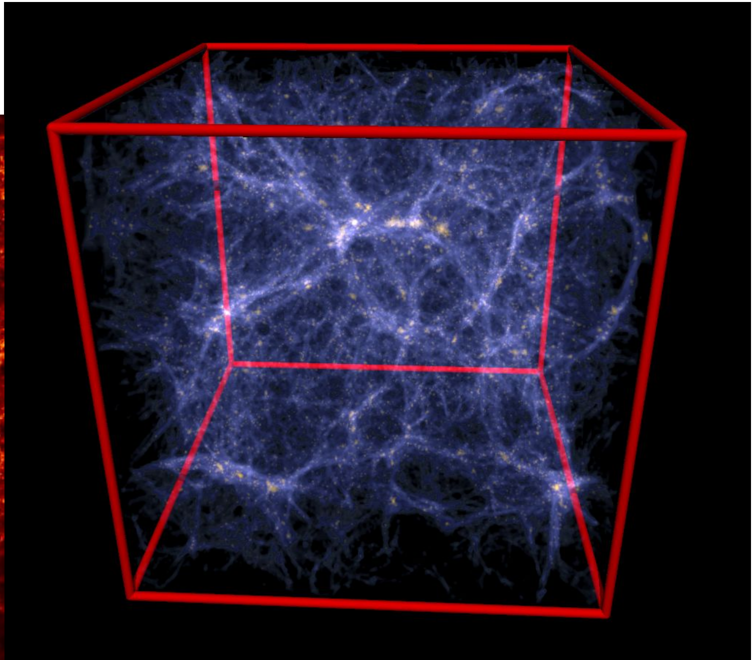


Introduction

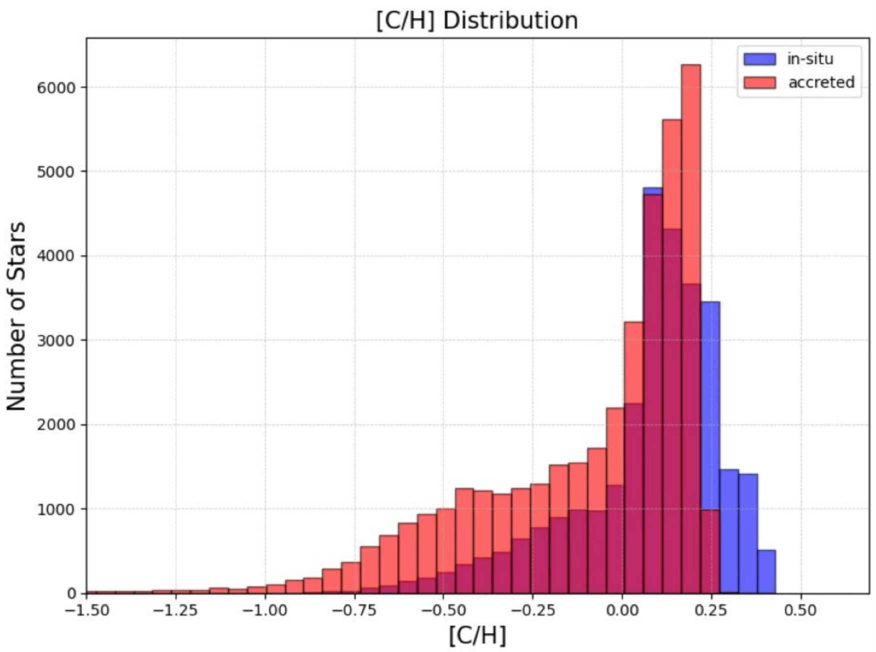
GADGET3

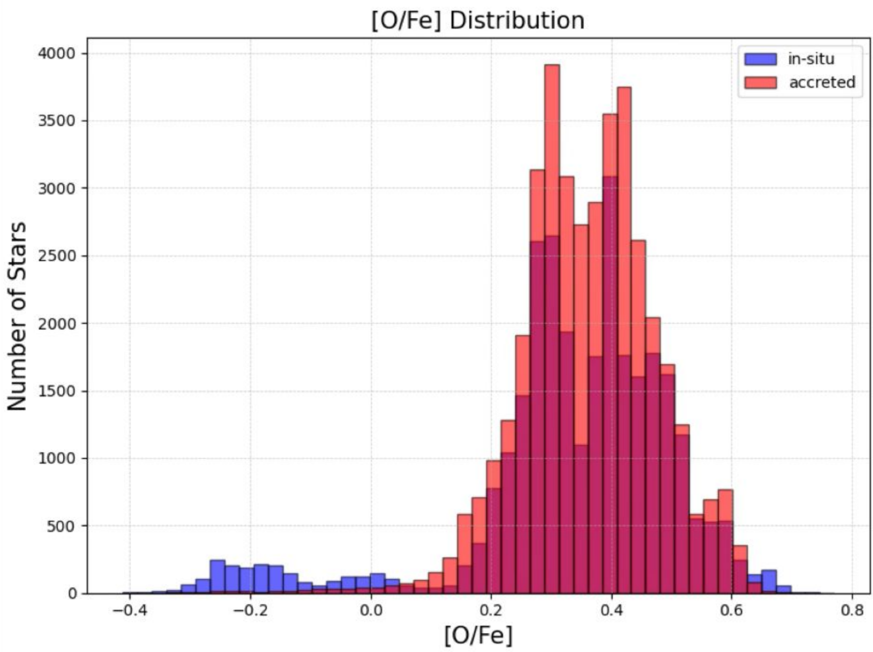
Cosmological N-body + Hydrodynamics Simulation Code
SPH(Smoothed-Particle Hydrodynamics)





Machine Learning(using GADGET3)





Machine Learning(using GADGET3)

Model	Train Acc	Val Acc	Test Acc	Precision	Recall	F1-Score
RF(50,3)	0.678	0.679	0.691	0.661	0.979	0.789
RF(100,3)	0.678	0.678	0.690	0.661	0.979	0.789
RF(50,5)	0.715	0.707	0.719	0.692	0.948	0.800
RF(100,5)	0.716	0.708	0.718	0.692	0.945	0.799
RF(50,10)	0.842	0.820	0.822	0.786	0.962	0.865
RF(100,10)	0.845	0.821	0.827	0.790	0.964	0.868

Machine Learning(using GADGET3)

	G01	G15	G17	G18	G19	G23	G24	G25	G27	G38	G40	G44
G01	0.68	0.60	0.52	0.66	0.44	0.60	0.59	0.45	0.81	0.48	0.76	0.66
G15	0.54	0.77	0.48	0.67	0.58	0.62	0.67	0.44	0.74	0.62	0.66	0.66
G17	0.58	0.64	0.62	0.62	0.62	0.70	0.61	0.47	0.73	0.59	0.67	0.71
G18	0.55	0.66	0.45	0.76	0.46	0.50	0.65	0.46	0.78	0.48	0.74	0.57
G19	0.54	0.67	0.47	0.64	0.74	0.71	0.66	0.44	0.73	0.62	0.66	0.65
G23	0.57	0.67	0.52	0.62	0.69	0.82	0.61	0.44	0.76	0.62	0.63	0.72
G24	0.52	0.66	0.43	0.67	0.48	0.56	0.78	0.45	0.79	0.53	0.75	0.61
G25	0.47	0.43	0.44	0.58	0.44	0.46	0.39	0.61	0.62	0.32	0.46	0.58
G27	0.48	0.49	0.35	0.63	0.25	0.36	0.47	0.42	0.88	0.53	0.76	0.50
G38	0.51	0.62	0.47	0.63	0.45	0.50	0.52	0.45	0.77	0.76	0.70	0.65
G40	0.47	0.43	0.33	0.63	0.16	0.29	0.43	0.39	0.78	0.29	0.83	0.46
G44	0.57	0.64	0.55	0.63	0.64	0.65	0.63	0.46	0.78	0.58	0.73	0.83

Table 2. PR-AUC scores for the MLP models trained on accreted and in-situ examples from the galaxy in the left-most column, and tested on the galaxies listed in the top row. Where the galaxy label is the same, stars in the validation dataset were considered.

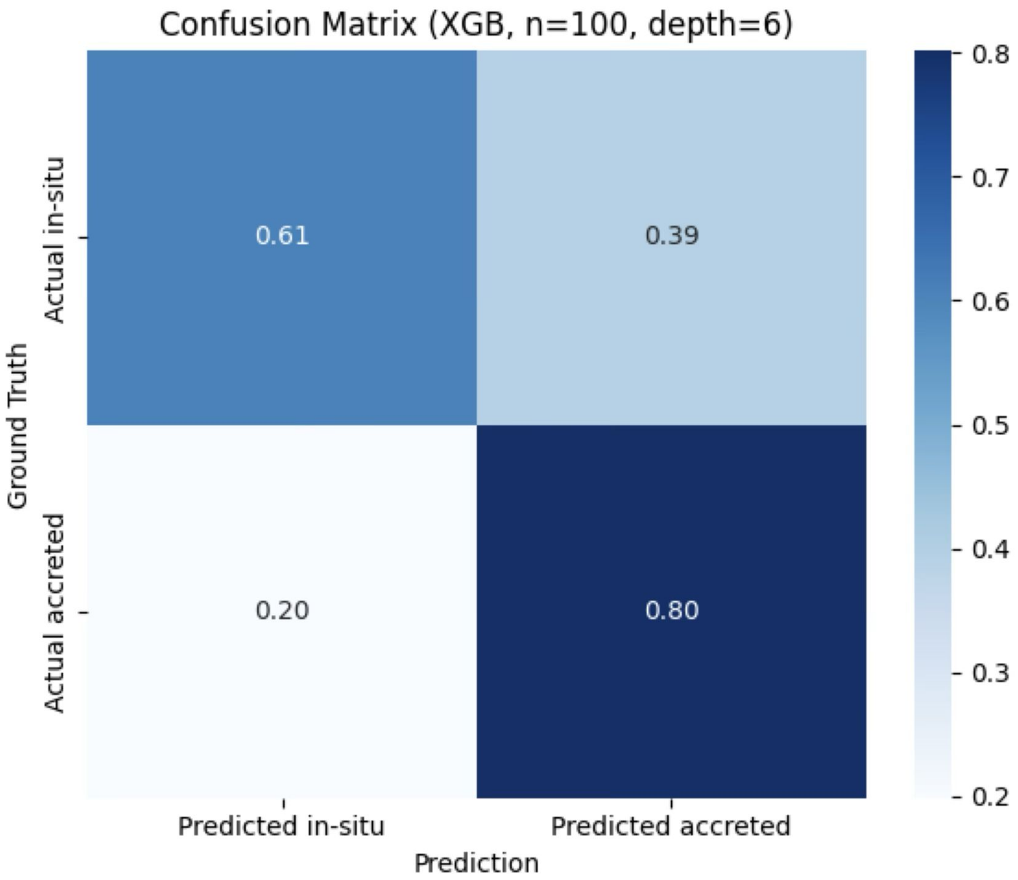
[1] Sante, A., Font, A. S., Ortega-Martorell, S., Olier, I., & McCarthy, I. G. 2024. **Monthly Notices of the Royal Astronomical Society**, **531(4)**, 4363–4382.

Machine Learning(using GADGET3)

Classifier	n_estimators	max_depth	Train	m0125	m0163	m0204
RandomForest	50	3	m0125	0.763	0.685	0.592
			m0163	0.726	0.694	0.578
			m0204	0.671	0.663	0.678
RandomForest	50	5	m0125	0.788	0.693	0.608
			m0163	0.714	0.747	0.588
			m0204	0.663	0.629	0.712
RandomForest	50	10	m0125	0.865	0.675	0.6
			m0163	0.689	0.883	0.589
			m0204	0.671	0.645	0.836
RandomForest	100	3	m0125	0.762	0.684	0.593
			m0163	0.726	0.693	0.578
			m0204	0.673	0.658	0.677
RandomForest	100	6	m0125	0.789	0.692	0.608
			m0163	0.717	0.75	0.59
			m0204	0.663	0.625	0.713
RandomForest	100	10	m0125	0.866	0.675	0.6
			m0163	0.691	0.887	0.592
			m0204	0.673	0.646	0.833

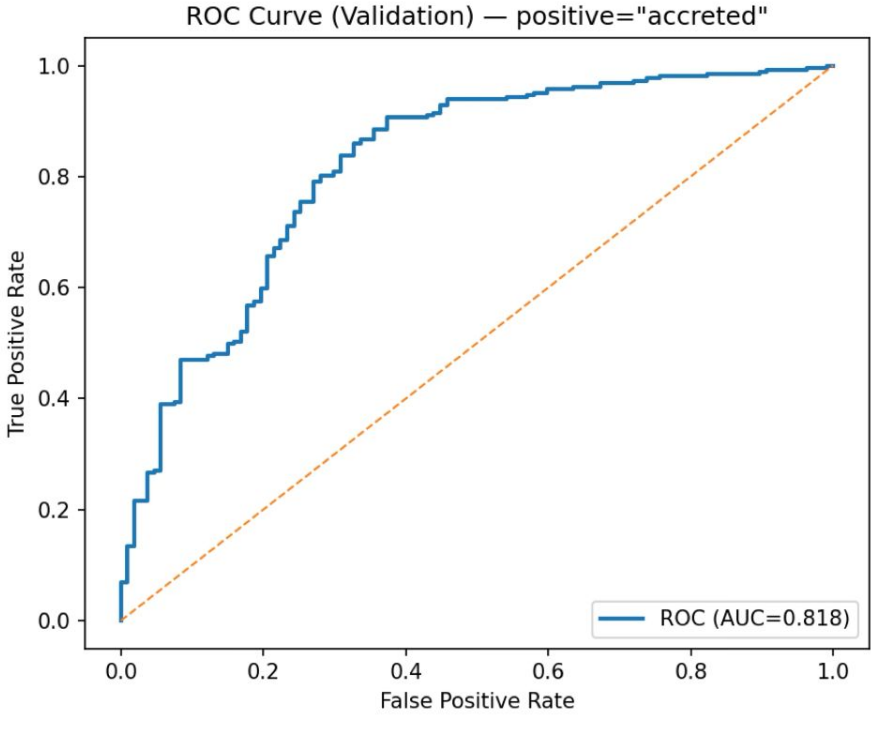
Classifier	n_estimators	max_depth	Train	m0125	m0163	m0204
XGBoost	50	3	m0125	0.856	0.676	0.596
			m0163	0.642	0.907	0.581
			m0204	0.661	0.655	0.791
XGBoost	50	5	m0125	0.899	0.669	0.591
			m0163	0.649	0.956	0.592
			m0204	0.655	0.65	0.857
XGBoost	50	10	m0125	0.941	0.625	0.58
			m0163	0.662	0.974	0.598
			m0204	0.635	0.631	0.924
XGBoost	100	3	m0125	0.881	0.676	0.599
			m0163	0.641	0.935	0.589
			m0204	0.657	0.668	0.827
XGBoost	100	6	m0125	0.923	0.651	0.588
			m0163	0.662	0.966	0.598
			m0204	0.651	0.648	0.887
XGBoost	100	10	m0125	0.947	0.616	0.578
			m0163	0.659	0.976	0.604
			m0204	0.635	0.632	0.934

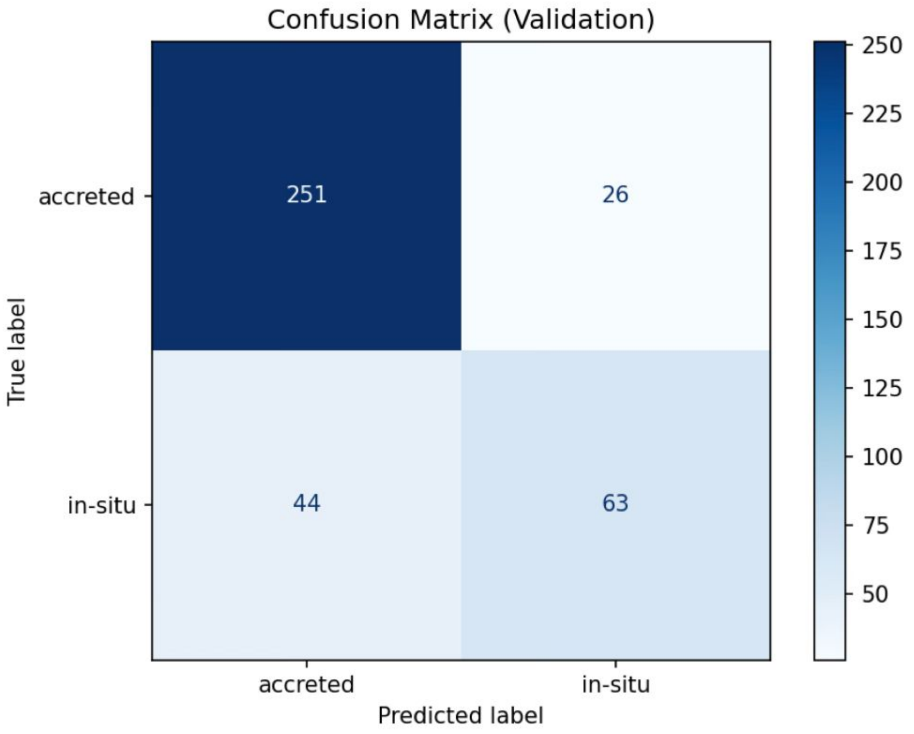
Machine Learning(using GADGET3)



Machine Learning(using GADGET3)

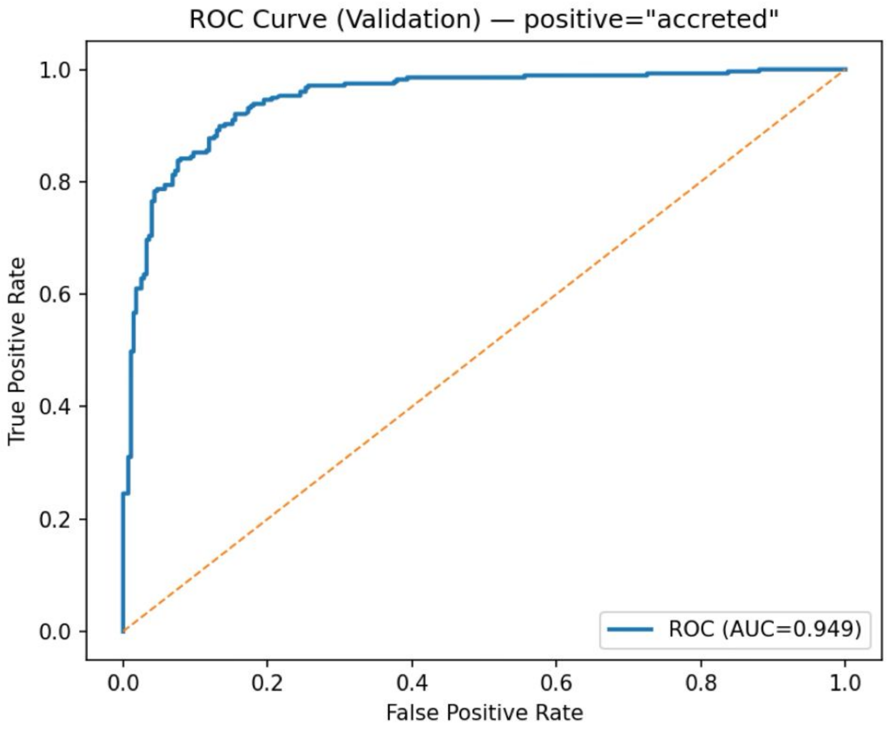
Train Acc	Val Acc	Test Acc	Val AUC	Test AUC	Val Precision	Test Precision	Val Recall	Test Recall
0.9978	0.8177	0.7461	0.8181	0.7358	0.8508	0.7922	0.9061	0.8777

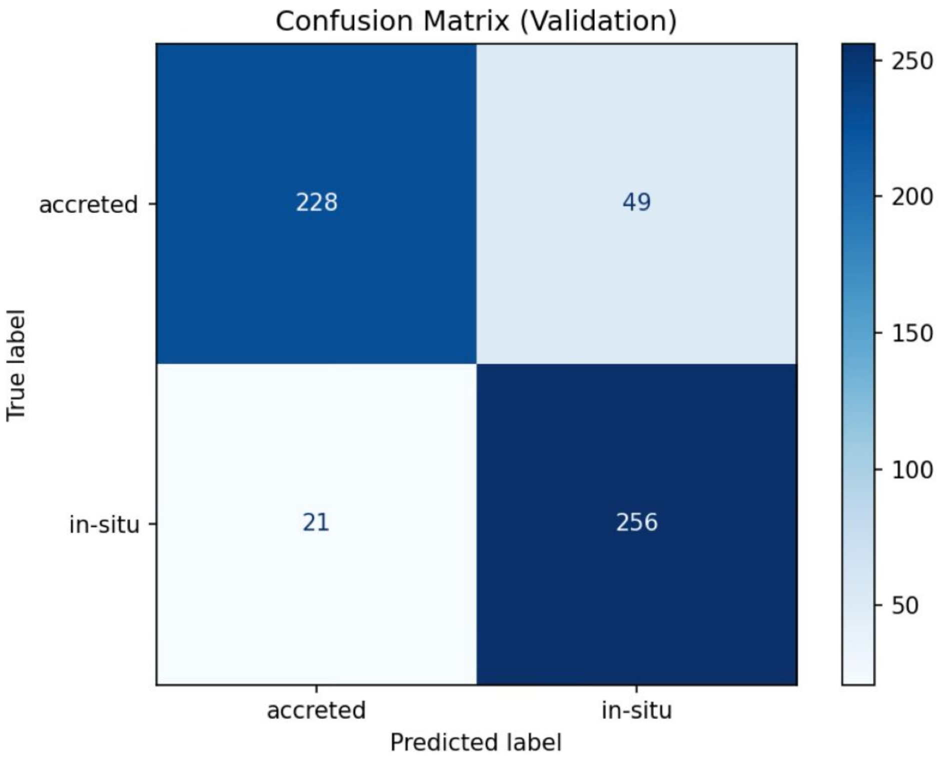




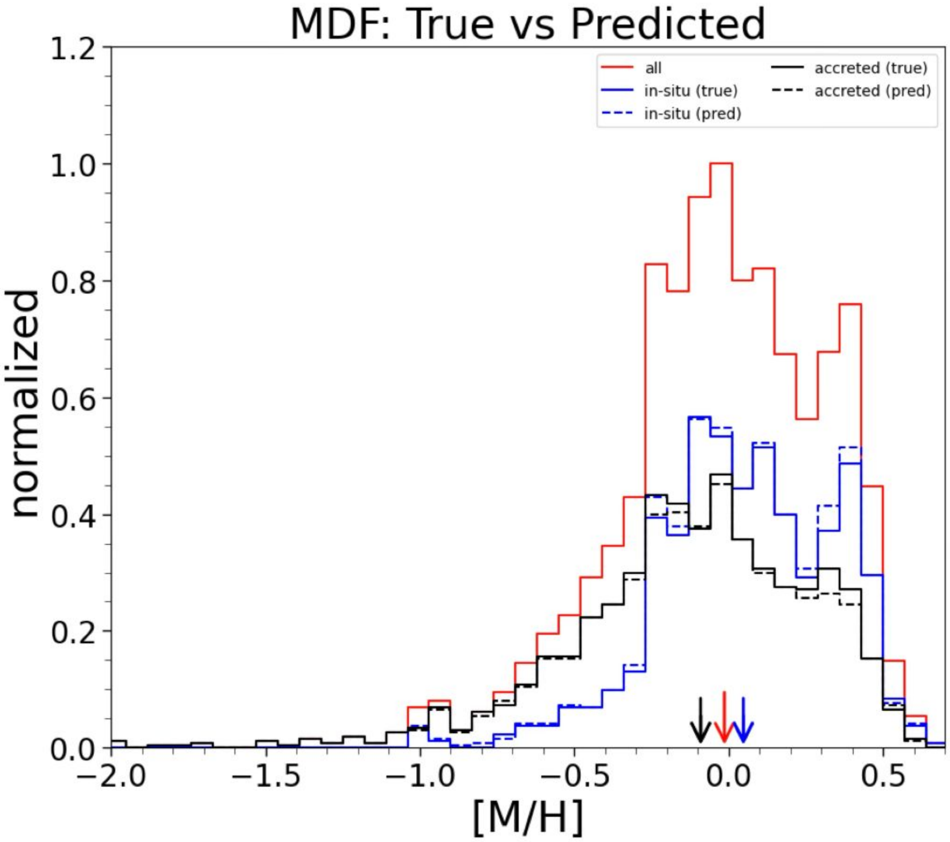
Machine Learning(using GADGET3)

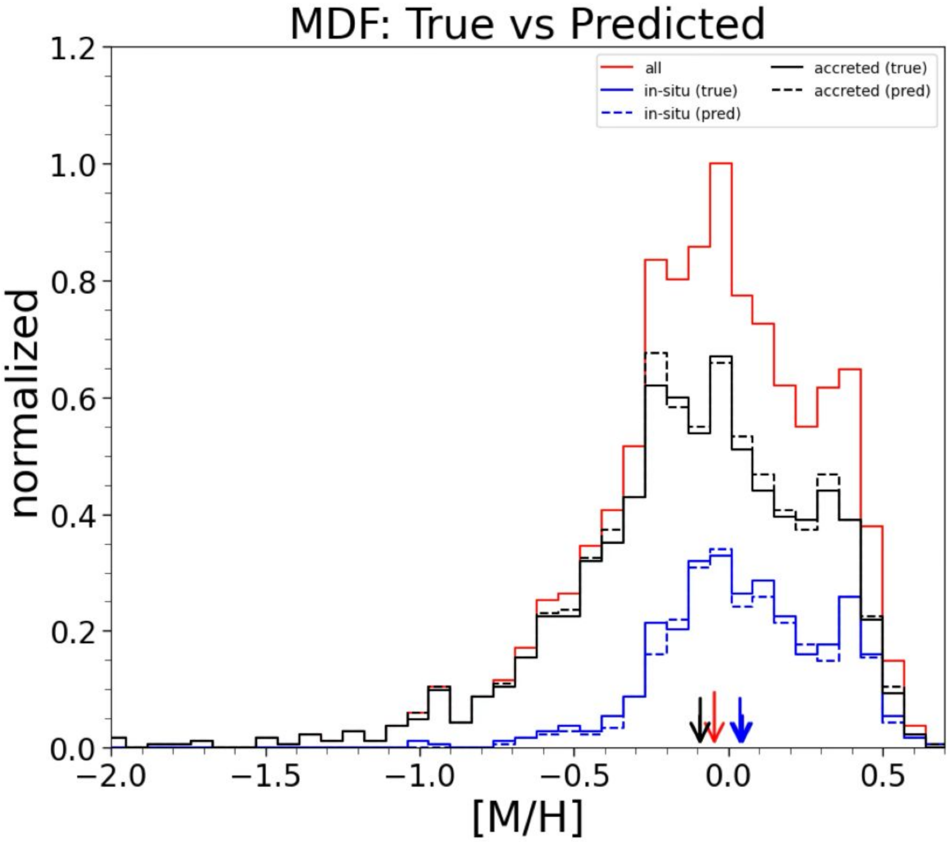
Train Acc	Val Acc	Test Acc	Val AUC	Test AUC	Val Precision	Test Precision	Val Recall	Test Recall
0.9979	0.8736	0.8412	0.9487	0.9241	0.9157	0.8992	0.8231	0.7698





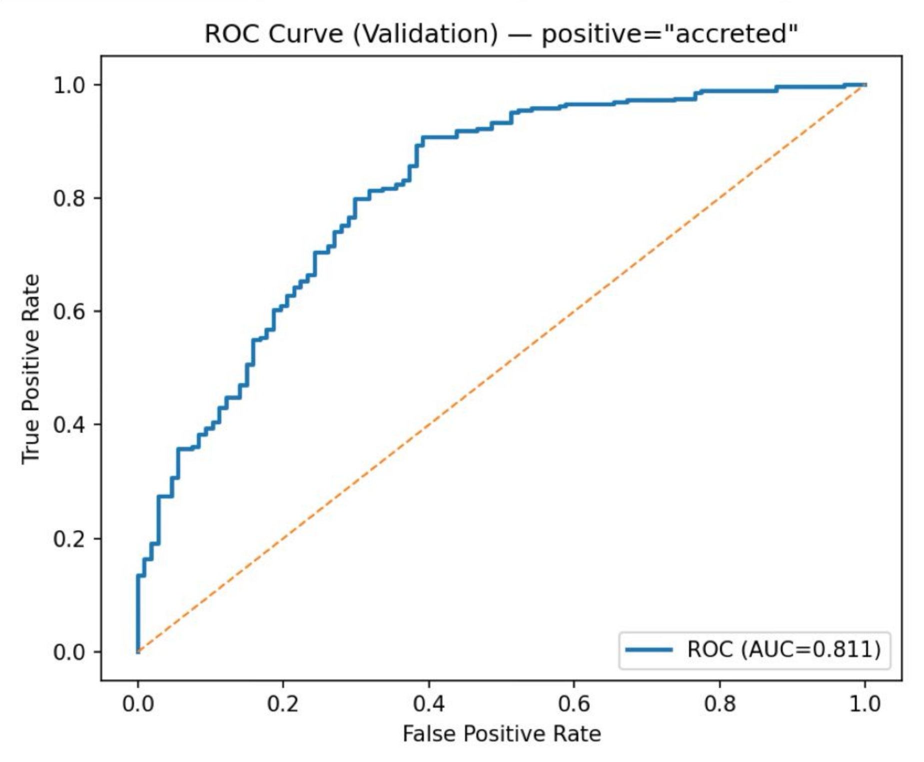
Machine Learning(using GADGET3)

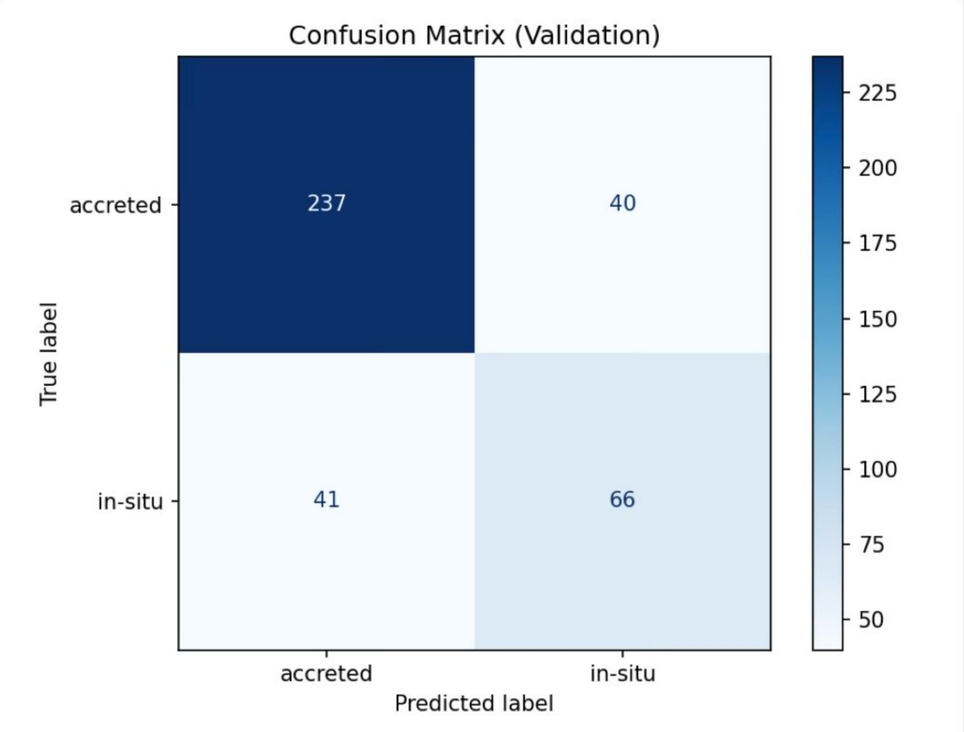




Machine Learning(using GADGET3)

Train Acc	Val Acc	Test Acc	Val AUC	Test AUC	Val Precision	Test Precision	Val Recall	Test Recall
0.9963	0.7891	0.7254	0.8112	0.7514	0.8525	0.7986	0.8556	0.8273



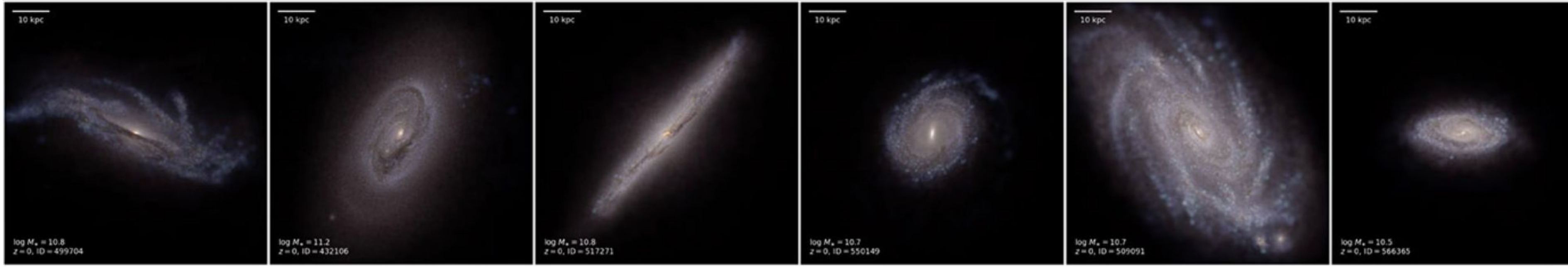


Machine Learning(using IllustrisTNG)

TNG 50-1

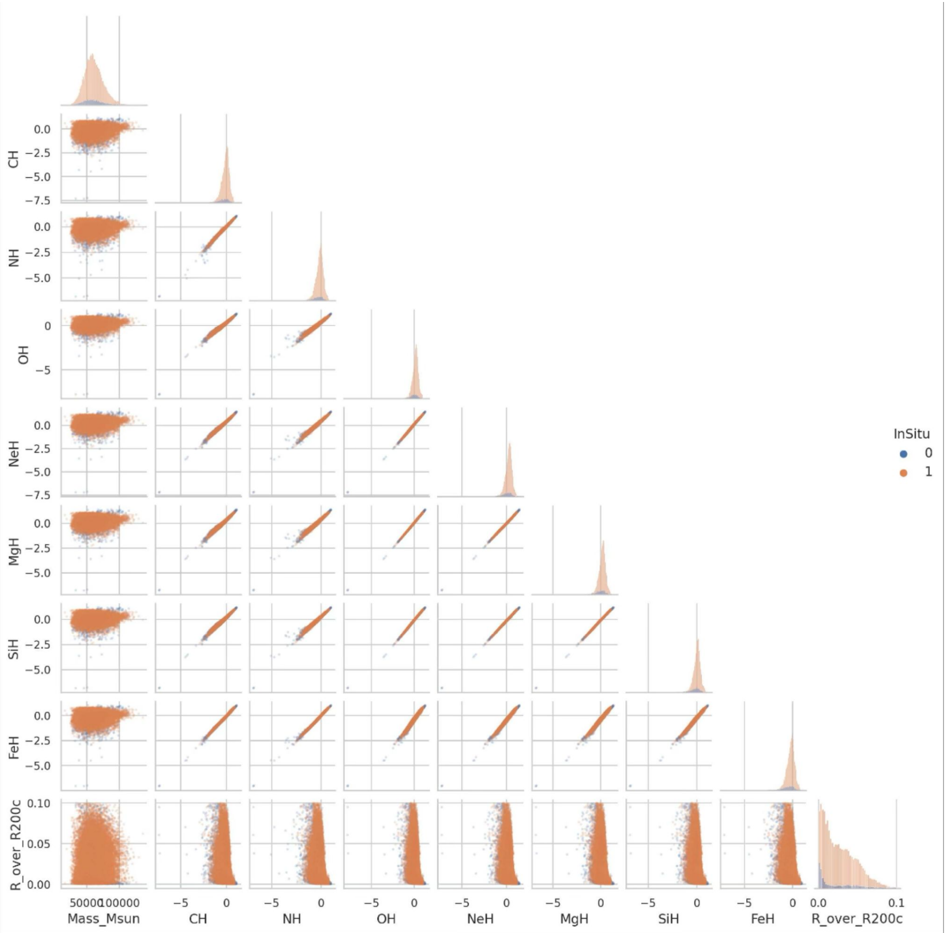
Simulation Name	L_{box} [Mpc]	N_{DM}	$m_{DM}[M_\odot]$	N_{snap}	$N_{subfind}(z=0)$
TNG50-1	51.7	2160^3	8.5×10^4	100	5688113

198 MW like galaxy Cutout data



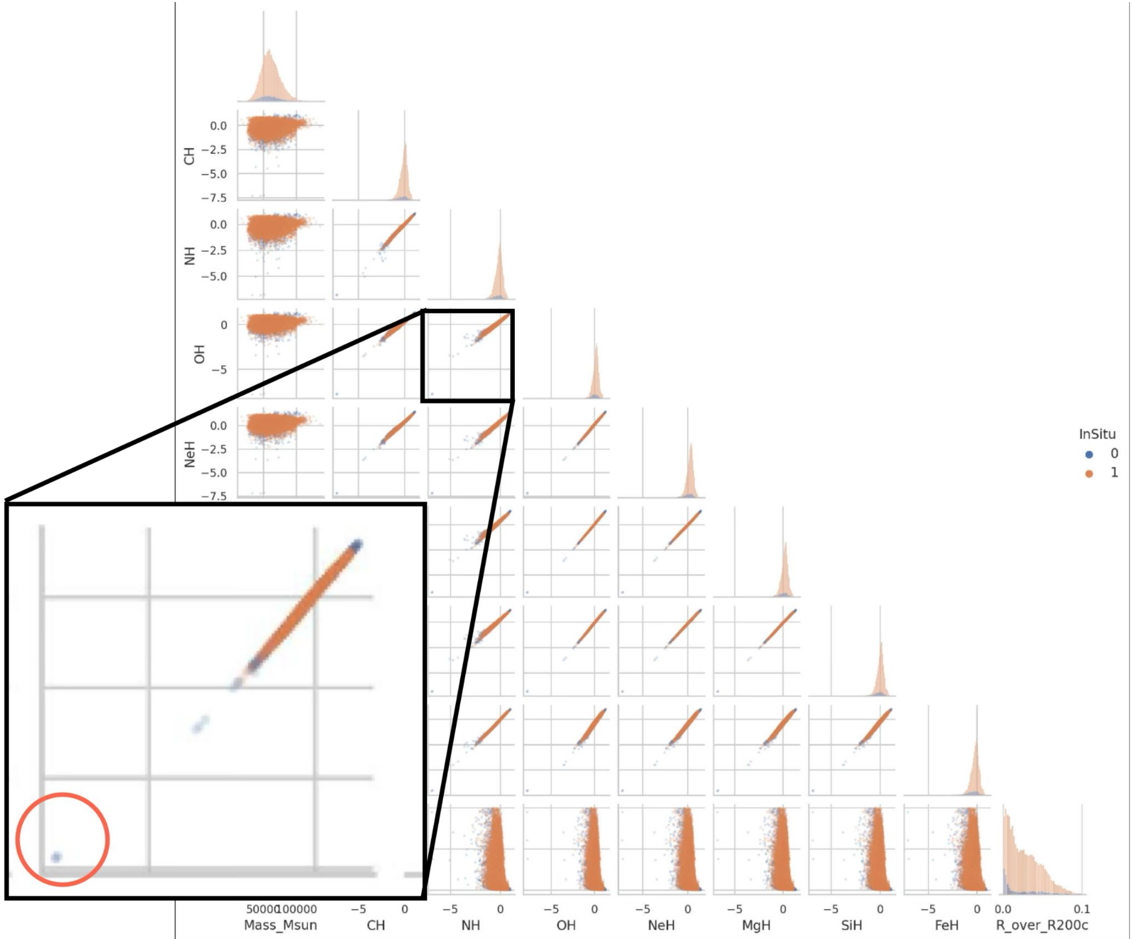
Source: TNG Project, **TNG50 Milky Way + Andromeda Sample**, https://www.tng-project.org/data/milkyway+andromeda/

Machine Learning(using IllustrisTNG)



Machine Learning(using IllustrisTNG)

대부분의 galaxy의 insitu fraction
0.8-0.95
Star particle의 metal 확인
[X/H]가 -7.0으로 낮은 particle
존재outlier로 판단
insitu_fraction이 낮은 경우도
outlier로 판단



Future Work

1. Optimization of machine-learning scores through systematic data preprocessing.
2. Establishment of hyperparameters for various machine-learning techniques using TNG simulation data.
3. Application of the trained models to observational data to verify consistency with current theoretical expectations for in-situ and accreted stellar populations.