

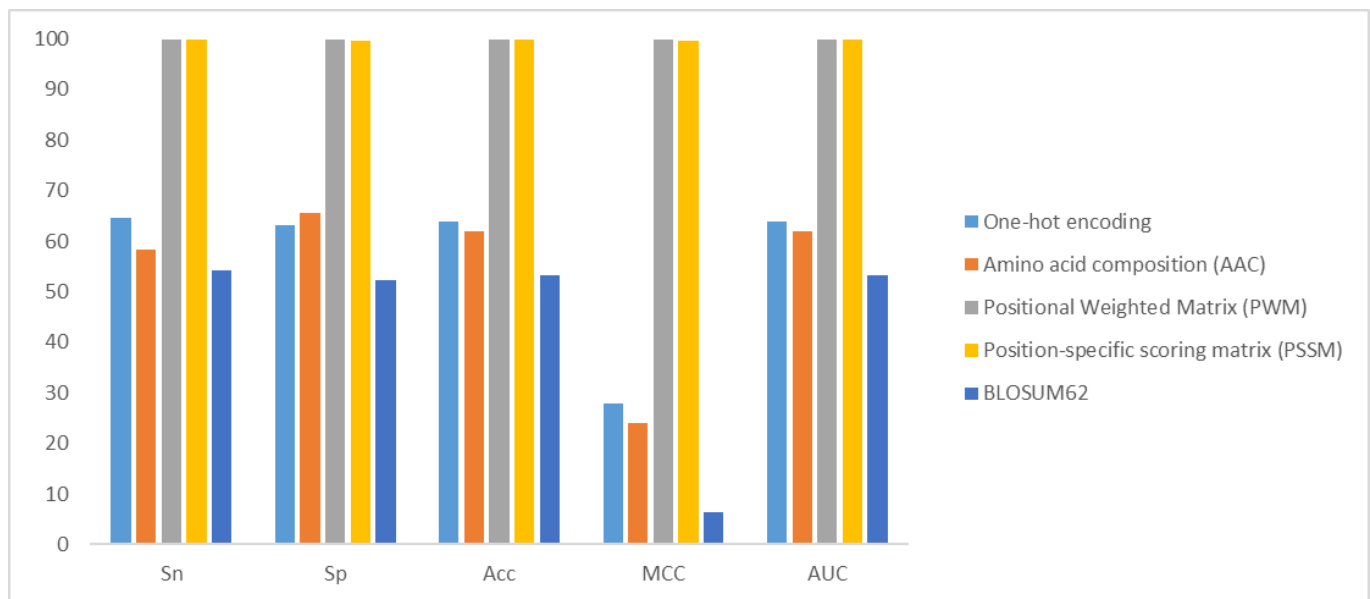
Machine Learning in Computational Biology Assignment #3

312351012 李延杰

專案代碼 : <https://github.com/Connection2Peter/NYCU-BioML-2023>

1. [Performance Comparison of Different Feature Encoding Methods]

Training features	Sn	Sp	Acc	MCC	AUC
One-hot encoding	64.734	63.213	63.973	27.96	63.973
Amino acid composition (AAC)	58.371	65.616	61.993	24.054	61.993
Positional Weighted Matrix (PWM)	100	99.812	99.906	99.813	99.906
Position-specific scoring matrix (PSSM)	100	99.7	99.85	99.7	99.85
BLOSUM62	54.26	52.177	53.219	6.442	53.218



2. [Performance Comparison of Different Supervised Learning Methods]

Training features	Sn	Sp	Acc	MCC	AUC
Decision Tree	99.512	99.287	99.399	98.8	99.399
Random Forest	100	99.737	99.869	99.738	99.869
Support Vector Machine	72.278	64.433	68.356	36.831	68.356
XGBoost	100	100	100	100	100
Multilayer Perceptron	80.254	75.338	77.797	55.775	77.796

