

Advancing Antimicrobial Polymer Development: A Novel Database and Accelerated Design via Machine Learning

Yuankai Zhao^a, Roger J Mulder^b Daniel J. Eyckens^b, Shadi Houshyar^a and Tu C. Le^{a,*}

^a School of Engineering, STEM College, RMIT University, GPO Box 2476, Melbourne, VIC 3001, Australia

^b CSIRO Manufacturing, Research Way, Clayton, VIC 3168, Australia

* Tu.Le@rmit.edu.au

Section 1: Decision Tree

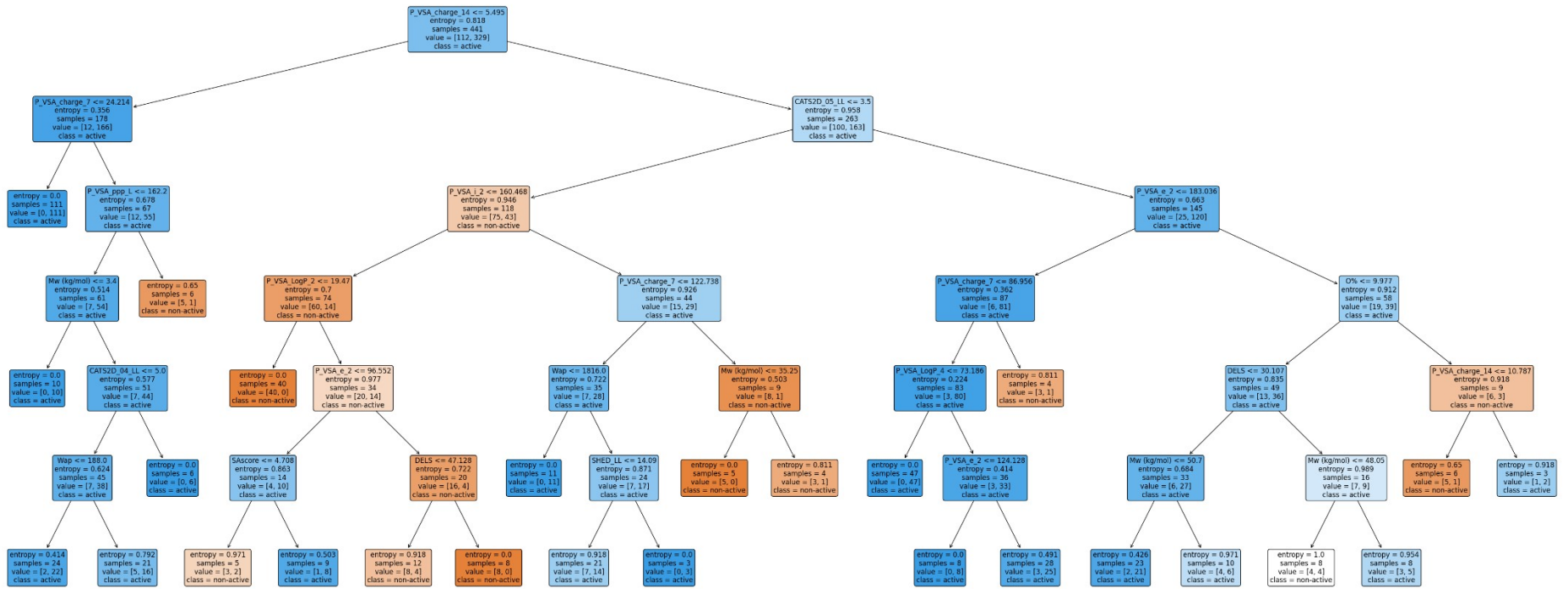
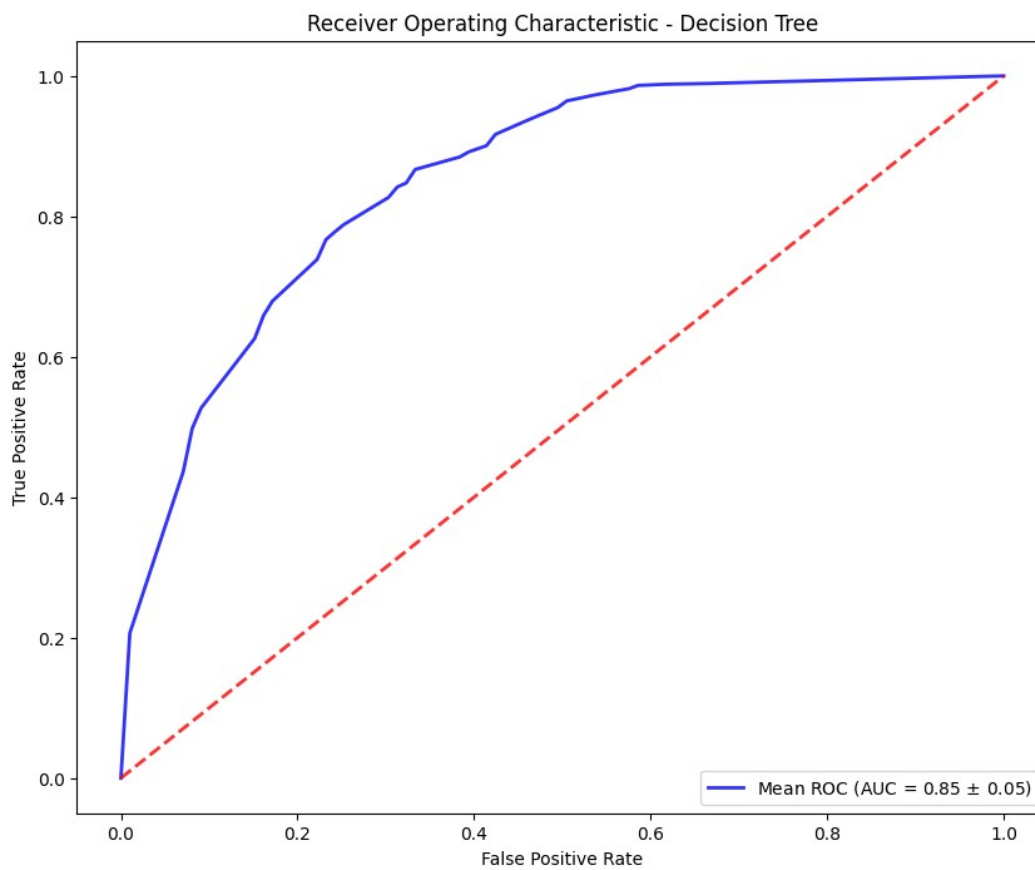
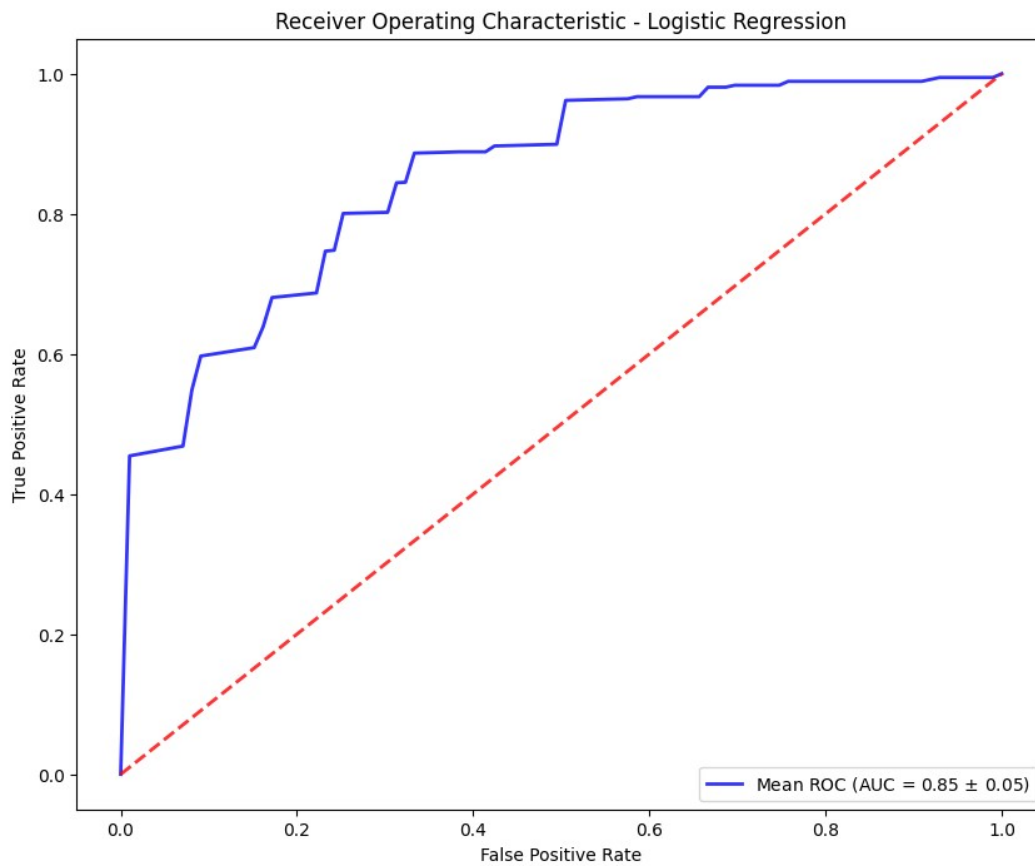
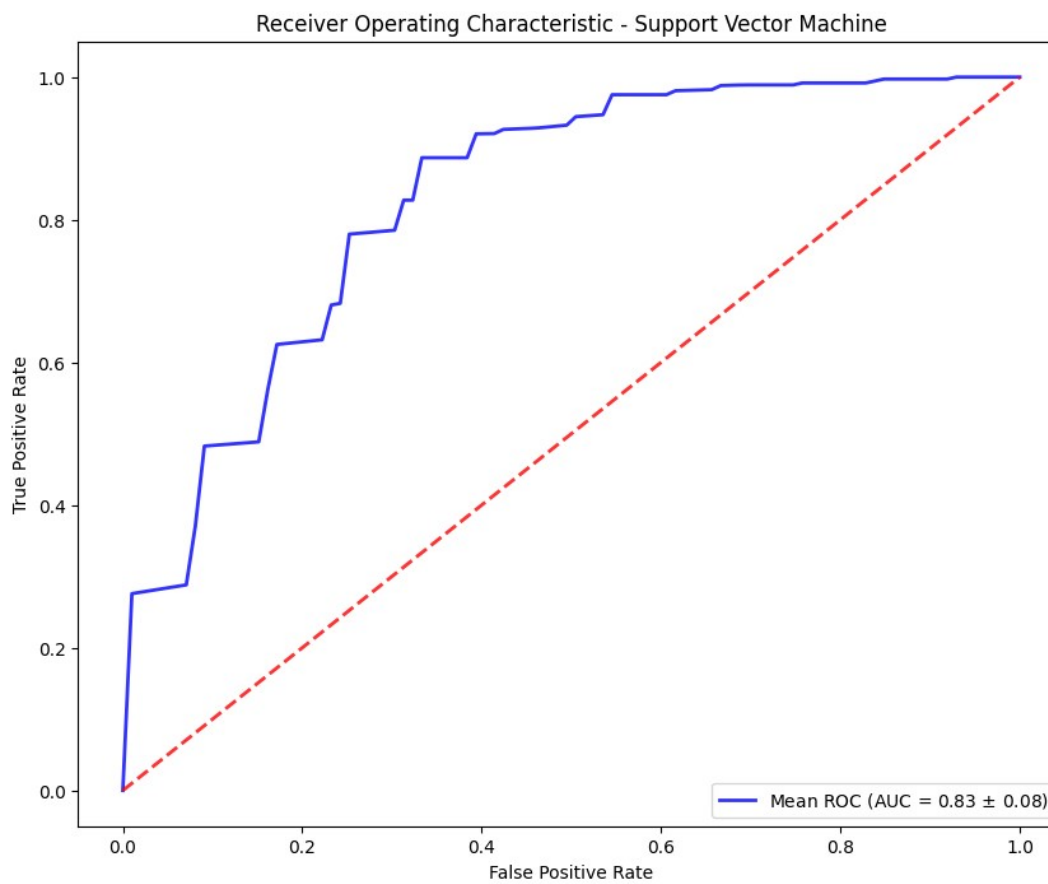
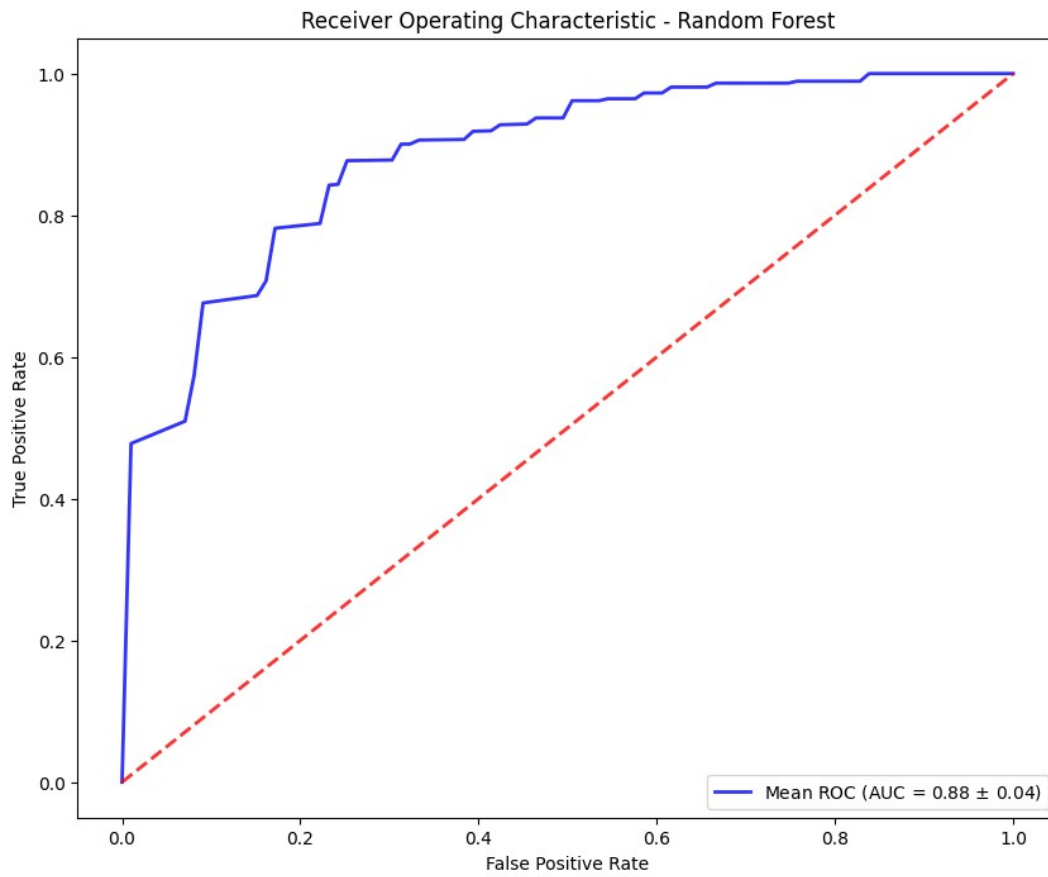


Figure S1. The Decision Tree structure shows how the model make decisions for AMPs.

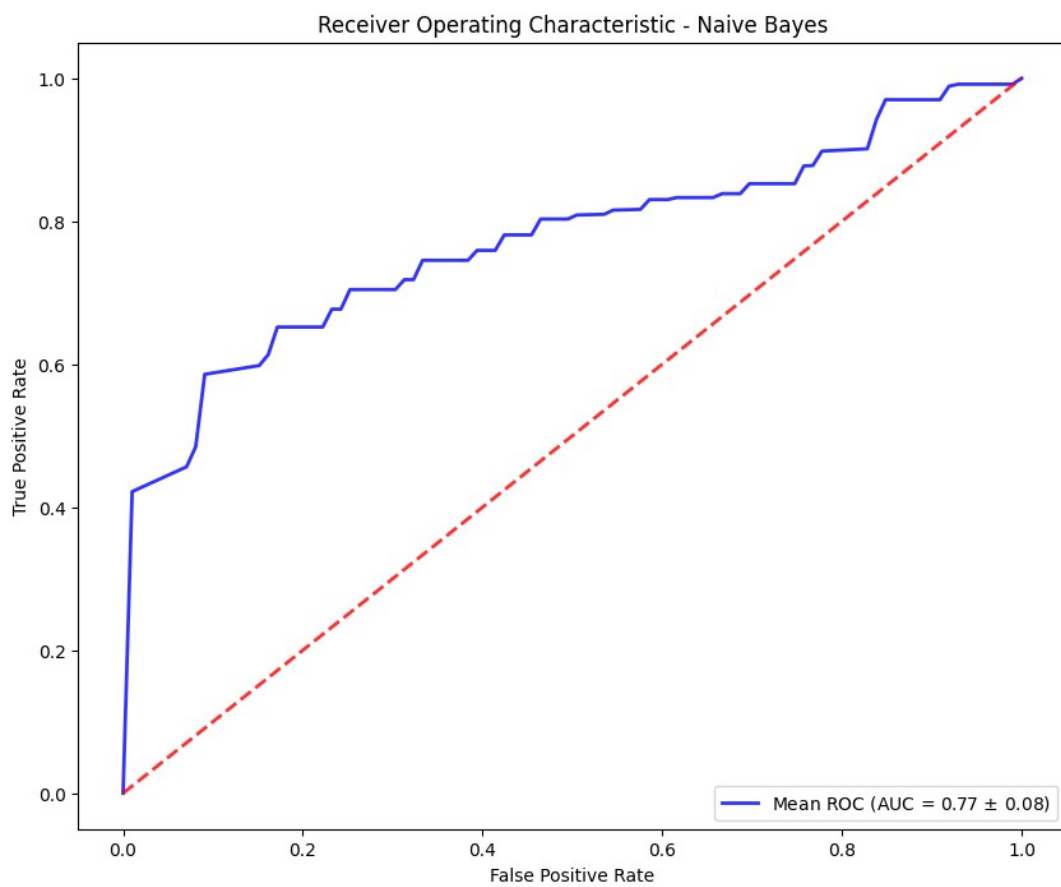
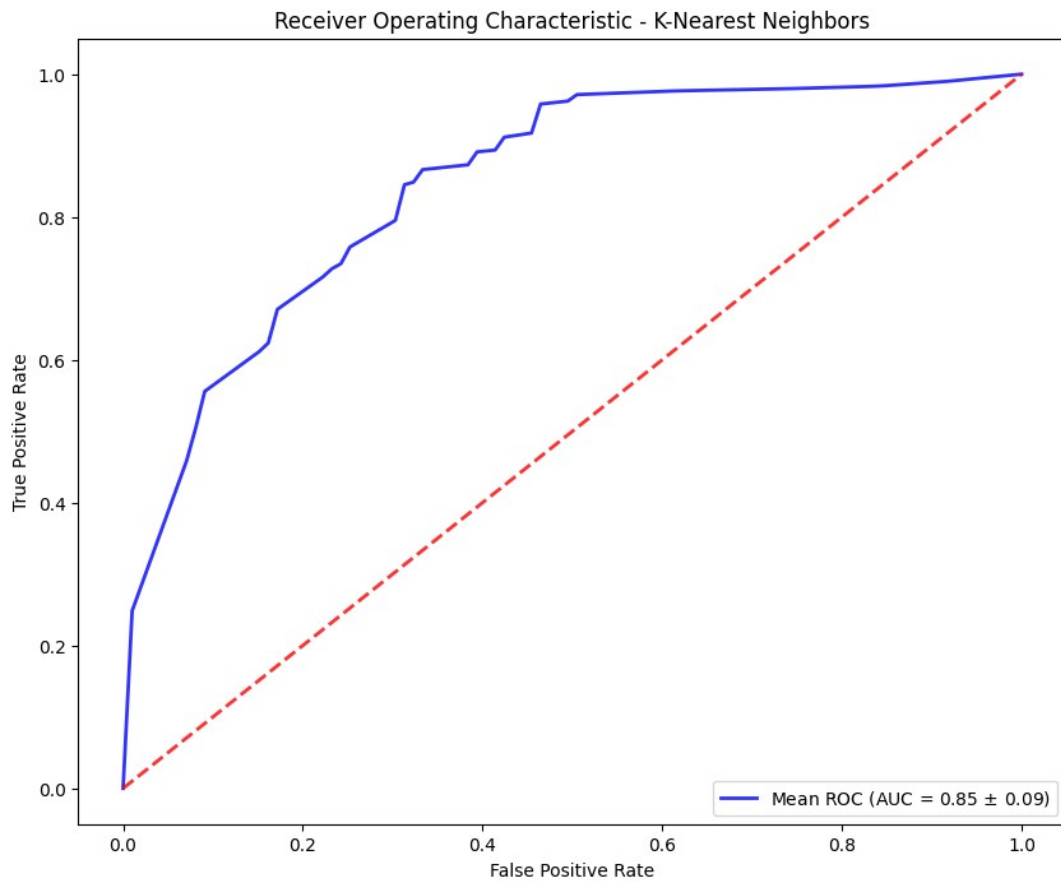
Section 2: ROC curves



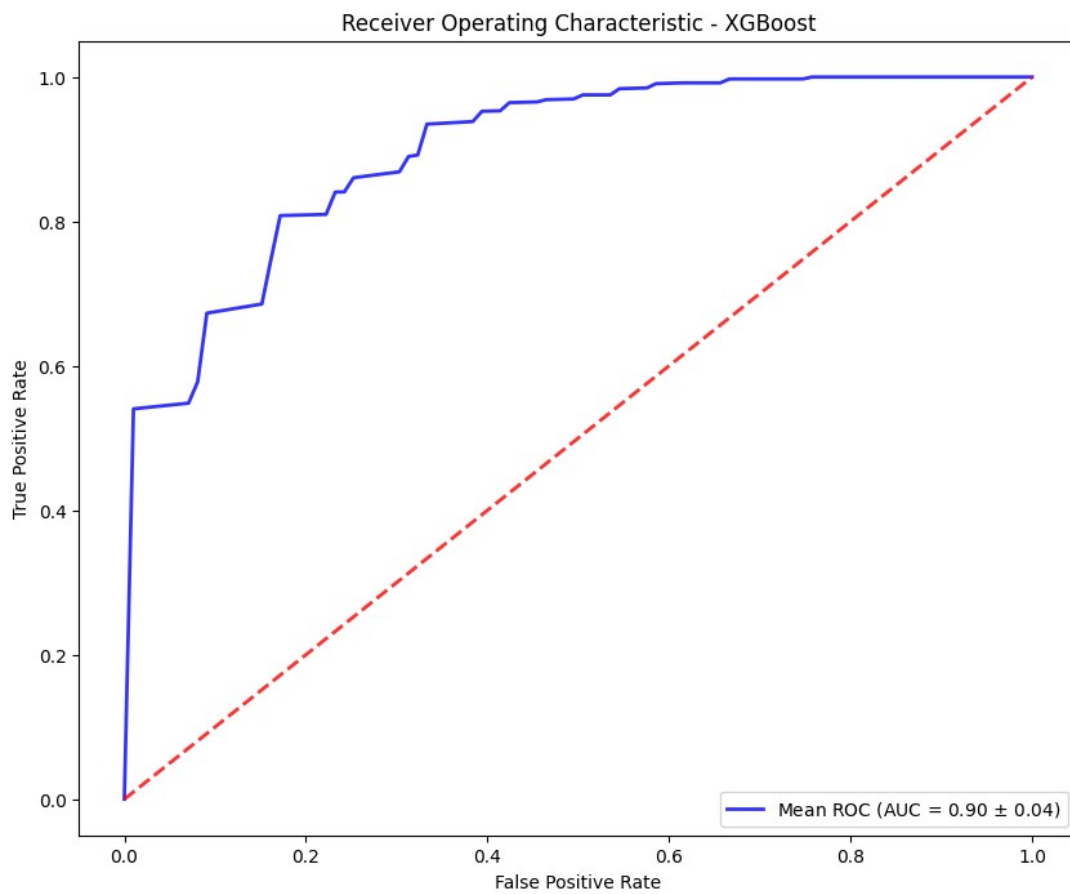
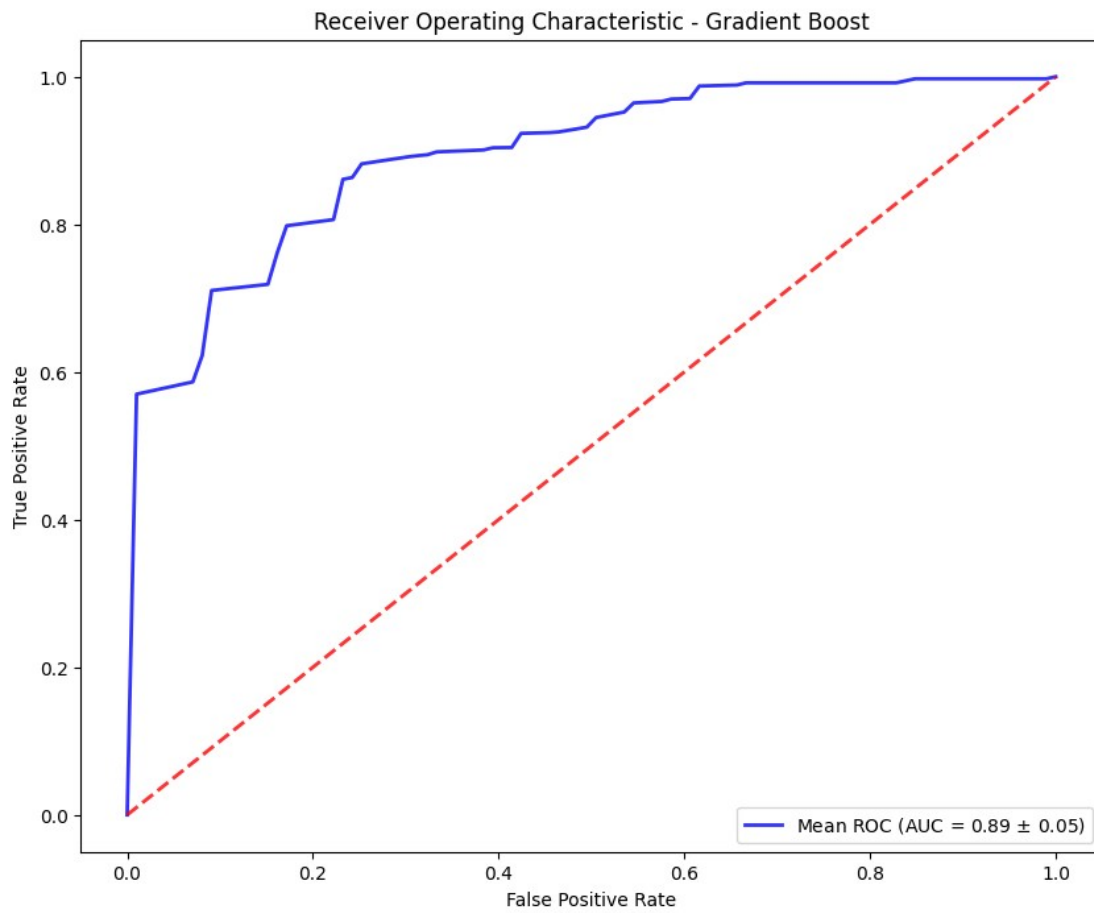
RMIT Classification: Trusted



RMIT Classification: Trusted



RMIT Classification: Trusted



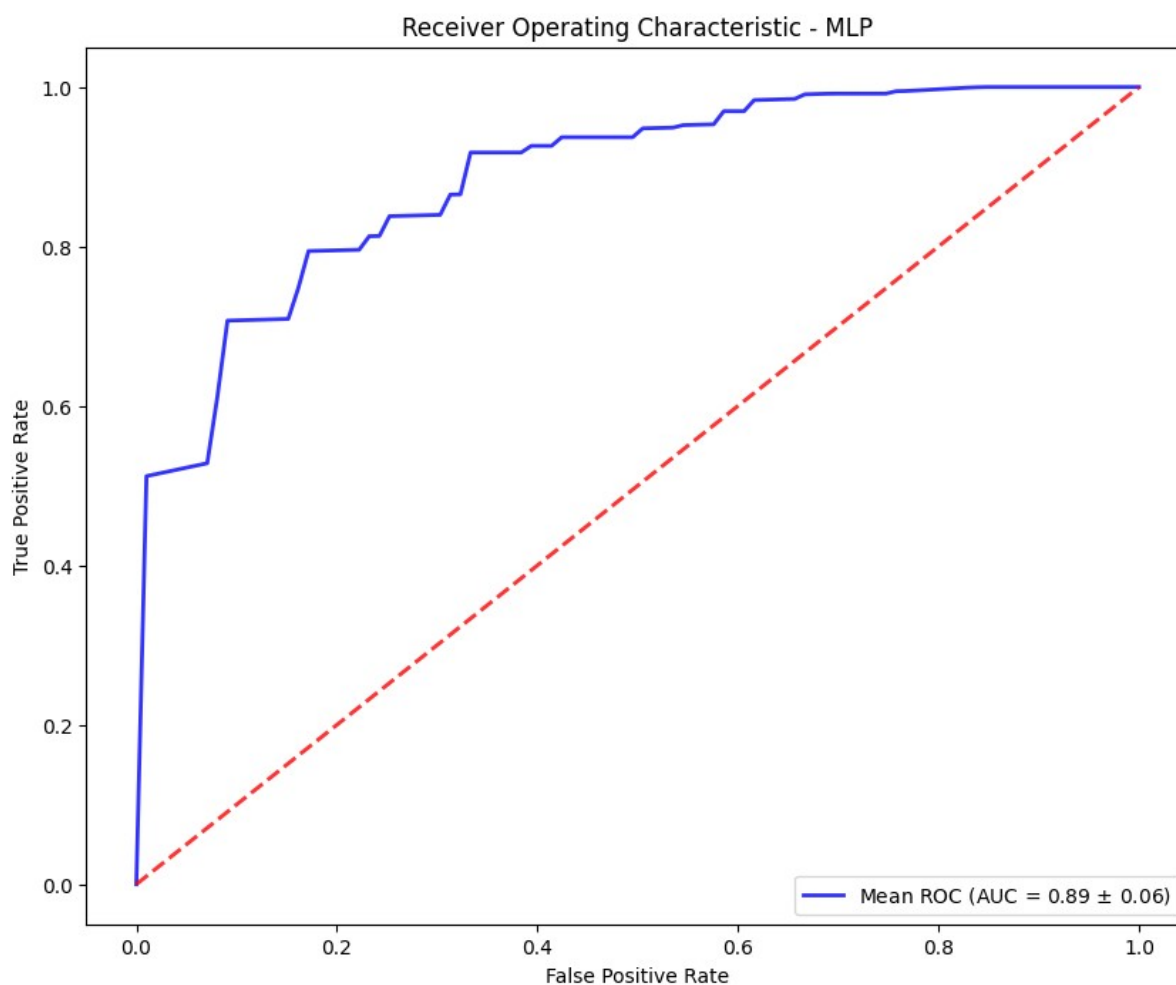


Figure S2. The ROC curves for different ML algorithms

Section 3: Descriptors importance ranking

Table S1. Descriptor lists and their importance rankings from Random Forest algorithm.

Descriptor	Importance
Mw (kg/mol)	0.083174
P_VSA_MR_6	0.064321
P_VSA_LogP_2	0.059293
P_VSA_charge_7	0.050354
CATS2D_05_LL	0.045839
P_VSA_m_2	0.042147
SHED_LL	0.04047
P_VSA_charge_9	0.038355
O%	0.037535
P_VSA_ppp_L	0.036251
P_VSA_i_2	0.035064
ALOGP	0.032593
CATS2D_04_LL	0.031622

AMR	0.03125
P_VSA_charge_2	0.030445
P_VSA_e_2	0.029582
P_VSA_MR_5	0.028341
SAscore	0.027054
J_B(i)	0.026974
Wap	0.025576
DELS	0.025043
IDET	0.022924
P_VSA_s_3	0.022489
P_VSA_LogP_4	0.022416
P_VSA_charge_14	0.022406
P_VSA_LogP_6	0.020709
P_VSA_LogP_5	0.019448
P_VSA_MR_8	0.01604
P_VSA_s_4	0.015038
P_VSA_ppp_hal	0.008767
nCsp3_endgroup1	0.004596
P_VSA_m_4	0.003881