

We selected the optimal dipeptide subset from 400 dipeptide features, which contained 144 features. These features are as follows:

['GI', 'SC', 'SS', 'CP', 'ID', 'IG', 'IK', 'FK', 'YY', 'YK', 'YR', 'PS', 'KY', 'EE', 'NI', 'IE', 'FD', 'SP', 'KI', 'ES', 'SA', 'PW', 'TY', 'EI', 'IA', 'GS', 'FY', 'CV', 'YA', 'TS', 'LK', 'RI', 'DG', 'VR', 'LM', 'RY', 'CS', 'IY', 'HC', 'SL', 'ME', 'RD', 'QI', 'RL', 'DI', 'YP', 'YD', 'SM', 'HS', 'GP', 'LC', 'AS', 'SG', 'KC', 'TF', 'PY', 'CR', 'LN', 'CD', 'GC', 'ED', 'EF', 'FR', 'ML', 'TC', 'SE', 'RG', 'IR', 'CM', 'AA', 'CL', 'YH', 'IV', 'NC', 'PI', 'DC', 'TG', 'YN', 'PG', 'PP', 'LQ', 'TI', 'QC', 'NY', 'GH', 'II', 'CT', 'ST', 'CG', 'DR', 'CE', 'VC', 'VI', 'QF', 'DY', 'WG', 'EC', 'IP', 'DS', 'YG', 'IN', 'TA', 'QV', 'AM', 'KL', 'GR', 'GF', 'PH', 'PE', 'IT', 'QE', 'WC', 'CQ', 'EP', 'EM', 'KF', 'RE', 'KG', 'WM', 'EY', 'AY', 'CN', 'AP', 'MS', 'WR', 'LW', 'GA', 'YI', 'KN', 'HW', 'YF', 'YQ', 'YC', 'AD', 'MT', 'DP', 'AE', 'VV', 'NM', 'MY', 'WS', 'YE', 'TW', 'QL']