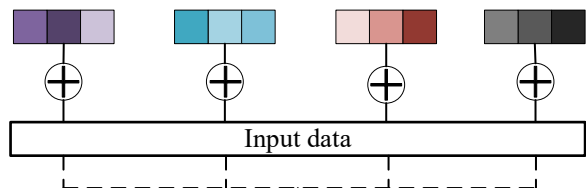


Mixture of Experts

Inputs



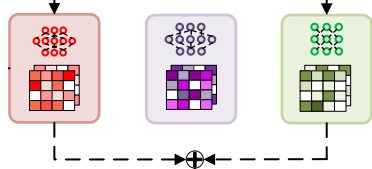
Layer normalization

Self-attention mechanism

Standardization

Gated network

Select an expert model

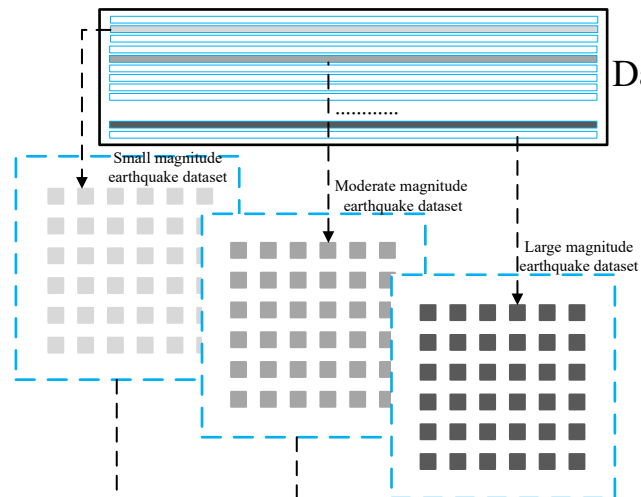


Weighting scheme

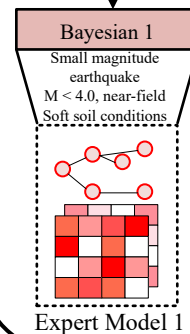


Determine the Expert Model

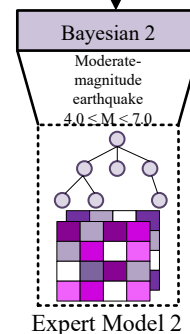
Dataset



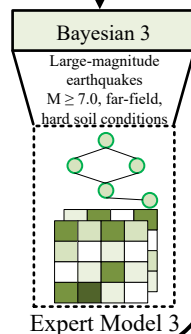
Bayesian Optimization



Expert Model 1



Expert Model 2



Expert Model 3

Output

