**Table S3** - Hyper-parameters of RPITER in comparison.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataseta | Batch size | The learning rate of the Adam optimizerb | The learning rate of the SGD optimizer | Epochs of the first training | Epochs of the second training |
| RPI2241 | 150 | 0.001 | 0.01 | [35,35,35,35,15]c | [10,10,10,10,10] |
| RPI369 | 80 | 0.001 | 0.01 | [20,20,20,20,10] | [10,10,10,10,15] |
| NPInter10412 | 150 | 0.001 | 0.005 | [35,35,35,35,15] | [15,15,15,15,15] |
| NPInter10412\_20% | 150 | 0.001 | 0.005 | [30,30,30,30,15] | [10,10,10,10,10] |
| NPInter10412\_40% | 150 | 0.001 | 0.005 | [33,33,33,33,15] | [10,10,10,10,10] |
| NPInter10412\_60% | 150 | 0.001 | 0.005 | [33,33,33,33,15] | [13,13,13,13,13] |
| NPInter10412\_80% | 150 | 0.001 | 0.005 | [33,33,33,33,15] | [15,15,15,15,15] |
| RPI7317 | 150 | 0.001 | 0.005 | [35,35,35,35,15] | [15,15,15,15,15] |

aNPInter10412\_20%, NPInter10412\_40%, NPInter10412\_60%, and NPInter10412\_80% were randomly selected from the NPInter10412 with 20%, 40%, 60%, and 80% ratio.

bRPITER contains two training stages. The Adam was used to obtain a quick converge in the first stage. The SGD (stochastic gradient descent) was used to fine tune the modules in the second stage.

cThe numbers respectively correspond to the training epochs of Conjoint-CNN Module, Conjoint-Struct-CNN Module, Conjoint-SAE Module, Conjoint-Struct-SAE Module, and Ensemble Module.