**Table 2. Overall Performance of MOA-DILI Model in Training and Test Sets**

| | | ACC | MCC | $F_1$ score |
|---|---|---|---|---|
| MOA-DILI model | training set | **0.757 (0.020)** | **0.505 (0.045)** | **0.717 (0.032)** |
| | test set | 0.695 (0.043) | 0.385 (0.087) | 0.640 (0.058) |
| MOA-DILI model with top 4 MOAs | training set | 0.703 (0.027) | 0.397 (0.056) | 0.649 (0.042) |
| | test set | **0.711 (0.040)** | **0.416 (0.080)** | **0.659 (0.052)** |
| standard QSAR model | training set | 0.658 (0.031) | 0.310 (0.062) | 0.626 (0.040) |
| | test set | 0.662 (0.041) | 0.322 (0.082) | 0.627 (0.049) |
| label permutated model | training set | 0.591 (0.039) | 0.200 (0.080) | 0.609 (0.046) |
| | test set | 0.582 (0.042) | 0.182 (0.084) | 0.604 (0.046) |

The 1000 repetitions resulted in 1000 different MOA-DILI models, which allow assessment of the frequency of each MOA selected by MOA-DILI models. As shown in Figure 2, most MOA-DILI models only used two or four MOAs; the average number of MOAs was around 3.56. Four MOAs (MOA-4, -10, -11, and -16), labeled as ARE-bla (antioxidant response element), ER-luc-bg1-4e2-antagonist (ER-alpha, BG1 cell line), gh3-tre-antagonist (thyroid receptor), and PPARG-bla-agonist (peroxisome proliferator-activated receptor gamma), were used by more than 30% of MOA-DILI models, much more often than other MOAs (Supporting Information Table S2). Consequently, we constructed the MOA-DILI model using only these four MOAs with the same process outlined in Figure 1. As summarized in Table 2, this final model yielded prediction accuracies of 0.697 in 5-fold CV and 0.711 in hold-out testing, which was comparable with other MOA-DILI models and much better than the QSAR model (*P* value <0.0001). The detailed prediction results of the top 4 assay specific models are provided in Supporting Information Table S3.