APPENDIX

In appendixes, we provide the detailed results for RQ1-RQ4 as the supplementary materials. First, for each problem and each time budget, we compare a pair of algorithms. Second, to compare the overall performance of the algorithms, we combine all objectives together by calculating average values of the objective functions (called *OFV*):

$$OFV = \frac{\sum_{i=1}^{n} Fitness_i}{n}$$

where n is the number of objectives for the prioritization problem, and $Fitness_i$ is the fitness value of the ith objective for the problem. Third, we used hypervolume (*HV*)–the most commonly used quality indicator to compare the overall performance of multi-objective search algorithms. Last, we calculated *Rank* and *Confidence* (as described in Section 4.1.5) for group comparison. Table 1 shows the navigation of RQ1-RQ4 results in the supplementary materials.

TABLE 1
Navigation of detailed results in the supplementary materials

Name	**Index and Table Numbers**
Use Case	AW1, AW2, AW3, AW4 and GS1 in Appendix A-E
RQ1	Comparison with RS with the Vargha and Delaney statistics ($\hat{A}_{12}$) and Mann-Whitney U Test (p-value) in Table i ($i = 1 \dots 10$) for problem i
RQ2	- Group comparison with the Kruskal-Wallis test in Table $11 + 3(i-1)(i = 1 \dots 10)$ for problem i - Pair comparison with the Vargha and Delaney statistics ($\hat{A}_{12}$) and Mann-Whitney U Test (p-value) in Table $12 + 3(i-1)(i = 1 \dots 10)$ for problem i - Rank and Confidence in Table $13 + 3(i-1)(i = 1 \dots 10)$ for problem i
RQ4	- Group comparison with the Kruskal-Wallis test in Table 41 for all the problems - Pair comparison using the Vargha and Delaney statistics ($\hat{A}_{12}$) and Mann-Whitney U Test (p-value) in Table 42 for all the problems