**SM3. Statistical Tests to Compare the Performances of the Algorithms**

**Parametric test for MinF1**

Since the MinF1 data set is normally distributed (p>0.05), one-way ANOVA is applied. While Table A1 shows that the homogeneity of variances is not significantly different from each other, Table A2 shows that the results obtained from the four algorithms have no significant difference in terms of MinF1 since the p-value is greater than 0.05 (sig. 0.976).

Table A1. Test of homogeneity of variances for Minf1 values

|  |  |  |  |
| --- | --- | --- | --- |
| Levene Statistic | df1 | df2 | Sig. |
| .129 | 3 | 220 | .943 |

Since Sig is .943 (>0.05), the variances of the groups are the same.

Table A2. One-way ANOVA for MinF1 values

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 2449.544 | 3 | 816.515 | .069 | .976 |
| Within Groups | 2603955.260 | 220 | 11836.160 |  |  |
| Total | 2606404.804 | 223 |  |  |  |

**Non-parametric Tests for MinF2 and MeanHV\_N**

Two non-parametric tests were used to statistically analyze the experimental results and validate the effectiveness of the proposed NSGAII\_GO. The first test is the Friedman test, which is designed to measure differences between the performances of two or more algorithms. Table A3 presents the Friedman test results for the compared algorithms based on MinF2. It is observed that the proposed NSGAII\_GO achieves the minimum rank, demonstrating its superior performance among all algorithms. Additionally, the p-value equals .002 indicating that the differences among four algorithms are statistically significant.

Table A3.Friedman test summary for MinF2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Algorithms | Rank | Mean | Std deviation | Min | Max |
| NSGAII\_RF | 2.85 | 39.3600 | 95.82712 | .00 | 485.51 |
| NSGAII\_RO | 2.62 | 41.2552 | 103.72764 | .00 | 502.44 |
| NSGAII\_GF | 2.36 | 26.9888 | 71.49753 | .00 | 396.00 |
| NSGAII\_GO | 2.18 | 29.4863 | 66.29029 | .00 | 323.49 |
| Chi-square | 14.781 |  |  |  |  |
| p-value | .002 |  |  |  |  |

The Wilcoxon signed-rank test, which is the second non-parametric test, aims to compare the distribution of paired samples to determine if there is a significant difference between them. The result of the Wilcoxon test is presented in Table A4 for MeanHV\_N values. In the table, R+ indicates that NSGAII\_GO outperforms the compared algorithms on R+ instances, while R− means that the other algorithms performed better than NSGAII\_GO. Since the value of R+ is considerably larger than that of R− in each comparison, NSGAII\_GO significantly outperforms the other three algorithms in pairwise comparisons.

Table A4. Wilcoxon signed-rank test summary for MeanHV\_N

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Comparisons | R+ | R- | Ties | p-value |
| NSGAII\_RO - NSGAII\_RF | 19 | 15 | 22 | .783 |
| NSGAII\_GF - NSGAII\_RF | 37 | 12 | 7 | .000 |
| NSGAII\_GO - NSGAII\_RF | 39 | 9 | 8 | .000 |
| NSGAII\_GF - NSGAII\_RO | 36 | 17 | 3 | .005 |
| NSGAII\_GO - NSGAII\_RO | 35 | 11 | 10 | .000 |
| NSGAII\_GO - NSGAII\_GF | 29 | 18 | 9 | .053 |