Supplementary material: Time complexity evaluation and further performance analysis of NSGAII-MIIP

An-Da Li[a,b], Zhen He[c], Qing Wang[a,b], Yang Zhang[a,b,*], Yanhui Ma[d,*]

[a] *School of Management, Tianjin University of Commerce, Tianjin 300134, China*
[b] *Research Center for Management Innovation and Evaluation, Tianjin University of Commerce, Tianjin 300134, China*
[c] *College of Management and Economics, Tianjin University, Tianjin 300072, China*
[d] *School of Management, Tianjin University of Technology, Tianjin 300384, China*

1. Introduction

In this article, the supplementary information of the manuscript is provided. Section 2 analyzes the time complexity of the proposed NSGAII-MIIP algorithm. Section 3 evaluates the effects of ω in the mutual information (MI) based feature importance measure. Section 4 gives supplementary information on the overall KPF selection performance of non-dominated sets found by the multi-objective feature selection (FS) methods on the manufacturing process datasets. Section 5 further evaluates the FS performance of NSGAII-MIIP on public benchmark datasets. Section 6 gives the experimental results of the ablation study, where NSGAII-MIIP is compared to its two variants to verify the effectiveness of the improvement strategy in NSGAII-MIIP.

2. Evaluating the time complexity of NSGAII-MIIP

The time-consuming parts of NSGAII-MIIP shown in Algorithm 3 of the manuscript include a) the entropy and MI calculation, b) the proposed k-medoids algorithm, and c) the genetic operators and improvement-phase-embedded ranking approach during the iteration process. Let M be the number of instances in the training set $\mathcal{D}^{tr}$, N be the number of features in $\mathbb{F}$, T be the number of iterations of NSGAII-MIIP, S be the population size, and V be the number of objectives (which is 2 in the proposed key process feature (KPF) selection model). Without considering the time of function evaluations, the time complexity of each of the three parts and the overall time complexity of NSGAII-MIIP are evaluated as follows.

First, the time complexities of calculating the entropy for each feature in $\mathbb{F}$ and calculating the MI between any two features in $\mathbb{F}$ in the first two lines of Algorithm 3 are $O(MN)$ and $O(MN^2)$ respectively.

Second, the time complexity of the proposed k-medoids algorithm is decided by the k-means++ algorithm, the distance calculation step, and the iteration process. The time to calculate the distances between any two features in $\mathbb{F}$ is $O(N^2)$. The k-means++ algorithm needs to calculate the probability of each feature to be selected as the next medoid, which requires a time of $O(kN)$ (k is the number of clusters, which is lower than N). The time complexity at each iteration of the k-medoids algorithm is decided by finding the minimal value of the sum of distances in each cluster $\mathbb{C}_i$, which requires a time of $O(N)$. Assuming that the number of iterations is I, the time complexity of the iteration process is $O(IN)$. Since $I < N$, the time complexity of the iteration process is $O(N^2)$ in the worst case. According to the above analysis, the time complexity of the k-medoids algorithm is $O(N^2) + O(kN) + O(N^2) \cong O(N^2)$.

Third, genetic operators and the improvement-phase-embedded ranking approach are two time-consuming parts in the iteration process of NSGAII-MIIP. The time complexities of selection, crossover, and mutation operations at an iteration are $O(S)$, $O(SN)$, and $O(SN)$. Therefore, the overall time complexity of the genetic operators is $O(SN)$. According to Algorithm 4 of the manuscript, the proposed ranking approach

*Corresponding authors
Email addresses: adli@tjcu.edu.cn (An-Da Li), zhhe@tju.edu.cn (Zhen He), wangqing@tjcu.edu.cn (Qing Wang), yzhang@tcju.edu.cn (Yang Zhang), mayanhui@email.tjut.edu.cn (Yanhui Ma)

is composed of dividing solutions in $\mathbb{R}^{new}$ into different fronts, calculating crowding distances for solutions in $\mathbb{R}^{new}$, and the MI-guided improvement strategy. Although the improvement phase is embedded in the ranking process, no additional operations are conducted when obtaining the non-dominated fronts compared with the fast non-dominated sorting approach. Therefore, the time complexity to obtain the non-dominated fronts for $\mathbb{R}^{new} = \mathbb{R} \cup \Gamma$ is $O(V(2S+L)^2)$ (Deb et al., 2002), where V, S, and L are the number of objectives, the population size, and the number of solutions in Γ. Moreover, the time complexity of crowding distance calculation is $O(V(2S+L)\log(2S+L))$ (Deb et al., 2002).

To estimate the time complexity of the MI-guided improvement strategy embedded in the ranking approach, we need to estimate the complexity of the three improvement operations. According to the add operation in Algorithm 2, in the worst case, if the operation needs to traverse all the k clusters, the total time required for finding sets $\mathbb{C}_s$ and $\mathbb{C}_e$ is $O(N)$. The time to find the feature with the maximal $W(f, \mathbb{C}_s)$ is $O(N_e)$, where $N_e \leq N$ is the number of features in $\mathbb{C}_e$. Therefore, the add operation requires a time of $O(N) + O(N_e) \cong O(N)$. Similarly, the time complexity for the eliminate operation is $O(N)$. According to the interchange operation in Algorithm 2, if the operation traverses all the k clusters, the total time to find set $\mathbb{C}_s$ and $\mathbb{C}_e$ is $O(N)$, and the time to obtain the two features $f_e = \arg\min_{f \in \mathbb{C}_s} W(f, \mathbb{C}_s \backslash \{f\})$ and $f_a = \arg\max_{f \in \mathbb{C}_e} W(f, \mathbb{C}_s \backslash \{f_e\})$ for interchanging in all clusters is $O(N)$. Therefore, the overall time complexity of applying the improvement operations to generate a set Γ of solutions is $O(L(N+N+N)) \cong O(LN)$, where L is the number of solutions in Γ.

According to the above analysis, the total time complexity of the proposed ranking approach is $O(V(2S+L)^2) + O(V(2S+L)\log(2S+L)) + O(LN) \cong O(V(2S+L)^2) + O(LN)$. The size of $\mathcal{F}_1$ at each iteration does not exceed the size of the ranking pool $2S$ (the size of $\mathcal{F}_1$ is generally substantially smaller than $2S$). Therefore, the number of solutions from the improvement phase does not exceed $6S$, i.e., $L < 6S$. So, the time complexity of the ranking approach can be further estimated as $O(V(2S+L)^2) + O(LN) \cong O(V(2S+6S)^2) + O(6SN) \cong O(VS^2) + O(SN)$, which is decided by both the population size S and the number of features N. Since the number of iterations is T, the total time required by the genetic operations and the ranking approach is $O(TSN) + (O(TVS^2) + O(TSN)) \cong O(TSN) + O(TVS^2)$. The proposed NSGAII-MIIP has a similar time complexity in the iteration process compared to NSGA-II if the same genetic operators are used.

Finally, summing the complexities of entropy and MI calculation, the k-medoids algorithm, and the iteration process, the overall time complexity of NSGAII-MIIP is $[O(MN)+O(MN^2)]+O(N^2)+[O(TSN)+O(TVS^2)] \cong O(MN^2) + O(TSN) + O(TVS^2)$.

3. Evaluating the effects of ω in the MI-based feature importance measure

The parameter ω of the feature importance measure in Eq. (6) of the manuscript reflects the relative importance of relevance compared to redundancy. In this section, we evaluate the effects of different values of ω on the search performance of NSGAII-MIIP and select a proper ω for NSGAII-MIIP with experiments. In the experiments, three candidate ω values (i.e., 1,2, and 3) are used in NSGAII-MIIP with the same experimental settings described in Section 5.3 of the manuscript. HV, IGD, and SC described in Section 7.1 of the manuscript are used as the performance metrics to compare the search performance of NSGAII-MIIP of different ω values. The performance metric results on the three ω values are shown in Table 1. The best HV, IGD, or SC value obtained by a ω is highlighted in bold. The Wilcoxon signed-rank test (Wilcoxon, 1945) is used to test if the HV, IGD, or SC value obtained by $\omega = 2$ is significantly different from $\omega = 1$ or $\omega = 2$, where $\upuparrows$ ($\uparrow$) or $\downdownarrows$ ($\downarrow$) denotes that $\omega = 2$ obtains a significantly better or worse result at a significance level of 0.05 (0.1).

According to the statistical significance test results in Table 1, the HV, IGD, and SC values of $\omega = 2$ are not significantly different from those of $\omega = 1$ and $\omega = 3$ in most cases. The only exceptions are that the HV, IGD, and SC values of $\omega = 2$ are significantly better than $\omega = 3$ on PAPER, and the IGD value of $\omega = 2$ is significantly worse than $\omega = 3$ on PAPER-F. These results denote that the search performance of NSGAII-MIIP is not sensitive to ω. Overall, $\omega = 2$ obtains the best results on the three metrics in 10 cases, $\omega = 1$ obtains the best results in 9 cases, while $\omega = 3$ obtains the best results in 5 cases. This denotes that $\omega = 2$

Table 1: The HV, IGD, and SC results of NSGAII-MIIP with different ω values.

Metric	Dataset	$\omega = 2$	$\omega = 1$	$\omega = 3$
HV	ADPN	1.003 ± 0.064	$\mathbf{1.006 \pm 0.062}$	1.006 ± 0.055
	LATEX	1.011 ± 0.045	1.013 ± 0.059	$\mathbf{1.018 \pm 0.054}$
	PAPER	$\mathbf{1.025 \pm 0.042}$	1.014 ± 0.046	$0.997 \pm 0.051 \upuparrows$
	SPIRA	$\mathbf{0.987 \pm 0.053}$	0.986 ± 0.040	0.979 ± 0.052
	ADPN-F	1.076 ± 0.064	$\mathbf{1.079 \pm 0.056}$	1.079 ± 0.061
	LATEX-F	$\mathbf{1.068 \pm 0.071}$	1.060 ± 0.086	1.063 ± 0.087
	PAPER-F	1.007 ± 0.057	$\mathbf{1.012 \pm 0.044}$	1.011 ± 0.061
	SPIRA-F	$\mathbf{1.013 \pm 0.046}$	1.012 ± 0.038	1.011 ± 0.046
IGD	ADPN	0.053 ± 0.027	$\mathbf{0.047 \pm 0.034}$	0.052 ± 0.027
	LATEX	0.055 ± 0.031	0.052 ± 0.035	$\mathbf{0.052 \pm 0.030}$
	PAPER	$\mathbf{0.037 \pm 0.025}$	0.047 ± 0.033	$0.061 \pm 0.037 \upuparrows$
	SPIRA	0.048 ± 0.028	$\mathbf{0.047 \pm 0.024}$	0.052 ± 0.023
	ADPN-F	0.056 ± 0.034	$\mathbf{0.055 \pm 0.030}$	0.060 ± 0.035
	LATEX-F	$\mathbf{0.067 \pm 0.035}$	0.070 ± 0.045	0.070 ± 0.040
	PAPER-F	0.063 ± 0.043	0.053 ± 0.030	$\mathbf{0.053 \pm 0.038} \downarrow$
	SPIRA-F	0.057 ± 0.028	$\mathbf{0.049 \pm 0.021}$	0.059 ± 0.023
SC	ADPN	0.395 ± 0.202	$\mathbf{0.415 \pm 0.279}$	0.384 ± 0.199
	LATEX	0.347 ± 0.182	0.367 ± 0.216	$\mathbf{0.373 \pm 0.169}$
	PAPER	$\mathbf{0.466 \pm 0.237}$	0.391 ± 0.167	$0.361 \pm 0.194 \upuparrows$
	SPIRA	$\mathbf{0.449 \pm 0.161}$	0.401 ± 0.160	0.390 ± 0.126
	ADPN-F	0.282 ± 0.197	$\mathbf{0.302 \pm 0.178}$	0.290 ± 0.206
	LATEX-F	$\mathbf{0.156 \pm 0.243}$	0.136 ± 0.211	0.075 ± 0.129
	PAPER-F	0.259 ± 0.148	0.268 ± 0.161	$\mathbf{0.319 \pm 0.182}$
	SPIRA-F	$\mathbf{0.310 \pm 0.177}$	0.284 ± 0.162	0.290 ± 0.191

obtains better search performance in more cases. Therefore, we use $\omega = 2$ to establish the NSGAII-MIIP algorithm in the experiments of the manuscript and the remaining parts of this supplementary material.

4. Further comparison of overall KPF selection performance of non-dominated sets found by the multi-objective FS methods

In the manuscript, we have compared the overall KPF selection performance of the non-dominated sets between NSGAII-MIIP and benchmark algorithms on the AUC (area under the curve) measure. In this section, we further list the comparison results on GM and F1-score measures.

Tables 2 and 3 show the comparison results of the overall KPF selection performance using GM and F1-score measures. Overall, the results are consistent with that of the AUC measure shown in the manuscript. First, according to the mean HV, IGD, and SC values on the two measures, NSGAII-MIIP obtains the best results in most cases. In terms of GM, NSGAII-MIIP obtains the best results on HV, IGD, and SC in 20 of 24 cases. In terms of the F1-score, NSGAII-MIIP obtains the best results in 16 of 24 cases. The significance test results denote that NSGAII-MIIP obtains significantly better results on HV, IGD, and SC than benchmark algorithms except for MOPSO-LS in most cases in terms of the GM and F1-score measures. No evidence indicates that NSGAII-MIIP obtains significantly worse results in terms of GM or the F1-score. Although NSGAII-MIIP obtains significantly better HV, IGD, and SC values than MOPSO-LS in relatively fewer cases, the mean HV, IGD, and SC values of NSGAII-MIIP are better than those of MOPSO-LS in most cases. This indicates that NSGAII-MIIP outperforms MOPSO-LS. Overall, the results in Tables 2 and 3 demonstrate that the non-dominated solutions obtained by NSGAII-MIIP have better overall KPF selection performance than benchmark algorithms.

Table 2: Comparison of overall KPF selection performance of obtained non-dominated sets between NSGAII-MIIP and benchmark algorithms using the GM measure.

Metric	Dataset	NSGAII-MIIP	GADMS-IPM	MOPSO-LS	MOFS-BDE	NSGAII-IPM	IDMS-IPM	NSGA-II/SDR	MOEA/D	SPEA2
HV	ADPN	0.977 ± 0.291	$0.892 \pm 0.415 \upuparrows$	0.864 ± 0.589	$\mathbf{0.998 \pm 0.231}$	$0.833 \pm 0.393 \upuparrows$	$0.178 \pm 0.159 \upuparrows$	0.864 ± 0.420	0.900 ± 0.545	$0.803 \pm 0.376 \upuparrows$
	LATEX	$\mathbf{0.996 \pm 0.131}$	$0.888 \pm 0.149 \upuparrows$	0.990 ± 0.137	$0.922 \pm 0.154 \uparrow$	$0.717 \pm 0.142 \upuparrows$	$0.254 \pm 0.087 \upuparrows$	$0.896 \pm 0.119 \upuparrows$	$0.903 \pm 0.143 \upuparrows$	$0.808 \pm 0.116 \upuparrows$
	PAPER	$\mathbf{1.067 \pm 0.097}$	1.051 ± 0.138	1.066 ± 0.085	$0.876 \pm 0.163 \upuparrows$	$0.910 \pm 0.180 \upuparrows$	$0.799 \pm 0.188 \upuparrows$	$0.974 \pm 0.190 \upuparrows$	$1.011 \pm 0.135 \upuparrows$	$0.996 \pm 0.156 \upuparrows$
	SPIRA	$\mathbf{0.982 \pm 0.154}$	$0.939 \pm 0.162 \uparrow$	0.970 ± 0.143	$0.932 \pm 0.177 \uparrow$	$0.830 \pm 0.155 \upuparrows$	$0.218 \pm 0.162 \upuparrows$	$0.894 \pm 0.189 \upuparrows$	$0.883 \pm 0.130 \upuparrows$	$0.875 \pm 0.135 \upuparrows$
	ADPN-F	1.004 ± 0.298	$0.806 \pm 0.382 \upuparrows$	$\mathbf{1.024 \pm 0.257}$	$0.908 \pm 0.284 \upuparrows$	$0.552 \pm 0.572 \upuparrows$	$0.147 \pm 0.167 \upuparrows$	$0.532 \pm 0.826 \upuparrows$	$0.859 \pm 0.384 \upuparrows$	$0.608 \pm 0.507 \upuparrows$
	LATEX-F	$\mathbf{1.032 \pm 0.106}$	$0.815 \pm 0.113 \upuparrows$	$0.985 \pm 0.102 \upuparrows$	$0.856 \pm 0.139 \upuparrows$	$0.655 \pm 0.110 \upuparrows$	$0.239 \pm 0.086 \upuparrows$	$0.810 \pm 0.130 \upuparrows$	$0.936 \pm 0.118 \upuparrows$	$0.727 \pm 0.132 \upuparrows$
	PAPER-F	$\mathbf{1.104 \pm 0.081}$	1.079 ± 0.094	1.073 ± 0.125	$1.016 \pm 0.144 \upuparrows$	$0.854 \pm 0.178 \upuparrows$	$0.376 \pm 0.208 \upuparrows$	$1.052 \pm 0.128 \uparrow$	$1.028 \pm 0.114 \upuparrows$	$0.969 \pm 0.140 \upuparrows$
	SPIRA-F	$\mathbf{1.028 \pm 0.112}$	$0.838 \pm 0.145 \upuparrows$	0.974 ± 0.198	$0.915 \pm 0.151 \upuparrows$	$0.686 \pm 0.145 \upuparrows$	$0.211 \pm 0.108 \upuparrows$	$0.871 \pm 0.150 \upuparrows$	$0.900 \pm 0.203 \upuparrows$	$0.726 \pm 0.156 \upuparrows$
IGD	ADPN	$\mathbf{0.157 \pm 0.191}$	$0.269 \pm 0.348 \upuparrows$	$0.263 \pm 0.465 \uparrow$	0.166 ± 0.131	$0.330 \pm 0.331 \upuparrows$	$0.961 \pm 0.310 \upuparrows$	0.253 ± 0.346	0.242 ± 0.434	$0.329 \pm 0.332 \upuparrows$
	LATEX	$\mathbf{0.114 \pm 0.064}$	$0.198 \pm 0.100 \upuparrows$	0.123 ± 0.080	$0.190 \pm 0.117 \upuparrows$	$0.314 \pm 0.115 \upuparrows$	$0.777 \pm 0.100 \upuparrows$	$0.195 \pm 0.100 \upuparrows$	$0.196 \pm 0.129 \upuparrows$	$0.265 \pm 0.095 \upuparrows$
	PAPER	$\mathbf{0.133 \pm 0.084}$	$0.158 \pm 0.105 \uparrow$	0.154 ± 0.096	$0.286 \pm 0.103 \upuparrows$	$0.264 \pm 0.117 \upuparrows$	$0.298 \pm 0.155 \upuparrows$	$0.234 \pm 0.138 \upuparrows$	$0.213 \pm 0.119 \upuparrows$	$0.182 \pm 0.099 \upuparrows$
	SPIRA	$\mathbf{0.131 \pm 0.084}$	$0.164 \pm 0.083 \upuparrows$	0.138 ± 0.071	0.158 ± 0.104	$0.245 \pm 0.080 \upuparrows$	$0.827 \pm 0.225 \upuparrows$	$0.187 \pm 0.105 \upuparrows$	$0.200 \pm 0.072 \upuparrows$	$0.220 \pm 0.074 \upuparrows$
	ADPN-F	$\mathbf{0.127 \pm 0.210}$	$0.335 \pm 0.336 \upuparrows$	0.147 ± 0.202	$0.205 \pm 0.196 \upuparrows$	$0.549 \pm 0.554 \upuparrows$	$0.982 \pm 0.319 \upuparrows$	$0.551 \pm 0.754 \upuparrows$	$0.261 \pm 0.341 \upuparrows$	$0.492 \pm 0.482 \upuparrows$
	LATEX-F	$\mathbf{0.125 \pm 0.089}$	$0.258 \pm 0.086 \upuparrows$	$0.153 \pm 0.074 \upuparrows$	$0.242 \pm 0.091 \upuparrows$	$0.373 \pm 0.085 \upuparrows$	$0.810 \pm 0.101 \upuparrows$	$0.263 \pm 0.084 \upuparrows$	$0.182 \pm 0.070 \upuparrows$	$0.336 \pm 0.083 \upuparrows$
	PAPER-F	$\mathbf{0.079 \pm 0.050}$	0.097 ± 0.057	0.097 ± 0.083	$0.134 \pm 0.078 \upuparrows$	$0.235 \pm 0.099 \upuparrows$	$0.658 \pm 0.192 \upuparrows$	$0.116 \pm 0.075 \upuparrows$	$0.124 \pm 0.073 \upuparrows$	$0.163 \pm 0.091 \upuparrows$
	SPIRA-F	$\mathbf{0.114 \pm 0.066}$	$0.241 \pm 0.080 \upuparrows$	0.142 ± 0.119	$0.204 \pm 0.084 \upuparrows$	$0.352 \pm 0.075 \upuparrows$	$0.862 \pm 0.126 \upuparrows$	$0.237 \pm 0.091 \upuparrows$	$0.219 \pm 0.116 \upuparrows$	$0.347 \pm 0.099 \upuparrows$
SC	ADPN	$\mathbf{0.478 \pm 0.452}$	$0.139 \pm 0.263 \upuparrows$	0.317 ± 0.402	$0.172 \pm 0.261 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.222 \pm 0.362 \upuparrows$	$0.056 \pm 0.216 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	LATEX	$\mathbf{0.239 \pm 0.196}$	$0.039 \pm 0.129 \upuparrows$	0.191 ± 0.199	$0.017 \pm 0.051 \upuparrows$	$0.011 \pm 0.061 \upuparrows$	$0.006 \pm 0.030 \upuparrows$	$0.090 \pm 0.134 \upuparrows$	$0.033 \pm 0.092 \upuparrows$	$0.012 \pm 0.047 \upuparrows$
	PAPER	$\mathbf{0.205 \pm 0.183}$	$0.104 \pm 0.146 \upuparrows$	0.169 ± 0.163	0.118 ± 0.206	0.118 ± 0.243	$0.044 \pm 0.121 \upuparrows$	0.120 ± 0.183	$0.146 \pm 0.164 \uparrow$	$0.061 \pm 0.118 \upuparrows$
	SPIRA	$\mathbf{0.278 \pm 0.253}$	$0.138 \pm 0.159 \upuparrows$	0.268 ± 0.252	0.216 ± 0.215	$0.038 \pm 0.122 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.142 \pm 0.203 \upuparrows$	$0.061 \pm 0.131 \upuparrows$	$0.018 \pm 0.070 \upuparrows$
	ADPN-F	$\mathbf{0.539 \pm 0.421}$	$0.011 \pm 0.061 \upuparrows$	$0.344 \pm 0.391 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.033 \pm 0.127 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	LATEX-F	$\mathbf{0.251 \pm 0.261}$	$0.000 \pm 0.000 \upuparrows$	$0.098 \pm 0.172 \upuparrows$	$0.013 \pm 0.052 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.015 \pm 0.057 \upuparrows$	$0.007 \pm 0.037 \upuparrows$
	PAPER-F	0.236 ± 0.197	$0.158 \pm 0.161 \upuparrows$	0.187 ± 0.194	$0.092 \pm 0.149 \upuparrows$	$0.025 \pm 0.101 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$\mathbf{0.252 \pm 0.267}$	$0.104 \pm 0.179 \upuparrows$	$0.024 \pm 0.093 \upuparrows$
	SPIRA-F	0.242 ± 0.275	$0.000 \pm 0.000 \upuparrows$	$\mathbf{0.269 \pm 0.337}$	$0.017 \pm 0.063 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.028 \pm 0.108 \upuparrows$	$0.000 \pm 0.000 \upuparrows$

Table 3: Comparison of overall KPF selection performance of obtained non-dominated sets between NSGAII-MIIP and benchmark algorithms using the F1-score measure.

Metric	Dataset	NSGAII-MIIP	GADMS-IPM	MOPSO-LS	MOFS-BDE	NSGAII-IPM	IDMS-IPM	NSGA-II/SDR	MOEA/D	SPEA2
HV	ADPN	$\mathbf{0.950 \pm 0.293}$	$0.843 \pm 0.421 \upuparrows$	$0.855 \pm 0.453 \uparrow$	0.941 ± 0.254	$0.805 \pm 0.394 \upuparrows$	$0.171 \pm 0.155 \upuparrows$	0.815 ± 0.409	0.885 ± 0.403	$0.771 \pm 0.376 \upuparrows$
	LATEX	$\mathbf{0.986 \pm 0.142}$	$0.867 \pm 0.158 \upuparrows$	0.964 ± 0.152	0.920 ± 0.163	$0.705 \pm 0.146 \upuparrows$	$0.247 \pm 0.091 \upuparrows$	$0.876 \pm 0.130 \upuparrows$	$0.882 \pm 0.151 \upuparrows$	$0.795 \pm 0.126 \upuparrows$
	PAPER	0.882 ± 0.135	$\mathbf{0.898 \pm 0.138}$	0.880 ± 0.133	0.839 ± 0.149	0.842 ± 0.139	$0.693 \pm 0.169 \upuparrows$	$0.833 \pm 0.186 \uparrow$	$0.834 \pm 0.142 \upuparrows$	$0.825 \pm 0.127 \upuparrows$
	SPIRA	$\mathbf{0.992 \pm 0.148}$	$0.936 \pm 0.158 \upuparrows$	0.977 ± 0.148	$0.934 \pm 0.178 \uparrow$	$0.827 \pm 0.150 \upuparrows$	$0.226 \pm 0.126 \upuparrows$	$0.895 \pm 0.193 \upuparrows$	$0.878 \pm 0.129 \upuparrows$	$0.869 \pm 0.129 \upuparrows$
	ADPN-F	0.977 ± 0.297	$0.752 \pm 0.388 \upuparrows$	$\mathbf{0.989 \pm 0.266}$	$0.872 \pm 0.277 \upuparrows$	$0.548 \pm 0.484 \upuparrows$	$0.146 \pm 0.127 \upuparrows$	$0.537 \pm 0.666 \upuparrows$	$0.797 \pm 0.367 \upuparrows$	$0.602 \pm 0.416 \upuparrows$
	LATEX-F	$\mathbf{1.016 \pm 0.105}$	$0.794 \pm 0.111 \upuparrows$	$0.964 \pm 0.118 \upuparrows$	$0.853 \pm 0.133 \upuparrows$	$0.649 \pm 0.109 \upuparrows$	$0.234 \pm 0.084 \upuparrows$	$0.794 \pm 0.132 \upuparrows$	$0.913 \pm 0.119 \upuparrows$	$0.712 \pm 0.136 \upuparrows$
	PAPER-F	0.918 ± 0.155	0.910 ± 0.127	0.891 ± 0.162	$\mathbf{0.931 \pm 0.114}$	$0.768 \pm 0.125 \upuparrows$	$0.358 \pm 0.188 \upuparrows$	0.902 ± 0.187	$0.827 \pm 0.108 \upuparrows$	$0.857 \pm 0.147 \upuparrows$
	SPIRA-F	$\mathbf{1.029 \pm 0.109}$	$0.832 \pm 0.142 \upuparrows$	0.972 ± 0.195	$0.910 \pm 0.149 \upuparrows$	$0.683 \pm 0.145 \upuparrows$	$0.209 \pm 0.108 \upuparrows$	$0.864 \pm 0.150 \upuparrows$	$0.898 \pm 0.202 \upuparrows$	$0.724 \pm 0.158 \upuparrows$
IGD	ADPN	$\mathbf{0.168 \pm 0.206}$	$0.294 \pm 0.353 \upuparrows$	0.256 ± 0.349	0.202 ± 0.162	$0.340 \pm 0.325 \upuparrows$	$0.967 \pm 0.300 \upuparrows$	0.273 ± 0.347	0.245 ± 0.299	$0.349 \pm 0.328 \upuparrows$
	LATEX	$\mathbf{0.116 \pm 0.065}$	$0.201 \pm 0.102 \upuparrows$	0.131 ± 0.089	$0.183 \pm 0.117 \upuparrows$	$0.314 \pm 0.117 \upuparrows$	$0.787 \pm 0.102 \upuparrows$	$0.194 \pm 0.101 \upuparrows$	$0.198 \pm 0.131 \upuparrows$	$0.268 \pm 0.094 \upuparrows$
	PAPER	0.201 ± 0.081	$\mathbf{0.200 \pm 0.094}$	0.211 ± 0.100	$0.257 \pm 0.080 \upuparrows$	$0.244 \pm 0.102 \upuparrows$	$0.326 \pm 0.129 \upuparrows$	$0.261 \pm 0.122 \upuparrows$	$0.259 \pm 0.121 \upuparrows$	$0.240 \pm 0.070 \upuparrows$
	SPIRA	$\mathbf{0.128 \pm 0.074}$	$0.171 \pm 0.076 \upuparrows$	0.140 ± 0.069	$0.161 \pm 0.102 \upuparrows$	$0.246 \pm 0.078 \upuparrows$	$0.811 \pm 0.148 \upuparrows$	$0.188 \pm 0.108 \upuparrows$	$0.205 \pm 0.073 \upuparrows$	$0.224 \pm 0.076 \upuparrows$
	ADPN-F	$\mathbf{0.132 \pm 0.203}$	$0.357 \pm 0.331 \upuparrows$	0.159 ± 0.200	$0.220 \pm 0.183 \upuparrows$	$0.541 \pm 0.449 \upuparrows$	$0.964 \pm 0.188 \upuparrows$	$0.516 \pm 0.604 \upuparrows$	$0.288 \pm 0.329 \upuparrows$	$0.475 \pm 0.379 \upuparrows$
	LATEX-F	$\mathbf{0.138 \pm 0.107}$	$0.270 \pm 0.095 \upuparrows$	$0.176 \pm 0.088 \upuparrows$	$0.243 \pm 0.097 \upuparrows$	$0.379 \pm 0.087 \upuparrows$	$0.798 \pm 0.099 \upuparrows$	$0.272 \pm 0.093 \upuparrows$	$0.198 \pm 0.083 \upuparrows$	$0.340 \pm 0.083 \upuparrows$
	PAPER-F	0.152 ± 0.084	0.153 ± 0.070	0.163 ± 0.093	$\mathbf{0.151 \pm 0.055}$	$0.254 \pm 0.077 \upuparrows$	$0.651 \pm 0.175 \upuparrows$	0.166 ± 0.091	$0.204 \pm 0.071 \upuparrows$	$0.226 \pm 0.073 \upuparrows$
	SPIRA-F	$\mathbf{0.116 \pm 0.069}$	$0.242 \pm 0.081 \upuparrows$	0.145 ± 0.117	$0.204 \pm 0.086 \upuparrows$	$0.354 \pm 0.076 \upuparrows$	$0.864 \pm 0.128 \upuparrows$	$0.238 \pm 0.088 \upuparrows$	$0.222 \pm 0.121 \upuparrows$	$0.345 \pm 0.100 \upuparrows$
SC	ADPN	$\mathbf{0.478 \pm 0.452}$	$0.139 \pm 0.263 \upuparrows$	0.317 ± 0.402	$0.172 \pm 0.261 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.222 \pm 0.362 \upuparrows$	$0.056 \pm 0.216 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	LATEX	$\mathbf{0.224 \pm 0.182}$	$0.033 \pm 0.127 \upuparrows$	0.205 ± 0.211	$0.035 \pm 0.073 \upuparrows$	$0.008 \pm 0.046 \upuparrows$	$0.005 \pm 0.026 \upuparrows$	$0.071 \pm 0.112 \upuparrows$	$0.032 \pm 0.091 \upuparrows$	$0.011 \pm 0.044 \upuparrows$
	PAPER	0.142 ± 0.168	$0.082 \pm 0.143 \upuparrows$	0.113 ± 0.155	$\mathbf{0.197 \pm 0.233}$	0.144 ± 0.267	$0.047 \pm 0.127 \upuparrows$	0.126 ± 0.180	0.124 ± 0.159	$0.046 \pm 0.106 \upuparrows$
	SPIRA	$\mathbf{0.301 \pm 0.272}$	$0.136 \pm 0.175 \upuparrows$	0.288 ± 0.272	0.227 ± 0.226	$0.038 \pm 0.122 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.151 \pm 0.228 \upuparrows$	$0.058 \pm 0.120 \upuparrows$	$0.018 \pm 0.070 \upuparrows$
	ADPN-F	$\mathbf{0.550 \pm 0.411}$	$0.011 \pm 0.061 \upuparrows$	$0.344 \pm 0.391 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.033 \pm 0.127 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	LATEX-F	$\mathbf{0.249 \pm 0.255}$	$0.000 \pm 0.000 \upuparrows$	$0.086 \pm 0.149 \upuparrows$	$0.022 \pm 0.069 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.007 \pm 0.037 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.007 \pm 0.037 \upuparrows$	$0.007 \pm 0.037 \upuparrows$
	PAPER-F	0.189 ± 0.199	0.131 ± 0.143	0.153 ± 0.181	0.092 ± 0.149	$0.025 \pm 0.101 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$\mathbf{0.238 \pm 0.269}$	0.101 ± 0.176	$0.036 \pm 0.109 \upuparrows$
	SPIRA-F	0.242 ± 0.275	$0.000 \pm 0.000 \upuparrows$	$\mathbf{0.269 \pm 0.337}$	$0.017 \pm 0.063 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.028 \pm 0.108 \upuparrows$	$0.000 \pm 0.000 \upuparrows$

5. Further analysis of the FS performance on public benchmark datasets

5.1. Experimental design

We apply four public datasets from the UCI machine learning repository (https://archive.ics.uci.edu/) (Kelly et al., 2024) to verify the proposed NSGAII-MIIP algorithm further. The details of the four datasets are shown in Table 4. The number of features in these datasets varies from 147 to 1558, higher than the four manufacturing process datasets in the manuscript. The number of instances in the datasets except for LSVT is much higher than that in the manufacturing process datasets. Moreover, the four datasets include both the binary and multi-class types. The datasets except for Mfeat are unbalanced and the imbalance ratios (ratio of the number of instances of the largest majority class to that of the smallest minority class) of these datasets are shown in Table 4.

The FS methods used in the experiments include NSGAII-MIIP, GADMS-IPM (Li et al., 2020), MOPSO-LS (He et al., 2022), MOFS-BDE (Zhang et al., 2020), NSGAII-IPM (Li et al., 2016), NSGA-II/SDR (Tian et al., 2019), MOEA/D (Zhang & Li, 2007), and SPEA2 (Zitzler et al., 2001), which are also used in the manuscript. The results in the manuscript have shown that IDMS-IPM converges much slower than NSGAII-MIIP. Therefore, IDMS-IPM is not used on the four public datasets with a higher dimensionality as it needs a large computational budget on these datasets. Note that MOFS-BDE and NSGAII-IPM are

Table 4: Details of the datasets from the UCI machine learning repository.

Dataset (Abbreviation)	Dataset (Full Name)	#Features	#Instances	#Classes	Imbalance Ratio
LSVT	LSVT Voice Rehabilitation	310	126	2	2.00
InterAd	Internet Advertisements	1558	3279	2	6.14
UrbanLC	Urban Land Cover	147	675	9	3.86
Mfeat	Multiple Features	649	2000	10	1.00

applied on the same KPF/feature selection model as other algorithms to facilitate the comparison, the same as the setting described in Section 7.3 of the manuscript. The parameter settings (except for the stopping criterion) of these algorithms are the same as those described in Section 5.3 of the manuscript. The maximum number of objective function evaluations (stopping criterion) is set as 5,000 for UrbanLC (the same as the manufacturing process datasets in the manuscript), and it is increased to 10,000 for LSVT, InterAd, and Mfeat because the three datasets, which have substantially more features, require more computational resources. As the four datasets generally have considerably more instances than the manufacturing process datasets in the manuscript, we use the holdout validation in the experiments. It divides a dataset into a training set (70%) and a test test (30%). The training set is input to the FS methods to select key features and the test test is used to evaluate the prediction performance of selected features. These FS methods run 30 times on each dataset, yielding 30 sets of experimental results. Similar to the manuscript, the Wilcoxon signed-rank test is used to compare the performance between NSGAII-MIIP and benchmark algorithms, the sign $\upuparrows$ ($\uparrow$) or $\downdownarrows$ ($\downarrow$) denote that NSGAII-MIIP is significantly better or worse than the benchmark algorithm at a significance level of 0.05 (0.1).

The performance metrics to evaluate the FS performance include GM, the F1-score, AUC, and the number of selected features (#SFs), the same as that used in the manuscript for the binary datasets LSVT and InterAd. For the multi-class datasets UrbanLC and Mfeat, the calculations of GM, F1-score, and AUC metrics are slightly different. Specifically, GM is calculated as the geometric mean of the recall values of the multiple classes. This GM measure is also used for objective function evaluation in FS methods for UrbanLC and Mfeat. The F1-score is the average of the F1-score values for the multiple classes, i.e., the macro-averaged F1-score is used. AUC is calculated as the average of the one-vs-rest AUC values in terms of multiple classes. A one-vs-rest AUC value is computed as the AUC of a class against the rest. For a detailed description of these measures, please refer to the user guide of scikit-learn (Pedregosa et al., 2011).

5.2. Comparison of FS performance of solutions selected by the ideal point method (IPM)

In all these multi-objective FS methods, the IPM is used in the second phase to select the final solution. In this section, the FS performance of these FS methods is compared based on the solutions obtained by the IPM. Table 5 shows the comparison results between NSGAII-MIIP and benchmark algorithms on the four UCI datasets. Overall, NSGAII-MIIP obtains better FS performance on all four datasets compared with GADMS-IPM, NSGAII-IPM, NSGA-II/SDR, MOEA/D, and SPEA2 because it obtains similar or significantly higher GM, F1-score, and AUC results while selecting significantly fewer features. Compared with MOPSO-LS, NGAII-MIIP obtains significantly better GM, F1-score, and AUC results and selects fewer features on the datasets except for InterAd. On InterAd, NSGAII-MIIP obtains better GM, F1-score, and AUC results while selecting more features than MOPSO-LS. This denotes that, on InterAd, the feature reduction performance of NSGAII-MIIP is not better than MOPSO-LS. Compared with MOFS-BDE, NSGAII-MIIP obtains similar or better GM, F1-score, and AUC results on LSVT and UrbanLC while selecting significantly fewer features. On InterAd, NSGAII-MIIP obtains significantly better F1-score and AUC values and a significantly lower GM value. Meanwhile, NSGAII-MIIP selects more features than MOFS-BDE. On Mfeat, MOFS-BDE shows better FS performance than NSGAII-MIIP as it obtains significantly better GM and F1-score results while selecting a similar number of features to NSGAII-MIIP. We find that on the two most high-dimensional datasets, InterAd and Mfeat, NSGAII-MIIP does not show better FS performance than MOFS-BDE.

Table 5: Comparison of FS selection performance of final solutions between NSGAII-MIIP and benchmark algorithms on the UCI datasets.

Dataset	Metric	NSGAII-MIIP	GADMS-IPM	MOPSO-LS	MOFS-BDE	NSGAII-IPM	NSGA-II/SDR	MOEA/D	SPEA2
LSVT	GM (%)	$\mathbf{82.21 \pm 6.02}$	$67.81 \pm 5.54 \upuparrows$	$74.96 \pm 7.06 \upuparrows$	79.89 ± 6.88	$68.28 \pm 2.47 \upuparrows$	$72.10 \pm 9.27 \upuparrows$	$68.26 \pm 2.30 \upuparrows$	$68.03 \pm 3.81 \upuparrows$
	F1-score (%)	$\mathbf{76.06 \pm 7.59}$	$57.57 \pm 5.86 \upuparrows$	$66.87 \pm 8.54 \upuparrows$	72.37 ± 9.08	$59.96 \pm 2.37 \upuparrows$	$62.82 \pm 10.77 \upuparrows$	$59.96 \pm 2.23 \upuparrows$	$58.37 \pm 3.91 \upuparrows$
	AUC (%)	$\mathbf{88.16 \pm 4.52}$	$80.39 \pm 4.14 \upuparrows$	$85.23 \pm 4.90 \upuparrows$	86.49 ± 4.85	$73.03 \pm 1.99 \upuparrows$	$83.16 \pm 7.37 \upuparrows$	$72.15 \pm 1.61 \upuparrows$	$81.08 \pm 4.50 \upuparrows$
	#SFs	$\mathbf{2.5 \pm 0.5}$	$15.5 \pm 5.0 \upuparrows$	$4.7 \pm 3.2 \upuparrows$	$12.0 \pm 2.8 \upuparrows$	$73.1 \pm 4.7 \upuparrows$	$6.6 \pm 3.1 \upuparrows$	$120.9 \pm 7.2 \upuparrows$	$27.2 \pm 3.5 \upuparrows$
InterAd	GM (%)	92.92 ± 0.74	93.15 ± 0.65	$91.72 \pm 1.33 \upuparrows$	$\mathbf{93.47 \pm 0.60} \downdownarrows$	$92.56 \pm 0.62 \uparrow$	92.95 ± 0.71	$91.40 \pm 0.87 \upuparrows$	92.78 ± 0.70
	F1-score (%)	90.11 ± 0.85	$\mathbf{90.15 \pm 0.91}$	$87.30 \pm 2.18 \upuparrows$	$89.43 \pm 1.08 \upuparrows$	89.89 ± 0.95	90.09 ± 0.86	$88.66 \pm 1.11 \upuparrows$	$89.83 \pm 0.86 \uparrow$
	AUC (%)	$\mathbf{97.51 \pm 0.45}$	97.50 ± 0.50	$97.14 \pm 0.64 \upuparrows$	$97.11 \pm 0.49 \upuparrows$	97.48 ± 0.51	97.26 ± 0.62	97.50 ± 0.49	97.43 ± 0.52
	#SFs	484.1 ± 16.2	$531.1 \pm 15.7 \upuparrows$	$\mathbf{452.2 \pm 14.3} \downdownarrows$	$459.3 \pm 17.5 \downdownarrows$	$611.4 \pm 11.8 \upuparrows$	$513.3 \pm 18.3 \upuparrows$	$699.3 \pm 17.3 \upuparrows$	$559.6 \pm 13.9 \upuparrows$
UrbanLC	GM (%)	84.14 ± 2.30	84.95 ± 2.35	$82.55 \pm 3.39 \uparrow$	84.80 ± 2.00	84.96 ± 2.26	83.98 ± 2.21	83.66 ± 2.02	$\mathbf{85.35 \pm 1.65} \downdownarrows$
	F1-score (%)	85.00 ± 2.16	85.43 ± 2.17	$83.71 \pm 3.02 \uparrow$	85.10 ± 1.88	85.45 ± 2.29	84.55 ± 1.87	84.27 ± 1.79	$\mathbf{85.54 \pm 1.57}$
	AUC (%)	$\mathbf{98.64 \pm 0.27}$	98.50 ± 0.39	$98.27 \pm 0.51 \upuparrows$	$98.44 \pm 0.39 \upuparrows$	$98.33 \pm 0.33 \upuparrows$	$98.36 \pm 0.36 \upuparrows$	$98.29 \pm 0.36 \upuparrows$	$98.32 \pm 0.36 \upuparrows$
	#SFs	$\mathbf{5.1 \pm 1.1}$	$11.5 \pm 2.3 \upuparrows$	5.3 ± 1.3	$13.5 \pm 2.2 \upuparrows$	$23.0 \pm 2.8 \upuparrows$	$13.6 \pm 4.2 \upuparrows$	$11.1 \pm 3.5 \upuparrows$	$19.6 \pm 3.1 \upuparrows$
Mfeat	GM (%)	96.52 ± 0.38	96.64 ± 0.37	$96.06 \pm 0.56 \upuparrows$	$\mathbf{97.08 \pm 0.33} \downdownarrows$	$96.17 \pm 0.39 \upuparrows$	$96.22 \pm 0.43 \upuparrows$	$95.90 \pm 0.45 \upuparrows$	96.40 ± 0.47
	F1-score (%)	96.57 ± 0.37	96.68 ± 0.36	$96.12 \pm 0.56 \upuparrows$	$\mathbf{97.11 \pm 0.32} \downdownarrows$	$96.22 \pm 0.38 \upuparrows$	$96.27 \pm 0.42 \upuparrows$	$95.96 \pm 0.45 \upuparrows$	96.45 ± 0.47
	AUC (%)	99.70 ± 0.11	$99.65 \pm 0.11 \uparrow$	$99.57 \pm 0.13 \upuparrows$	$\mathbf{99.71 \pm 0.08}$	$99.56 \pm 0.09 \upuparrows$	$99.57 \pm 0.13 \upuparrows$	$99.49 \pm 0.07 \upuparrows$	$99.58 \pm 0.10 \upuparrows$
	#SFs	126.9 ± 12.5	$152.6 \pm 7.3 \upuparrows$	$138.5 \pm 10.4 \upuparrows$	$\mathbf{124.9 \pm 7.8}$	$220.3 \pm 9.8 \upuparrows$	$169.5 \pm 10.4 \upuparrows$	$275.4 \pm 9.9 \upuparrows$	$182.0 \pm 10.6 \upuparrows$

Table 6: Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP and benchmark algorithms on the UCI datasets using the AUC measure.

Metric	Dataset	NSGAII-MIIP	GADMS-IPM	MOPSO-LS	MOFS-BDE	NSGAII-IPM	NSGA-II/SDR	MOEA/D	SPEA2
HV	LSVT	$\mathbf{1.122 \pm 0.016}$	$0.833 \pm 0.104 \upuparrows$	$1.040 \pm 0.086 \upuparrows$	$1.032 \pm 0.090 \upuparrows$	$0.383 \pm 0.045 \upuparrows$	$0.918 \pm 0.188 \upuparrows$	$0.200 \pm 0.029 \upuparrows$	$0.770 \pm 0.087 \upuparrows$
	InterAd	0.789 ± 0.098	$0.675 \pm 0.089 \upuparrows$	$\mathbf{0.954 \pm 0.083} \downdownarrows$	$0.906 \pm 0.076 \downdownarrows$	$0.492 \pm 0.056 \upuparrows$	$0.679 \pm 0.097 \upuparrows$	$0.275 \pm 0.055 \upuparrows$	$0.607 \pm 0.072 \upuparrows$
	UrbanLC	$\mathbf{1.167 \pm 0.007}$	$1.023 \pm 0.043 \upuparrows$	$1.152 \pm 0.010 \upuparrows$	$0.996 \pm 0.042 \upuparrows$	$0.773 \pm 0.065 \upuparrows$	$0.986 \pm 0.064 \upuparrows$	$1.085 \pm 0.038 \upuparrows$	$0.864 \pm 0.064 \upuparrows$
	Mfeat	0.863 ± 0.159	$0.702 \pm 0.122 \upuparrows$	$0.780 \pm 0.100 \upuparrows$	$\mathbf{0.965 \pm 0.071} \downdownarrows$	$0.383 \pm 0.081 \upuparrows$	$0.579 \pm 0.108 \upuparrows$	$0.179 \pm 0.041 \upuparrows$	$0.530 \pm 0.111 \upuparrows$
IGD	LSVT	$\mathbf{0.053 \pm 0.008}$	$0.242 \pm 0.088 \upuparrows$	$0.104 \pm 0.064 \upuparrows$	$0.116 \pm 0.052 \upuparrows$	$0.631 \pm 0.051 \upuparrows$	$0.201 \pm 0.145 \upuparrows$	$0.879 \pm 0.041 \upuparrows$	$0.282 \pm 0.066 \upuparrows$
	InterAd	0.247 ± 0.034	$0.306 \pm 0.044 \upuparrows$	$\mathbf{0.159 \pm 0.039} \downdownarrows$	$0.194 \pm 0.037 \downdownarrows$	$0.471 \pm 0.035 \upuparrows$	$0.302 \pm 0.055 \upuparrows$	$0.693 \pm 0.050 \upuparrows$	$0.357 \pm 0.035 \upuparrows$
	UrbanLC	0.076 ± 0.019	$0.129 \pm 0.023 \upuparrows$	$\mathbf{0.075 \pm 0.019}$	$0.145 \pm 0.025 \upuparrows$	$0.303 \pm 0.055 \upuparrows$	$0.156 \pm 0.047 \upuparrows$	$0.103 \pm 0.020 \upuparrows$	$0.234 \pm 0.049 \upuparrows$
	Mfeat	0.189 ± 0.128	$0.314 \pm 0.103 \upuparrows$	$0.275 \pm 0.090 \upuparrows$	$\mathbf{0.127 \pm 0.045} \downarrow$	$0.621 \pm 0.080 \upuparrows$	$0.415 \pm 0.098 \upuparrows$	$0.888 \pm 0.064 \upuparrows$	$0.465 \pm 0.097 \upuparrows$
SC	LSVT	$\mathbf{0.067 \pm 0.096}$	$0.000 \pm 0.000 \upuparrows$	$0.013 \pm 0.051 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.020 \pm 0.061 \uparrow$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	InterAd	0.000 ± 0.000	0.000 ± 0.000	$\mathbf{0.015 \pm 0.041}$	0.009 ± 0.026	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.009 ± 0.051
	UrbanLC	$\mathbf{0.081 \pm 0.071}$	$0.000 \pm 0.000 \upuparrows$	$0.007 \pm 0.028 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	Mfeat	$\mathbf{0.030 \pm 0.099}$	0.000 ± 0.000	0.000 ± 0.000	0.003 ± 0.018	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000

Table 7: Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP and benchmark algorithms on the UCI datasets using the GM measure.

Metric	Dataset	NSGAII-MIIP	GADMS-IPM	MOPSO-LS	MOFS-BDE	NSGAII-IPM	NSGA-II/SDR	MOEA/D	SPEA2
HV	LSVT	$\mathbf{1.070 \pm 0.054}$	$0.686 \pm 0.108 \upuparrows$	$0.944 \pm 0.123 \upuparrows$	$0.990 \pm 0.089 \upuparrows$	$0.419 \pm 0.038 \upuparrows$	$0.828 \pm 0.200 \upuparrows$	$0.225 \pm 0.033 \upuparrows$	$0.626 \pm 0.058 \upuparrows$
	InterAd	0.842 ± 0.062	$0.730 \pm 0.064 \upuparrows$	$0.959 \pm 0.052 \downdownarrows$	$\mathbf{1.050 \pm 0.049} \downdownarrows$	$0.494 \pm 0.047 \upuparrows$	$0.758 \pm 0.057 \upuparrows$	$0.242 \pm 0.053 \upuparrows$	$0.649 \pm 0.039 \upuparrows$
	UrbanLC	$\mathbf{1.117 \pm 0.028}$	$0.993 \pm 0.054 \upuparrows$	$1.095 \pm 0.034 \upuparrows$	$0.974 \pm 0.049 \upuparrows$	$0.762 \pm 0.063 \upuparrows$	$0.950 \pm 0.074 \upuparrows$	$1.039 \pm 0.054 \upuparrows$	$0.848 \pm 0.064 \upuparrows$
	Mfeat	0.842 ± 0.103	$0.745 \pm 0.082 \upuparrows$	0.793 ± 0.092	$\mathbf{1.032 \pm 0.063} \downdownarrows$	$0.415 \pm 0.058 \upuparrows$	$0.622 \pm 0.086 \upuparrows$	$0.198 \pm 0.044 \upuparrows$	$0.601 \pm 0.078 \upuparrows$
IGD	LSVT	$\mathbf{0.070 \pm 0.029}$	$0.343 \pm 0.097 \upuparrows$	$0.154 \pm 0.091 \upuparrows$	$0.124 \pm 0.047 \upuparrows$	$0.575 \pm 0.039 \upuparrows$	$0.250 \pm 0.150 \upuparrows$	$0.830 \pm 0.045 \upuparrows$	$0.376 \pm 0.052 \upuparrows$
	InterAd	0.218 ± 0.037	$0.319 \pm 0.044 \upuparrows$	$0.151 \pm 0.031 \downdownarrows$	$\mathbf{0.121 \pm 0.020} \downdownarrows$	$0.535 \pm 0.038 \upuparrows$	$0.284 \pm 0.044 \upuparrows$	$0.787 \pm 0.054 \upuparrows$	$0.391 \pm 0.031 \upuparrows$
	UrbanLC	$\mathbf{0.118 \pm 0.044}$	0.130 ± 0.026	0.128 ± 0.039	0.128 ± 0.023	$0.224 \pm 0.043 \upuparrows$	$0.153 \pm 0.033 \upuparrows$	0.120 ± 0.036	$0.177 \pm 0.031 \upuparrows$
	Mfeat	0.188 ± 0.056	$0.261 \pm 0.044 \upuparrows$	$0.252 \pm 0.059 \upuparrows$	$\mathbf{0.116 \pm 0.028} \downdownarrows$	$0.570 \pm 0.048 \upuparrows$	$0.353 \pm 0.061 \upuparrows$	$0.831 \pm 0.059 \upuparrows$	$0.386 \pm 0.057 \upuparrows$
SC	LSVT	$\mathbf{0.083 \pm 0.120}$	$0.000 \pm 0.000 \upuparrows$	$0.025 \pm 0.101 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.008 \pm 0.046 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	InterAd	0.000 ± 0.000	0.000 ± 0.000	0.006 ± 0.023	$\mathbf{0.027 \pm 0.054} \downdownarrows$	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	UrbanLC	$\mathbf{0.017 \pm 0.034}$	0.011 ± 0.048	0.003 ± 0.015	0.003 ± 0.015	$0.000 \pm 0.000 \upuparrows$	0.003 ± 0.015	0.003 ± 0.015	$0.000 \pm 0.000 \upuparrows$
	Mfeat	0.007 ± 0.025	0.000 ± 0.000	0.000 ± 0.000	$\mathbf{0.027 \pm 0.052}$	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000

5.3. Comparison of overall FS performance of obtained non-dominated sets

In this section, the FS performance of these algorithms is compared based on the non-dominated sets found in the first phase. The comparison results between NSGAII-MIIP and benchmark algorithms on the AUC, GM, and F1-score measures are shown in Tables 6, 7, and 8 respectively. As the comparison results on the three measures are consistent, we take the results of AUC in Table 6 to analyze. According to Table 6, NSGAII-MIIP generally obtains significantly better FS performance than the benchmark algorithms except for MOPSO-LS and MOFS-BDE on all four UCI datasets. Specifically, NSGAII-MIIP obtains significantly better HV and IGD results on all four datasets, and obtains significantly better SC results on two of the four datasets. Compared with MOPSO-LS, the HV results of NSGAII-MIIP are significantly better on the datasets except for InterAd, and the IGD and SC results of NSGAII-MIIP are significantly better on two datasets. NSGAII-MIIP only obtains significantly worse HV and IGD values than MOPSO-LS on InterAd. Compared with MOFS-BDE, NSGAII-MIIP obtains significantly better HV, IGD, and SC results on two datasets, LSVT and UrbanLC. On InterAd and Mfeat, NSGAII-MIIP obtains significantly worse HV and IGD values than MOFS-BDE. This denotes that the FS performance of NSGAII-MIIP is worse than MOFS-BDE on the two most high-dimensional datasets, InterAd and Mfeat.

Table 8: Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP and benchmark algorithms on the UCI datasets using the F1-score measure.

Metric	Dataset	NSGAII-MIIP	GADMS-IPM	MOPSO-LS	MOFS-BDE	NSGAII-IPM	NSGA-II/SDR	MOEA/D	SPEA2
HV	LSVT	$\mathbf{1.022 \pm 0.054}$	$0.578 \pm 0.107 \upuparrows$	$0.855 \pm 0.145 \upuparrows$	$0.937 \pm 0.113 \upuparrows$	$0.366 \pm 0.035 \upuparrows$	$0.729 \pm 0.220 \upuparrows$	$0.198 \pm 0.028 \upuparrows$	$0.532 \pm 0.054 \upuparrows$
	InterAd	0.828 ± 0.059	$0.708 \pm 0.062 \upuparrows$	$0.895 \pm 0.087 \downdownarrows$	$\mathbf{0.976 \pm 0.055} \downdownarrows$	$0.492 \pm 0.052 \upuparrows$	$0.744 \pm 0.059 \upuparrows$	$0.243 \pm 0.052 \upuparrows$	$0.632 \pm 0.050 \upuparrows$
	UrbanLC	$\mathbf{1.125 \pm 0.019}$	$0.993 \pm 0.049 \upuparrows$	$1.109 \pm 0.026 \upuparrows$	$0.969 \pm 0.045 \upuparrows$	$0.758 \pm 0.060 \upuparrows$	$0.955 \pm 0.066 \upuparrows$	$1.046 \pm 0.046 \upuparrows$	$0.845 \pm 0.065 \upuparrows$
	Mfeat	0.842 ± 0.102	$0.746 \pm 0.082 \upuparrows$	0.792 ± 0.091	$\mathbf{1.032 \pm 0.063} \downdownarrows$	$0.415 \pm 0.058 \upuparrows$	$0.622 \pm 0.085 \upuparrows$	$0.198 \pm 0.044 \upuparrows$	$0.602 \pm 0.079 \upuparrows$
IGD	LSVT	$\mathbf{0.079 \pm 0.029}$	$0.404 \pm 0.098 \upuparrows$	$0.194 \pm 0.109 \upuparrows$	$0.143 \pm 0.062 \upuparrows$	$0.602 \pm 0.038 \upuparrows$	$0.301 \pm 0.168 \upuparrows$	$0.847 \pm 0.043 \upuparrows$	$0.429 \pm 0.051 \upuparrows$
	InterAd	0.250 ± 0.025	$0.326 \pm 0.033 \upuparrows$	$0.187 \pm 0.048 \downdownarrows$	$\mathbf{0.184 \pm 0.024} \downdownarrows$	$0.498 \pm 0.033 \upuparrows$	$0.297 \pm 0.033 \upuparrows$	$0.731 \pm 0.053 \upuparrows$	$0.378 \pm 0.029 \upuparrows$
	UrbanLC	0.142 ± 0.038	$0.170 \pm 0.022 \upuparrows$	$\mathbf{0.135 \pm 0.030}$	$0.179 \pm 0.019 \upuparrows$	$0.283 \pm 0.042 \upuparrows$	$0.194 \pm 0.025 \upuparrows$	0.150 ± 0.029	$0.233 \pm 0.032 \upuparrows$
	Mfeat	0.191 ± 0.054	$0.263 \pm 0.043 \upuparrows$	$0.252 \pm 0.057 \upuparrows$	$\mathbf{0.122 \pm 0.028} \downdownarrows$	$0.570 \pm 0.048 \upuparrows$	$0.354 \pm 0.060 \upuparrows$	$0.831 \pm 0.060 \upuparrows$	$0.388 \pm 0.057 \upuparrows$
SC	LSVT	$\mathbf{0.083 \pm 0.120}$	$0.000 \pm 0.000 \upuparrows$	$0.025 \pm 0.101 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.008 \pm 0.046 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	InterAd	0.000 ± 0.000	0.000 ± 0.000	0.007 ± 0.025	$\mathbf{0.023 \pm 0.057} \downarrow$	0.003 ± 0.018	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	UrbanLC	$\mathbf{0.023 \pm 0.043}$	0.007 ± 0.025	0.007 ± 0.025	$0.003 \pm 0.018 \uparrow$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.003 \pm 0.018 \uparrow$	$0.000 \pm 0.000 \upuparrows$
	Mfeat	0.004 ± 0.020	0.000 ± 0.000	0.000 ± 0.000	$\mathbf{0.030 \pm 0.058} \downarrow$	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000

Table 9: Search performance comparison between NSGAII-MIIP and benchmark algorithms on the UCI datasets.

Metric	Dataset	NSGAII-MIIP	GADMS-IPM	MOPSO-LS	MOFS-BDE	NSGAII-IPM	NSGA-II/SDR	MOEA/D	SPEA2
HV	LSVT	$\mathbf{1.160 \pm 0.019}$	$0.835 \pm 0.085 \upuparrows$	$0.939 \pm 0.092 \upuparrows$	$1.068 \pm 0.032 \upuparrows$	$0.416 \pm 0.021 \upuparrows$	$0.931 \pm 0.118 \upuparrows$	$0.225 \pm 0.024 \upuparrows$	$0.720 \pm 0.033 \upuparrows$
	InterAd	0.883 ± 0.047	$0.776 \pm 0.048 \upuparrows$	$0.926 \pm 0.049 \downdownarrows$	$\mathbf{1.102 \pm 0.040} \downdownarrows$	$0.540 \pm 0.033 \upuparrows$	$0.792 \pm 0.048 \upuparrows$	$0.280 \pm 0.048 \upuparrows$	$0.702 \pm 0.035 \upuparrows$
	UrbanLC	$\mathbf{1.143 \pm 0.009}$	$1.022 \pm 0.037 \upuparrows$	$1.119 \pm 0.011 \upuparrows$	$0.999 \pm 0.037 \upuparrows$	$0.796 \pm 0.059 \upuparrows$	$0.981 \pm 0.056 \upuparrows$	$1.064 \pm 0.034 \upuparrows$	$0.871 \pm 0.058 \upuparrows$
	Mfeat	0.924 ± 0.082	$0.811 \pm 0.041 \upuparrows$	$0.784 \pm 0.047 \upuparrows$	$\mathbf{1.079 \pm 0.040} \downdownarrows$	$0.449 \pm 0.044 \upuparrows$	$0.706 \pm 0.037 \upuparrows$	$0.194 \pm 0.031 \upuparrows$	$0.671 \pm 0.050 \upuparrows$
IGD	LSVT	$\mathbf{0.016 \pm 0.008}$	$0.226 \pm 0.061 \upuparrows$	$0.166 \pm 0.070 \upuparrows$	$0.086 \pm 0.021 \upuparrows$	$0.583 \pm 0.020 \upuparrows$	$0.168 \pm 0.077 \upuparrows$	$0.836 \pm 0.037 \upuparrows$	$0.311 \pm 0.027 \upuparrows$
	InterAd	0.202 ± 0.030	$0.293 \pm 0.039 \upuparrows$	0.189 ± 0.031	$\mathbf{0.104 \pm 0.022} \downdownarrows$	$0.495 \pm 0.029 \upuparrows$	$0.265 \pm 0.040 \upuparrows$	$0.749 \pm 0.050 \upuparrows$	$0.359 \pm 0.030 \upuparrows$
	UrbanLC	$\mathbf{0.065 \pm 0.011}$	$0.116 \pm 0.016 \upuparrows$	$0.079 \pm 0.010 \upuparrows$	$0.129 \pm 0.016 \upuparrows$	$0.249 \pm 0.043 \upuparrows$	$0.141 \pm 0.026 \upuparrows$	$0.108 \pm 0.017 \upuparrows$	$0.199 \pm 0.036 \upuparrows$
	Mfeat	0.143 ± 0.050	$0.218 \pm 0.029 \upuparrows$	$0.282 \pm 0.045 \upuparrows$	$\mathbf{0.071 \pm 0.024} \downdownarrows$	$0.539 \pm 0.044 \upuparrows$	$0.307 \pm 0.035 \upuparrows$	$0.850 \pm 0.038 \upuparrows$	$0.345 \pm 0.040 \upuparrows$
SC	LSVT	$\mathbf{0.344 \pm 0.157}$	$0.000 \pm 0.000 \upuparrows$	$0.011 \pm 0.042 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	InterAd	0.000 ± 0.000	0.000 ± 0.000	0.006 ± 0.024	$\mathbf{0.027 \pm 0.104}$	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	UrbanLC	$\mathbf{0.035 \pm 0.061}$	$0.002 \pm 0.011 \upuparrows$	$0.006 \pm 0.019 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$	$0.000 \pm 0.000 \upuparrows$
	Mfeat	$\mathbf{0.017 \pm 0.071}$	0.000 ± 0.000	0.000 ± 0.000	0.017 ± 0.043	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000

5.4. Comparison of search performance

In this section, similar to the manuscript, we compare the search performance of these multi-objective evolutionary algorithms (MOEA) based on the objective function values (evaluated on the training set) of the found non-dominated solutions. The comparison results between NSGAII-MIIP and benchmark algorithms are shown in Table 9. Overall, NSGAII-MIIP shows better search performance than benchmark algorithms except for MOPSO-LS and MOFS-BDE on all four datasets. Specifically, NSGAII-MIIP obtains significantly better HV and IGD values on all four datasets than these benchmark algorithms. Compared with MOPSO-LS, NSGAII-MIIP shows better search performance on three datasets and obtains worse search performance on InterAd because the HV value of NSGAII-MIIP is significantly lower than that of MOPSO-LS on InterAd. Compared with MOFS-BDE, NSGAII-MIIP shows better search performance on two datasets, LSVT and UrbanLC. Meanwhile, NSGAII-MIIP shows worse search performance than MOFS-BDE on InterAd and Mfeat according to the results of HV and IGD. This denotes that the search performance of NSGAII-MIIP is not as good as MOFS-BDE on the two most high-dimensional datasets, InterAd and Mfeat.

5.5. Discussion

The above experimental results on the four UCI datasets indicate that NSGAII-MIIP performs most effectively on LSVT and UrbanLC which have a moderate number of features. However, for the two most high-dimensional datasets InterAd and Mfeat, it seems that NSGAII-MIIP does not obtain the best FS performance. On Mfeat, the FS performance of NSGAII-MIIP is slightly worse than MOFS-BDE. On InterAd, the FS performance of NSGAII-MIIP is slightly worse than MOPSO-LS and MOFS-BDE. The search performance results demonstrate that NSGAII-MIIP does not show a better search capability on these two datasets, which explains the worse FS performance of NSGAII-MIIP. The reason that MOPSO-LS shows better search performance than NSGAII-MIIP is that the initialized particles of MOPSO-LS select around 40% of features, which is fewer than that of the initialized solutions of NSGAII-MIIP (50% of features). Therefore, MOPSO-LS can obtain better FS performance on the most high-dimensional dataset InterAd given a relatively limited computational budget. The reason that NSGAII-MIIP does not obtain better FS results than MOFS-BDE is that the mutation rate for NSGAII-MIIP is set too low for these two high-dimensional datasets with the current computational budget. In MOFS-BDE, a turbulence coefficient of 0.01 (which has the same effect as the mutation operator in other MOEAs) was used. In comparison, the mutation rate of NSGAII-MIIP is set as $1/N$ (N is the number of original features), which is much

Table 10: Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP* and benchmark algorithms on InterAd and Mfeat using the AUC measure.

Metric	Dataset	NSGAII-MIIP*	MOPSO-LS	MOFS-BDE
HV	InterAd	$\mathbf{0.857 \pm 0.115}$	$0.430 \pm 0.048 \upuparrows$	$0.404 \pm 0.046 \upuparrows$
	Mfeat	$\mathbf{0.933 \pm 0.143}$	$0.344 \pm 0.060 \upuparrows$	$0.501 \pm 0.050 \upuparrows$
IGD	InterAd	$\mathbf{0.212 \pm 0.076}$	$0.573 \pm 0.045 \upuparrows$	$0.581 \pm 0.047 \upuparrows$
	Mfeat	$\mathbf{0.161 \pm 0.109}$	$0.675 \pm 0.084 \upuparrows$	$0.504 \pm 0.044 \upuparrows$
SC	InterAd	$\mathbf{0.033 \pm 0.103}$	0.000 ± 0.000	0.000 ± 0.000
	Mfeat	$\mathbf{0.033 \pm 0.097}$	0.000 ± 0.000	0.000 ± 0.000

Table 11: Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP* and benchmark algorithms on InterAd and Mfeat using the GM measure.

Metric	Dataset	NSGAII-MIIP*	MOPSO-LS	MOFS-BDE
HV	InterAd	$\mathbf{0.944 \pm 0.081}$	$0.488 \pm 0.039 \upuparrows$	$0.533 \pm 0.044 \upuparrows$
	Mfeat	$\mathbf{0.862 \pm 0.143}$	$0.335 \pm 0.054 \upuparrows$	$0.516 \pm 0.047 \upuparrows$
IGD	InterAd	0.309 ± 0.049	$0.354 \pm 0.036 \upuparrows$	$\mathbf{0.285 \pm 0.023} \downdownarrows$
	Mfeat	$\mathbf{0.252 \pm 0.106}$	$0.784 \pm 0.073 \upuparrows$	$0.578 \pm 0.046 \upuparrows$
SC	InterAd	$\mathbf{0.020 \pm 0.081}$	0.000 ± 0.000	0.013 ± 0.051
	Mfeat	$\mathbf{0.033 \pm 0.183}$	0.000 ± 0.000	0.000 ± 0.000

lower than the turbulence coefficient value of MOFS-BDE. This setting makes every mutation operation of NSGAII-MIIP only slightly change the selected features in a solution, limiting the algorithm's ability to escape from local optima on high-dimensional FS problems.

To justify our explanations above, we conduct a set of new experiments on InterAd and Mfeat. Specifically, the parameter setting of NSGAII-MIIP is adjusted by increasing the mutation rate to 0.01. The NSGAII-MIIP algorithm with this new parameter setting is denoted by NSGAII-MIIP*. The experimental results of NSGAII-MIIP* are further collected and compared to MOPSO-LS and MOFS-BDE. With this new setting, NSGAII-MIIP* outperforms MOPSO-LS and MOFS-BDE on both InterAd and Mfeat. The detailed results and analysis are shown below.

Tables 10, 11, and 12 show the comparison results of FS performance of obtained non-dominated sets between NSGAII-MIIP* and the two benchmark algorithms (MOPSO-LS and MOFS-BDE) on the AUC, GM, and F1-score measures. Overall, the results in the three tables indicate that the non-dominated solutions obtained by NSGAII-MIIP* show significantly better overall FS performance than those obtained by MOPSO-LS and MOFS-BDE, which demonstrates that NSGAII-MIIP* outperforms the two benchmark algorithms. Specifically, regarding AUC and the F1-score, NSGAII-MIIP* obtains significantly better HV and IGD results on both InterAd and Mfeat. Regarding GM, NSGAII-MIIP* obtains significantly better HV and IGD results than MOPSO-LS and MOFS-BDE, except that it obtains a significantly worse IGD value than MOFS-BDE on InterAd. Moreover, the SC results of NSGAII-MIIP* are better than those of MOPSO-LS and MOFS-BDE in all cases.

Table 13 shows the comparison results of search performance between NSGAII-MIIP* and the two benchmark algorithms. It is obvious that the HV and IGD values obtained by NSGAII-MIIP* are significantly better than MOPSO-LS and MOFS-BDE on both InterAd and Mfeat. Meanwhile, the SC values of NSGAII-MIIP* are higher than MOPSO-LS and MOFS-BDE on the two datasets. These results demonstrate that NSGAII-MIIP* has better search performance than MOPSO-LS and MOFS-BDE on the two high-dimensional datasets.

To conclude, the additional results in Tables 10 to 13 verify the effectiveness of the proposed NSGAII-MIIP algorithm. Moreover, these results also justify the conclusion that the mutation rate should be increased for an MOEA-based FS method on high-dimensional data to improve the FS performance if the computational budget is limited.

Table 12: Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP* and benchmark algorithms on InterAd and Mfeat using the F1-score measure.

Metric	Dataset	NSGAII-MIIP*	MOPSO-LS	MOFS-BDE
HV	InterAd	$\mathbf{0.883 \pm 0.086}$	$0.458 \pm 0.049 \upuparrows$	$0.500 \pm 0.043 \upuparrows$
	Mfeat	$\mathbf{0.861 \pm 0.143}$	$0.335 \pm 0.054 \upuparrows$	$0.516 \pm 0.047 \upuparrows$
IGD	InterAd	$\mathbf{0.241 \pm 0.044}$	$0.521 \pm 0.054 \upuparrows$	$0.445 \pm 0.038 \upuparrows$
	Mfeat	$\mathbf{0.252 \pm 0.106}$	$0.784 \pm 0.073 \upuparrows$	$0.578 \pm 0.047 \upuparrows$
SC	InterAd	$\mathbf{0.033 \pm 0.134}$	0.000 ± 0.000	0.000 ± 0.000
	Mfeat	$\mathbf{0.033 \pm 0.183}$	0.000 ± 0.000	0.000 ± 0.000

Table 13: Search performance comparison between NSGAII-MIIP* and benchmark algorithms on InterAd and Mfeat.

Metric	Dataset	NSGAII-MIIP*	MOPSO-LS	MOFS-BDE
HV	InterAd	$\mathbf{1.016 \pm 0.065}$	$0.549 \pm 0.036 \upuparrows$	$0.654 \pm 0.039 \upuparrows$
	Mfeat	$\mathbf{1.032 \pm 0.076}$	$0.490 \pm 0.044 \upuparrows$	$0.723 \pm 0.039 \upuparrows$
IGD	InterAd	$\mathbf{0.163 \pm 0.028}$	$0.382 \pm 0.036 \upuparrows$	$0.258 \pm 0.026 \upuparrows$
	Mfeat	$\mathbf{0.108 \pm 0.040}$	$0.558 \pm 0.051 \upuparrows$	$0.351 \pm 0.030 \upuparrows$
SC	InterAd	$\mathbf{0.022 \pm 0.056}$	$0.000 \pm 0.000 \uparrow$	0.011 ± 0.045
	Mfeat	$\mathbf{0.032 \pm 0.145}$	0.000 ± 0.000	0.002 ± 0.010

6. Ablation study

In this section, we establish two variants of NSGAII-MIIP, denoted by NSGAII-MIIP-N and NSGAII-MIIP-R, to verify the effectiveness of the MI-guided improvement strategy. NSGAII-MIIP-N does not use an improvement phase to purify the non-dominated solutions during the iteration process. NSGAII-MIIP-R replaces the MI-guided improvement strategy in NSGAII-MIIP with the random search-based improvement strategy in MOPSO-LS, i.e., randomly adding and eliminating a feature for each non-dominated solution. Other settings in NSGAII-MIIP-N and NSGAII-MIIP-R are the same as NSGAII-MIIP. The two variants are tested on the eight (original and synthetic) CMP datasets with the same experimental settings described in Section 5.3 of the manuscript. Moreover, the two variants are tested on the four UCI datasets with the same settings introduced in Section 5.1 of this article. Note that on InterAd and Mfeat, the mutation rate is set as 0.01 for NSGAII-MIIP and the two variants because the experimental results in Section 5.5 have shown that 0.01 is a more reasonable mutation rate on these two high-dimensional datasets. Similar to the above comparisons, the KPF/feature selection performance and the search performance of the three algorithms are compared. For each comparison metric, the mean and standard deviation values over the 30 experimental runs are recorded. The Wilcoxon signed-rank test is used to compare NSGAII-MIIP with the two variants, where $\upuparrows$ ($\uparrow$) or $\downdownarrows$ ($\downarrow$) denotes that NSGAII-MIIP obtains a significantly better or worse result at a significance level of 0.05 (0.1).

6.1. Comparison of FS performance of solutions selected by IPM between NSGAII-MIIP and the variants

Table 14 shows the FS performance results of the final solutions obtained by IPM of NSGAII-MMIP and the two variants. Overall, NSGAII-MIIP obtains higher mean values of GM, F1-score, and AUC in most cases. Specifically, on the datasets except for PAPER, InterAd, and UrbanLC, NSGAII-MIIP can obtain the best results on at least two performance metrics. The significance test denotes that NSGAII-MIIP obtains a significantly better GM, F1-score, or AUC value than at least one of the two variants on LATEX, SPIRA, PAPER-F, SPIRA-F, LSVT, InterAd, UrbanLC, and Mfeat. In one case, i.e., on InterAd, NSGAII-MIIP obtains significantly worse GM and AUC results than NSGAII-MIIP-N. According to the #KPFs (#SFs) results, NSGAII-MIIP generally selects significantly fewer features than NSGAII-MIIP-N and a similar number of features to NSGAII-MIIP. This denotes that the feature reduction performance of NSGAII-MIIP is much improved compared with NSGAII-MIIP-N. It should be note noted that NSGAII-MIIP selects 314.7 features on average on InterAd, which is much fewer than that of NSGAII-MIIP-N. From the FS perspective, NSGAII-MIIP outperforms NSGAII-MIIP-N on InterAd because NSGAII-MIIP substantially reduces the number of features with a slight decrease of classification performance.

Table 14: Comparison of FS performance between NSGAII-MIIP and the variants.

Dataset	Metric	NSGAII-MIIP	NSGAII-MIIP-N	NSGAII-MIIP-R
ADPN	GM (%)	76.48 ± 23.46	$\mathbf{77.18 \pm 14.08}$	77.16 ± 20.67
	F1-score (%)	$\mathbf{69.11 \pm 23.16}$	67.56 ± 16.08	68.56 ± 20.51
	AUC (%)	$\mathbf{83.44 \pm 26.21}$	81.56 ± 19.55	83.00 ± 18.54
	#KPFs	$\mathbf{2.4 \pm 0.5}$	$3.0 \pm 1.1 \upuparrows$	2.4 ± 0.6
LATEX	GM (%)	77.47 ± 10.46	$\mathbf{77.60 \pm 10.05}$	75.20 ± 11.91
	F1-score (%)	$\mathbf{68.75 \pm 13.17}$	68.69 ± 12.44	65.69 ± 14.21
	AUC (%)	$\mathbf{88.02 \pm 6.76}$	$85.33 \pm 6.51 \uparrow$	$85.34 \pm 5.95 \upuparrows$
	#KPFs	3.8 ± 0.9	$6.2 \pm 2.1 \upuparrows$	$\mathbf{3.8 \pm 0.9}$
PAPER	GM (%)	87.91 ± 7.14	$\mathbf{89.82 \pm 6.81}$	89.52 ± 7.21
	F1-score (%)	55.87 ± 12.47	57.34 ± 13.13	$\mathbf{57.57 \pm 13.27}$
	AUC (%)	92.67 ± 4.98	$\mathbf{92.75 \pm 5.52}$	92.34 ± 5.02
	#KPFs	3.1 ± 0.5	$\mathbf{2.9 \pm 0.5} \downdownarrows$	3.1 ± 0.7
SPIRA	GM (%)	$\mathbf{81.26 \pm 8.84}$	$71.20 \pm 15.57 \upuparrows$	$74.17 \pm 10.86 \upuparrows$
	F1-score (%)	$\mathbf{76.78 \pm 10.36}$	$64.76 \pm 15.73 \upuparrows$	$67.87 \pm 12.90 \upuparrows$
	AUC (%)	$\mathbf{86.79 \pm 9.51}$	$83.10 \pm 8.09 \upuparrows$	$81.87 \pm 10.08 \upuparrows$
	#KPFs	$\mathbf{3.1 \pm 0.8}$	$4.1 \pm 1.0 \upuparrows$	$3.5 \pm 0.9 \uparrow$
ADPN-F	GM (%)	$\mathbf{77.95 \pm 11.38}$	72.65 ± 23.03	72.93 ± 22.93
	F1-score (%)	$\mathbf{70.33 \pm 12.86}$	64.74 ± 22.96	65.22 ± 22.31
	AUC (%)	$\mathbf{85.67 \pm 16.84}$	82.44 ± 16.00	82.39 ± 18.26
	#KPFs	3.0 ± 0.7	$8.5 \pm 2.5 \upuparrows$	$\mathbf{2.4 \pm 0.6} \downdownarrows$
LATEX-F	GM (%)	$\mathbf{76.67 \pm 10.93}$	75.25 ± 10.20	75.51 ± 11.07
	F1-score (%)	$\mathbf{67.64 \pm 13.24}$	65.63 ± 12.64	65.80 ± 13.48
	AUC (%)	$\mathbf{86.29 \pm 5.52}$	84.73 ± 6.47	85.25 ± 7.16
	#KPFs	8.0 ± 4.1	$20.6 \pm 4.4 \upuparrows$	$\mathbf{6.8 \pm 3.4}$
PAPER-F	GM (%)	$\mathbf{88.65 \pm 7.70}$	$85.86 \pm 7.84 \uparrow$	87.93 ± 8.76
	F1-score (%)	$\mathbf{57.67 \pm 15.18}$	$53.61 \pm 11.22 \uparrow$	57.12 ± 12.34
	AUC (%)	92.38 ± 5.44	$\mathbf{92.79 \pm 4.94}$	91.16 ± 6.82
	#KPFs	3.5 ± 0.8	$\mathbf{3.4 \pm 0.9}$	3.6 ± 0.8
SPIRA-F	GM (%)	$\mathbf{78.71 \pm 7.42}$	74.76 ± 12.88	$71.65 \pm 13.06 \upuparrows$
	F1-score (%)	$\mathbf{73.64 \pm 8.76}$	$68.07 \pm 16.19 \uparrow$	$64.70 \pm 16.09 \upuparrows$
	AUC (%)	$\mathbf{85.30 \pm 10.57}$	82.64 ± 10.66	$80.33 \pm 12.05 \upuparrows$
	#KPFs	$\mathbf{3.9 \pm 1.2}$	$10.0 \pm 2.4 \upuparrows$	4.1 ± 1.2
LSVT	GM (%)	$\mathbf{82.21 \pm 6.02}$	$73.76 \pm 9.40 \upuparrows$	79.75 ± 6.81
	F1-score (%)	$\mathbf{76.06 \pm 7.59}$	$64.80 \pm 11.07 \upuparrows$	$72.44 \pm 8.79 \uparrow$
	AUC (%)	$\mathbf{88.16 \pm 4.52}$	$81.86 \pm 7.29 \upuparrows$	$86.40 \pm 4.98 \upuparrows$
	#SFs	$\mathbf{2.5 \pm 0.5}$	$6.4 \pm 2.4 \upuparrows$	2.5 ± 0.6
InterAd	GM (%)	93.14 ± 0.45	$\mathbf{93.40 \pm 0.48} \downarrow$	93.01 ± 0.61
	F1-score (%)	89.50 ± 1.07	$\mathbf{89.74 \pm 0.92}$	89.22 ± 1.05
	AUC (%)	97.37 ± 0.52	$\mathbf{97.65 \pm 0.58} \downarrow$	$97.11 \pm 0.67 \uparrow$
	#SFs	$\mathbf{314.7 \pm 24.5}$	$363.6 \pm 24.6 \upuparrows$	322.2 ± 23.9
UrbanLC	GM (%)	84.14 ± 2.30	$\mathbf{84.50 \pm 2.83}$	$82.38 \pm 2.61 \upuparrows$
	F1-score (%)	85.00 ± 2.16	$\mathbf{85.18 \pm 2.48}$	$83.40 \pm 2.24 \upuparrows$
	AUC (%)	$\mathbf{98.64 \pm 0.27}$	98.49 ± 0.40	$98.23 \pm 0.43 \upuparrows$
	#SFs	5.1 ± 1.1	$11.5 \pm 3.0 \upuparrows$	$\mathbf{4.8 \pm 1.1}$
Mfeat	GM (%)	$\mathbf{96.85 \pm 0.50}$	96.83 ± 0.32	96.64 ± 0.55
	F1-score (%)	$\mathbf{96.89 \pm 0.49}$	96.87 ± 0.31	96.68 ± 0.54
	AUC (%)	$\mathbf{99.78 \pm 0.12}$	$99.69 \pm 0.11 \upuparrows$	$99.69 \pm 0.13 \upuparrows$
	#SFs	$\mathbf{59.7 \pm 12.2}$	$85.3 \pm 10.9 \upuparrows$	60.9 ± 14.4

Overall, the above results demonstrate that NSGAII-MIIP outperforms the two variants in FS. Compared with NSGAII-MIIP-N, NSGAII-MIIP obtains substantially fewer features on high-dimensional datasets. This indicates that the improvement phase in NSGAII-MIIP effectively improves the feature reduction performance. NSGAII-MIIP-R can generally obtain a similar number of features to NSGAII-MIIP. However, the classification performance of NSGAII-MIIP-R is worse than NSGAII-MIIP in most cases. This shows that NSGAII-MIIP can select real key features more precisely than NSGAII-MIIP-R.

Table 15: Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP and the variants using the AUC measure.

Metric	Dataset	NSGAII-MIIP	NSGAII-MIIP-N	NSGAII-MIIP-R
HV	ADPN	$\mathbf{1.194 \pm 0.008}$	$1.175 \pm 0.026 \upuparrows$	1.191 ± 0.013
	LATEX	1.103 ± 0.049	1.082 ± 0.072	$\mathbf{1.126 \pm 0.051} \downdownarrows$
	PAPER	$\mathbf{1.147 \pm 0.062}$	$1.117 \pm 0.086 \upuparrows$	$1.109 \pm 0.110 \upuparrows$
	SPIRA	$\mathbf{1.164 \pm 0.053}$	$1.124 \pm 0.065 \upuparrows$	$1.139 \pm 0.084 \upuparrows$
	ADPN-F	$\mathbf{1.199 \pm 0.009}$	$1.104 \pm 0.041 \upuparrows$	1.197 ± 0.012
	LATEX-F	1.110 ± 0.050	$0.924 \pm 0.064 \upuparrows$	$\mathbf{1.139 \pm 0.045} \downdownarrows$
	PAPER-F	$\mathbf{1.123 \pm 0.045}$	$1.094 \pm 0.072 \upuparrows$	$1.090 \pm 0.079 \upuparrows$
	SPIRA-F	$\mathbf{1.176 \pm 0.047}$	$1.057 \pm 0.067 \upuparrows$	1.165 ± 0.084
	LSVT	$\mathbf{1.060 \pm 0.019}$	$0.670 \pm 0.227 \upuparrows$	$1.006 \pm 0.101 \upuparrows$
	InterAd	$\mathbf{0.744 \pm 0.136}$	$0.624 \pm 0.094 \upuparrows$	0.686 ± 0.140
	UrbanLC	$\mathbf{1.122 \pm 0.013}$	$0.880 \pm 0.104 \upuparrows$	$1.108 \pm 0.024 \upuparrows$
	Mfeat	$\mathbf{0.880 \pm 0.170}$	$0.630 \pm 0.121 \upuparrows$	$0.738 \pm 0.204 \upuparrows$
IGD	ADPN	$\mathbf{0.023 \pm 0.040}$	$0.091 \pm 0.091 \upuparrows$	$0.077 \pm 0.097 \upuparrows$
	LATEX	0.071 ± 0.038	$0.139 \pm 0.068 \upuparrows$	$\mathbf{0.054 \pm 0.033} \downdownarrows$
	PAPER	$\mathbf{0.000 \pm 0.000}$	$0.047 \pm 0.086 \upuparrows$	$0.062 \pm 0.105 \upuparrows$
	SPIRA	$\mathbf{0.049 \pm 0.044}$	$0.085 \pm 0.055 \upuparrows$	$0.076 \pm 0.069 \uparrow$
	ADPN-F	$\mathbf{0.020 \pm 0.036}$	$0.187 \pm 0.083 \upuparrows$	$0.087 \pm 0.101 \upuparrows$
	LATEX-F	0.088 ± 0.042	$0.211 \pm 0.045 \upuparrows$	$\mathbf{0.063 \pm 0.040} \downdownarrows$
	PAPER-F	$\mathbf{0.038 \pm 0.050}$	$0.066 \pm 0.072 \uparrow$	$0.082 \pm 0.083 \upuparrows$
	SPIRA-F	$\mathbf{0.054 \pm 0.043}$	$0.167 \pm 0.083 \upuparrows$	0.059 ± 0.060
	LSVT	$\mathbf{0.125 \pm 0.014}$	$0.361 \pm 0.178 \upuparrows$	0.151 ± 0.071
	InterAd	$\mathbf{0.318 \pm 0.090}$	0.324 ± 0.086	0.346 ± 0.120
	UrbanLC	0.156 ± 0.036	$0.271 \pm 0.063 \upuparrows$	$\mathbf{0.133 \pm 0.045} \downarrow$
	Mfeat	$\mathbf{0.194 \pm 0.133}$	$0.368 \pm 0.108 \upuparrows$	$0.310 \pm 0.176 \upuparrows$
SC	ADPN	$\mathbf{0.683 \pm 0.445}$	$0.433 \pm 0.388 \upuparrows$	0.600 ± 0.403
	LATEX	0.200 ± 0.223	0.156 ± 0.189	$\mathbf{0.346 \pm 0.264} \downdownarrows$
	PAPER	$\mathbf{1.000 \pm 0.000}$	$0.733 \pm 0.430 \upuparrows$	$0.717 \pm 0.429 \upuparrows$
	SPIRA	$\mathbf{0.533 \pm 0.380}$	$0.244 \pm 0.324 \upuparrows$	0.382 ± 0.298
	ADPN-F	$\mathbf{0.633 \pm 0.414}$	$0.000 \pm 0.000 \upuparrows$	0.567 ± 0.365
	LATEX-F	0.102 ± 0.161	$0.007 \pm 0.037 \upuparrows$	$\mathbf{0.316 \pm 0.301} \downdownarrows$
	PAPER-F	$\mathbf{0.717 \pm 0.322}$	$0.439 \pm 0.454 \upuparrows$	$0.450 \pm 0.436 \upuparrows$
	SPIRA-F	0.290 ± 0.369	$0.017 \pm 0.091 \upuparrows$	$\mathbf{0.400 \pm 0.418}$
	LSVT	0.000 ± 0.000	0.025 ± 0.137	$\mathbf{0.067 \pm 0.112} \downdownarrows$
	InterAd	$\mathbf{0.021 \pm 0.064}$	0.010 ± 0.033	0.003 ± 0.014
	UrbanLC	$\mathbf{0.067 \pm 0.082}$	$0.000 \pm 0.000 \upuparrows$	0.033 ± 0.061
	Mfeat	$\mathbf{0.033 \pm 0.097}$	0.000 ± 0.000	0.000 ± 0.000

6.2. Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP and the variants

Table 15 compares the overall FS performance of the non-dominated sets obtained by NSGAII-MIIP and the two variants using the AUC measure. Overall, NSGAII-MIIP obtains the best mean HV, IGD, and SC results on most datasets. Specifically, compared with NSGAII-MIIP-N, the significance test results indicate that NSGAII-MIIP obtains significantly better HV and IGD values on 11 of 12 datasets, and obtains better SC values on 8 of 12 datasets. Compared with NSGAII-MIIP-R, NSGAII-MIIP obtains significantly better or similar FS performance in most cases. It obtains significantly better HV, IGD, or SC results on 8 of 12 datasets (except for LATEX, LATEX-F, SPIRA-F, and InterAd). On a few datasets, i.e., LATEX, LATEX-F, LSVT, and UrbanLC, NSGAII-MIIP obtains a significantly worse HV, IGD, or SC value than NSGAII-MIIP-R.

Table 16 compares the overall FS performance between NSGAII-MIIP and the two variants on the GM measure. Overall, NSGAII-MIIP obtains the best mean HV, IGD, and SC results on most datasets.

Table 16: Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP and the variants using the GM measure.

Metric	Dataset	NSGAII-MIIP	NSGAII-MIIP-N	NSGAII-MIIP-R
HV	ADPN	0.988 ± 0.270	1.017 ± 0.214	$\mathbf{1.073 \pm 0.178}$
	LATEX	1.007 ± 0.127	$0.941 \pm 0.129 \upuparrows$	$\mathbf{1.022 \pm 0.137}$
	PAPER	1.068 ± 0.093	$\mathbf{1.077 \pm 0.070}$	1.059 ± 0.112
	SPIRA	$\mathbf{0.973 \pm 0.144}$	0.955 ± 0.132	$0.925 \pm 0.142 \uparrow$
	ADPN-F	$\mathbf{0.996 \pm 0.304}$	$0.784 \pm 0.412 \upuparrows$	0.963 ± 0.311
	LATEX-F	$\mathbf{1.043 \pm 0.111}$	$0.808 \pm 0.120 \upuparrows$	1.027 ± 0.104
	PAPER-F	$\mathbf{1.115 \pm 0.058}$	$1.048 \pm 0.162 \upuparrows$	1.067 ± 0.146
	SPIRA-F	$\mathbf{1.011 \pm 0.119}$	$0.883 \pm 0.161 \upuparrows$	$0.950 \pm 0.186 \uparrow$
	LSVT	$\mathbf{1.004 \pm 0.064}$	$0.631 \pm 0.229 \upuparrows$	$0.964 \pm 0.098 \uparrow$
	InterAd	$\mathbf{0.895 \pm 0.098}$	$0.753 \pm 0.081 \upuparrows$	0.862 ± 0.071
	UrbanLC	$\mathbf{1.102 \pm 0.024}$	$0.909 \pm 0.083 \upuparrows$	$1.092 \pm 0.030 \upuparrows$
	Mfeat	$\mathbf{0.791 \pm 0.169}$	$0.630 \pm 0.094 \upuparrows$	0.736 ± 0.129
IGD	ADPN	0.106 ± 0.175	0.126 ± 0.150	$\mathbf{0.086 \pm 0.131}$
	LATEX	0.078 ± 0.060	$0.141 \pm 0.087 \upuparrows$	$\mathbf{0.067 \pm 0.049}$
	PAPER	0.093 ± 0.077	$\mathbf{0.085 \pm 0.073}$	0.106 ± 0.095
	SPIRA	$\mathbf{0.087 \pm 0.057}$	0.113 ± 0.082	$0.120 \pm 0.091 \upuparrows$
	ADPN-F	$\mathbf{0.140 \pm 0.224}$	$0.327 \pm 0.368 \upuparrows$	0.171 ± 0.263
	LATEX-F	0.096 ± 0.057	$0.249 \pm 0.065 \upuparrows$	$\mathbf{0.096 \pm 0.055}$
	PAPER-F	$\mathbf{0.057 \pm 0.040}$	0.105 ± 0.101	0.094 ± 0.102
	SPIRA-F	$\mathbf{0.076 \pm 0.040}$	$0.200 \pm 0.072 \upuparrows$	$0.109 \pm 0.078 \uparrow$
	LSVT	$\mathbf{0.142 \pm 0.037}$	$0.360 \pm 0.176 \upuparrows$	0.162 ± 0.063
	InterAd	0.230 ± 0.040	$\mathbf{0.224 \pm 0.043}$	0.238 ± 0.064
	UrbanLC	0.097 ± 0.031	$0.156 \pm 0.044 \upuparrows$	$\mathbf{0.088 \pm 0.023}$
	Mfeat	$\mathbf{0.304 \pm 0.130}$	$0.444 \pm 0.087 \upuparrows$	$0.357 \pm 0.111 \uparrow$
SC	ADPN	$\mathbf{0.628 \pm 0.398}$	$0.372 \pm 0.410 \upuparrows$	$0.450 \pm 0.427 \uparrow$
	LATEX	$\mathbf{0.401 \pm 0.264}$	$0.111 \pm 0.188 \upuparrows$	0.373 ± 0.211
	PAPER	0.442 ± 0.264	$\mathbf{0.454 \pm 0.270}$	0.405 ± 0.245
	SPIRA	$\mathbf{0.477 \pm 0.358}$	0.336 ± 0.310	$0.364 \pm 0.253 \uparrow$
	ADPN-F	$\mathbf{0.483 \pm 0.466}$	$0.033 \pm 0.127 \upuparrows$	0.400 ± 0.437
	LATEX-F	$\mathbf{0.257 \pm 0.231}$	$0.000 \pm 0.000 \upuparrows$	0.152 ± 0.224
	PAPER-F	$\mathbf{0.397 \pm 0.264}$	0.273 ± 0.224	0.310 ± 0.204
	SPIRA-F	$\mathbf{0.434 \pm 0.214}$	$0.045 \pm 0.125 \upuparrows$	$0.329 \pm 0.223 \uparrow$
	LSVT	0.060 ± 0.093	$0.013 \pm 0.073 \uparrow$	$\mathbf{0.093 \pm 0.126}$
	InterAd	$\mathbf{0.020 \pm 0.081}$	0.013 ± 0.051	0.000 ± 0.000
	UrbanLC	0.014 ± 0.044	0.005 ± 0.026	$\mathbf{0.024 \pm 0.066}$
	Mfeat	$\mathbf{0.033 \pm 0.183}$	0.000 ± 0.000	0.000 ± 0.000

Compared with NSGAII-MIIP-N, the significance test indicates that NSGAII-MIIP obtains significantly better HV results on 9 of 12 datasets, obtains significantly better IGD results on 7 datasets, and obtains significantly better SC results on 6 datasets. No evidence indicates NSGAII-MIIP obtains significantly worse HV, IGD, or SC results than NSGAII-MIIP-N. Compared with NSGAII-MIIP-R, NSGAII-MIIP obtains a significantly better HV, IGD, or SC value on 6 of 12 datasets. No evidence indicates that NSGAII-MIIP obtains significantly worse results than NSGAII-MIIP-R.

Table 17 compares the overall FS performance between NSGAII-MIIP and the two variants on the F1-score measure. Overall, the comparison results on the F1-score are similar to that on AUC and GM. Compared with NSGAII-MIIP-N, NSGAII-MIIP obtains significantly better HV, IGD, and SC results in most cases, indicating that the overall FS performance of non-dominated solutions found by NSGAII-MIIP is better than NSGAII-MIIP-N. Compared with NSAGAII-MIIP-R, NSGAII-MIIP obtains better mean HV, IGD, and SC results in most cases. On 7 of 12 datasets, NSGAII-MIIP obtains a significantly better HV, IGD, or SC value than NSGAII-MIIP-R. NSGAII-MIIP only obtains a significantly worse IGD value than

Table 17: Comparison of overall FS performance of obtained non-dominated sets between NSGAII-MIIP and the variants using the F1-score measure.

Metric	Dataset	NSGAII-MIIP	NSGAII-MIIP-N	NSGAII-MIIP-R
HV	ADPN	0.959 ± 0.282	0.962 ± 0.265	$\mathbf{1.026 \pm 0.233}$
	LATEX	0.986 ± 0.138	$0.905 \pm 0.138 \upuparrows$	$\mathbf{0.991 \pm 0.146}$
	PAPER	0.892 ± 0.122	0.891 ± 0.127	$\mathbf{0.893 \pm 0.136}$
	SPIRA	$\mathbf{0.972 \pm 0.131}$	0.951 ± 0.125	$0.912 \pm 0.141 \upuparrows$
	ADPN-F	$\mathbf{0.968 \pm 0.305}$	$0.742 \pm 0.397 \upuparrows$	0.916 ± 0.308
	LATEX-F	$\mathbf{1.036 \pm 0.108}$	$0.793 \pm 0.120 \upuparrows$	1.012 ± 0.105
	PAPER-F	$\mathbf{0.910 \pm 0.145}$	0.872 ± 0.165	0.880 ± 0.153
	SPIRA-F	$\mathbf{1.011 \pm 0.109}$	$0.875 \pm 0.154 \upuparrows$	$0.947 \pm 0.181 \uparrow$
	LSVT	$\mathbf{0.928 \pm 0.063}$	$0.523 \pm 0.243 \upuparrows$	$0.871 \pm 0.132 \uparrow$
	InterAd	$\mathbf{0.815 \pm 0.100}$	$0.672 \pm 0.071 \upuparrows$	$0.771 \pm 0.072 \uparrow$
	UrbanLC	$\mathbf{1.067 \pm 0.022}$	$0.874 \pm 0.081 \upuparrows$	$1.058 \pm 0.028 \upuparrows$
	Mfeat	$\mathbf{0.790 \pm 0.169}$	$0.629 \pm 0.094 \upuparrows$	0.735 ± 0.129
IGD	ADPN	0.089 ± 0.156	$0.132 \pm 0.168 \uparrow$	$\mathbf{0.082 \pm 0.146}$
	LATEX	0.080 ± 0.058	$0.146 \pm 0.090 \upuparrows$	$\mathbf{0.071 \pm 0.048}$
	PAPER	0.118 ± 0.084	$\mathbf{0.104 \pm 0.067}$	0.119 ± 0.078
	SPIRA	$\mathbf{0.086 \pm 0.064}$	$0.115 \pm 0.083 \uparrow$	$0.121 \pm 0.087 \upuparrows$
	ADPN-F	$\mathbf{0.131 \pm 0.203}$	$0.342 \pm 0.350 \upuparrows$	0.186 ± 0.261
	LATEX-F	$\mathbf{0.104 \pm 0.055}$	$0.256 \pm 0.063 \upuparrows$	0.106 ± 0.059
	PAPER-F	$\mathbf{0.092 \pm 0.049}$	0.115 ± 0.057	0.108 ± 0.062
	SPIRA-F	$\mathbf{0.072 \pm 0.040}$	$0.200 \pm 0.072 \upuparrows$	$0.106 \pm 0.075 \uparrow$
	LSVT	$\mathbf{0.155 \pm 0.037}$	$0.428 \pm 0.196 \upuparrows$	$0.192 \pm 0.081 \upuparrows$
	InterAd	$\mathbf{0.288 \pm 0.058}$	$0.339 \pm 0.046 \upuparrows$	0.311 ± 0.053
	UrbanLC	0.169 ± 0.042	$0.226 \pm 0.045 \upuparrows$	$\mathbf{0.152 \pm 0.030} \downarrow$
	Mfeat	$\mathbf{0.304 \pm 0.130}$	$0.445 \pm 0.087 \upuparrows$	$0.358 \pm 0.111 \uparrow$
SC	ADPN	$\mathbf{0.661 \pm 0.365}$	$0.383 \pm 0.413 \upuparrows$	$0.461 \pm 0.419 \upuparrows$
	LATEX	$\mathbf{0.407 \pm 0.254}$	$0.091 \pm 0.175 \upuparrows$	0.384 ± 0.236
	PAPER	0.410 ± 0.277	$\mathbf{0.432 \pm 0.239}$	0.390 ± 0.248
	SPIRA	$\mathbf{0.502 \pm 0.385}$	$0.342 \pm 0.321 \uparrow$	$0.373 \pm 0.257 \uparrow$
	ADPN-F	$\mathbf{0.517 \pm 0.441}$	$0.033 \pm 0.127 \upuparrows$	0.411 ± 0.430
	LATEX-F	$\mathbf{0.211 \pm 0.240}$	$0.011 \pm 0.061 \upuparrows$	0.167 ± 0.251
	PAPER-F	$\mathbf{0.369 \pm 0.228}$	$0.257 \pm 0.205 \uparrow$	0.309 ± 0.204
	SPIRA-F	$\mathbf{0.429 \pm 0.225}$	$0.042 \pm 0.115 \upuparrows$	0.344 ± 0.222
	LSVT	0.093 ± 0.101	$0.000 \pm 0.000 \upuparrows$	$\mathbf{0.113 \pm 0.125}$
	InterAd	$\mathbf{0.033 \pm 0.134}$	0.000 ± 0.000	0.000 ± 0.000
	UrbanLC	0.015 ± 0.038	0.007 ± 0.041	$\mathbf{0.030 \pm 0.065}$
	Mfeat	$\mathbf{0.033 \pm 0.183}$	0.000 ± 0.000	0.000 ± 0.000

NSGAII-MII-R on UrbanLC.

In summary, the above comparison results of the overall FS performance justify the effectiveness of the MI-guided improvement strategy in NSGAII-MIIP. Compared with NSGAII-MIIP-N with no improvement strategy, NSGAII-MIIP shows substantially better FS performance. Compared with NSGAII-MIIP-R with the random improvement strategy, NSGAII-MIIP shows better FS performance in most cases. Thus, we can conclude that the MI-guided improvement strategy is more effective than the random improvement strategy.

6.3. Comparison of the search performance between NSGAII-MIIP and the variants

Table 18 compares the search performance between NSGAII-MIIP and the two variants. NSGAII-MIIP obtains better mean HV, IGD, and SC results than the two variants in most cases. Compared with NSGAII-MIIP-N, NSGAII-MIIP obtains significantly better results in most cases according to the significance test results. This indicates that NSGAII-MIIP has significantly better search performance in dealing with the defined FS model than NSGAII-MIIP-N. Compared with NSGAII-MIIP-R, NSGAII-MIIP

Table 18: Comparison of the search performance between NSGAII-MIIP and the variants.

Metric	Dataset	NSGAII-MIIP	NSGAII-MIIP-N	NSGAII-MIIP-R
HV	ADPN	1.161 ± 0.025	$1.145 \pm 0.035 \Uparrow$	$\mathbf{1.165 \pm 0.022}$
	LATEX	$\mathbf{1.152 \pm 0.032}$	$1.117 \pm 0.039 \Uparrow$	1.144 ± 0.033
	PAPER	$\mathbf{1.132 \pm 0.020}$	$1.128 \pm 0.022 \Uparrow$	1.130 ± 0.024
	SPIRA	$\mathbf{1.140 \pm 0.026}$	$1.126 \pm 0.041 \Uparrow$	1.140 ± 0.033
	ADPN-F	$\mathbf{1.160 \pm 0.023}$	$1.093 \pm 0.038 \Uparrow$	1.156 ± 0.028
	LATEX-F	1.140 ± 0.037	$0.960 \pm 0.048 \Uparrow$	$\mathbf{1.142 \pm 0.034}$
	PAPER-F	$\mathbf{1.137 \pm 0.021}$	1.133 ± 0.026	1.128 ± 0.031
	SPIRA-F	$\mathbf{1.118 \pm 0.034}$	$1.037 \pm 0.057 \Uparrow$	1.112 ± 0.050
	LSVT	$\mathbf{1.125 \pm 0.018}$	$0.745 \pm 0.160 \Uparrow$	$1.098 \pm 0.034 \Uparrow$
	InterAd	$\mathbf{0.947 \pm 0.088}$	$0.800 \pm 0.080 \Uparrow$	0.929 ± 0.084
	UrbanLC	1.117 ± 0.010	$0.924 \pm 0.081 \Uparrow$	$\mathbf{1.118 \pm 0.009}$
	Mfeat	$\mathbf{0.948 \pm 0.108}$	$0.763 \pm 0.081 \Uparrow$	0.923 ± 0.113
IGD	ADPN	$\mathbf{0.020 \pm 0.012}$	$0.049 \pm 0.025 \Uparrow$	$0.033 \pm 0.025 \uparrow$
	LATEX	$\mathbf{0.019 \pm 0.013}$	$0.048 \pm 0.022 \Uparrow$	0.024 ± 0.015
	PAPER	$\mathbf{0.026 \pm 0.018}$	$0.032 \pm 0.022 \uparrow$	0.027 ± 0.021
	SPIRA	$\mathbf{0.027 \pm 0.014}$	$0.040 \pm 0.020 \Uparrow$	0.030 ± 0.019
	ADPN-F	$\mathbf{0.029 \pm 0.017}$	$0.099 \pm 0.035 \Uparrow$	$0.041 \pm 0.022 \Uparrow$
	LATEX-F	0.040 ± 0.027	$0.145 \pm 0.038 \Uparrow$	$\mathbf{0.039 \pm 0.019}$
	PAPER-F	$\mathbf{0.029 \pm 0.016}$	0.036 ± 0.022	0.036 ± 0.022
	SPIRA-F	$\mathbf{0.030 \pm 0.013}$	$0.101 \pm 0.037 \Uparrow$	0.034 ± 0.023
	LSVT	$\mathbf{0.035 \pm 0.011}$	$0.262 \pm 0.107 \Uparrow$	$0.060 \pm 0.024 \Uparrow$
	InterAd	$\mathbf{0.145 \pm 0.052}$	$0.255 \pm 0.065 \Uparrow$	0.156 ± 0.055
	UrbanLC	0.075 ± 0.012	$0.146 \pm 0.039 \Uparrow$	$\mathbf{0.074 \pm 0.016}$
	Mfeat	$\mathbf{0.151 \pm 0.057}$	$0.281 \pm 0.061 \Uparrow$	0.173 ± 0.069
SC	ADPN	$\mathbf{0.400 \pm 0.229}$	$0.222 \pm 0.244 \Uparrow$	0.381 ± 0.165
	LATEX	$\mathbf{0.415 \pm 0.168}$	$0.076 \pm 0.116 \Uparrow$	$0.314 \pm 0.173 \Uparrow$
	PAPER	0.395 ± 0.204	$0.310 \pm 0.164 \Uparrow$	$\mathbf{0.396 \pm 0.177}$
	SPIRA	$\mathbf{0.388 \pm 0.143}$	$0.195 \pm 0.165 \Uparrow$	$0.294 \pm 0.165 \Uparrow$
	ADPN-F	0.274 ± 0.172	$0.018 \pm 0.055 \Uparrow$	$\mathbf{0.290 \pm 0.204}$
	LATEX-F	0.179 ± 0.265	$0.000 \pm 0.000 \Uparrow$	$\mathbf{0.187 \pm 0.225}$
	PAPER-F	$\mathbf{0.261 \pm 0.168}$	0.233 ± 0.222	0.253 ± 0.201
	SPIRA-F	$\mathbf{0.316 \pm 0.178}$	$0.042 \pm 0.094 \Uparrow$	0.247 ± 0.245
	LSVT	$\mathbf{0.328 \pm 0.135}$	$0.000 \pm 0.000 \Uparrow$	$0.150 \pm 0.160 \Uparrow$
	InterAd	0.013 ± 0.073	0.000 ± 0.000	$\mathbf{0.020 \pm 0.081}$
	UrbanLC	0.018 ± 0.035	$0.000 \pm 0.000 \Uparrow$	$\mathbf{0.029 \pm 0.049}$
	Mfeat	$\mathbf{0.028 \pm 0.137}$	0.003 ± 0.018	0.002 ± 0.009

obtains a significantly better HV, IGD, or SC value on 5 of 12 datasets. No results from the significance test indicate that NSGAII-MIIP has worse search performance than NSGAII-MIIP-R.

Overall, the results in Table 18 demonstrate that NSGAII-MIIP has better search performance than NSGAII-MIIP-N and NSGAII-MIIP-R. This explains why NSGAII-MIIP obtains better FS performance than the two variants. Moreover, it is worth noting that the search performance results in Table 18 are slightly inconsistent with the FS performance results in Section 6.2. We take the FS performance results with the AUC measure as an example. NSGAII-MIIP obtains significantly better HV results than NSGAII-MIIP-R on 6 datasets according to Table 15. In comparison, it obtains a significantly better HV result than NSGAII-MIIP-R on 1 dataset according to Table 18. This reveals that the features selected by NSGAII-MIIP have a better generalization ability than those of NSGAII-MIIP-R as NSGAII-MIIP shows much better prediction performances on the test sets. This demonstrates that by considering feature relevance and redundancy in the MI-based feature importance measure, the MI-guided improvement strategy can substantially improve the FS performance.

References

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197.

He, Z., Hu, H., Zhang, M., Zhang, Y., & Li, A.-D. (2022). A decomposition-based multi-objective particle swarm optimization algorithm with a local search strategy for key quality characteristic identification in production processes. *Computers & Industrial Engineering*, *172*, 108617.

Kelly, M., Longjohn, R., & Nottingham, K. (2024). UCI Machine Learning Repository. `https://archive.ics.uci.edu/`.

Li, A.-D., He, Z., & Zhang, Y. (2016). Bi-objective variable selection for key quality characteristics selection based on a modified NSGA-II and the ideal point method. *Computers in Industry*, *82*, 95 – 103.

Li, A.-D., Xue, B., & Zhang, M. (2020). Multi-objective feature selection using hybridization of a genetic algorithm and direct multisearch for key quality characteristic selection. *Information Sciences*, *523*, 245 – 265.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Tian, Y., Cheng, R., Zhang, X., Su, Y., & Jin, Y. (2019). A strengthened dominance relation considering convergence and diversity for evolutionary many-objective optimization. *IEEE Transactions on Evolutionary Computation*, *23*(2), 331–345.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80 – 83.

Zhang, Q., & Li, H. (2007). MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, *11*(6), 712–731.

Zhang, Y., Gong, D., Gao, X., Tian, T., & Sun, X. (2020). Binary differential evolution with self-learning for multi-objective feature selection. *Information Sciences*, *507*, 67–85.

Zitzler, E., Laumanns, M., & Thiele, L. (2001). SPEA2: Improving the strength pareto evolutionary algorithm. In *Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems. Proceedings of the EUROGEN'2001. Athens. Greece, September 19-21* (pp. 95–100).