Response to Reviewer 1 Comments

**Point 1:** The results of the experiments should be added in brief in the abstract.

**Response 1:** Thanks for the comments. We now have added the results of the experiments in abstract: “Analysis of the generalization ability of BetaDCE revealed that it was a trade-off between bias and variance. The robustness of BetaDCE is demonstrated by experiments on two types of data. For XOR-like dataset, the false discovery rate of BetaDCE was significantly smaller than that of other methods. For leukemia dataset, the AUC of BetaDCE was 0.93 with only four selected features” (line 19-23)

**Point 2:** In line 20-21, “it also more accurately predicted the class labels with smaller number of features”. Please state by what percent is more accurately predicted.

**Response 2:** Thanks for the suggestion. We now have revised it as “For leukemia dataset, the AUC of BetaDCE on test set was 0.93 with only four selected features, which indicated that BetaDCE not only detected the irrelevant and redundant features precisely, it also more accurately predicted the class labels with smaller number of features than the original method whose AUC was 0.83 with 50 features.” (line 22-25)

**Point 3:** The work is based on past research/results. Is there a more recent attempt? If available, referring to more recent work will make the paper more convincing that it addresses a problem that has not yet been addressed. The paper can benefit by the addition of a “related work” section.

**Response 3:** Thanks for the comments. We now have added a section of the recent related work: “Many of recent studies on feature selection focused on the methods of optimization for feature subsets. New searching methods have been applied to select the optimal feature subset. For example, whale optimization-related approaches were used for wrapper feature selection [28]. Other new approaches include evolutionary population dynamics and grasshopper optimization [29], Binary dragonfly optimization [30], artificial bee colony optimization [31], etc. Some studies used the combined methods of optimization for feature selection. For example, Mafarja et al combined the whale optimization with simulated annealing [32]. Faris et al combined binary salp swarm algorithm with crossover scheme for feature selection [33]. However, few studies have been conducted on evaluation function, which is another important element for feature selection.” (line 61-69). The references have been updated (line 376-390).

**Point 4:** Consider replacing line 126 with a reference.

**Response 4:** Thanks for the suggestion. We now have replaced line 126 with a reference. (line 394).

**Point 5:** In figure 7, the labels of the red points are difficult to read. Consider making them bold.
Response 5: Thanks for the suggestion. We now have made the labels bold and enlarged the size. (line 281)

Point 6: Consider adding a conclusions and limitations section.

Response 6: Thanks for the suggestion. We have added the limitations section, with a description of future work. (line 300-305). The conclusions section has also been added. (line 311-315).