Supplementary Information
Classification of Two and Multi-Class Cancer Data Reliably Using Multi-Objective Evolutionary Algorithms

Kalyanmoy Deb and A. Raji Reddy

Kanpur Genetic Algorithms Laboratory (KanGAL)
Indian Institute of Technology Kanpur
Kanpur, PIN 208 016, India
{deb,arreddy}@iitk.ac.in
http://www.iitk.ac.in/kangal

1 50-Gene Leukemia Dataset

1.1 Minimization of Gene Subset Size with $\theta = 0$

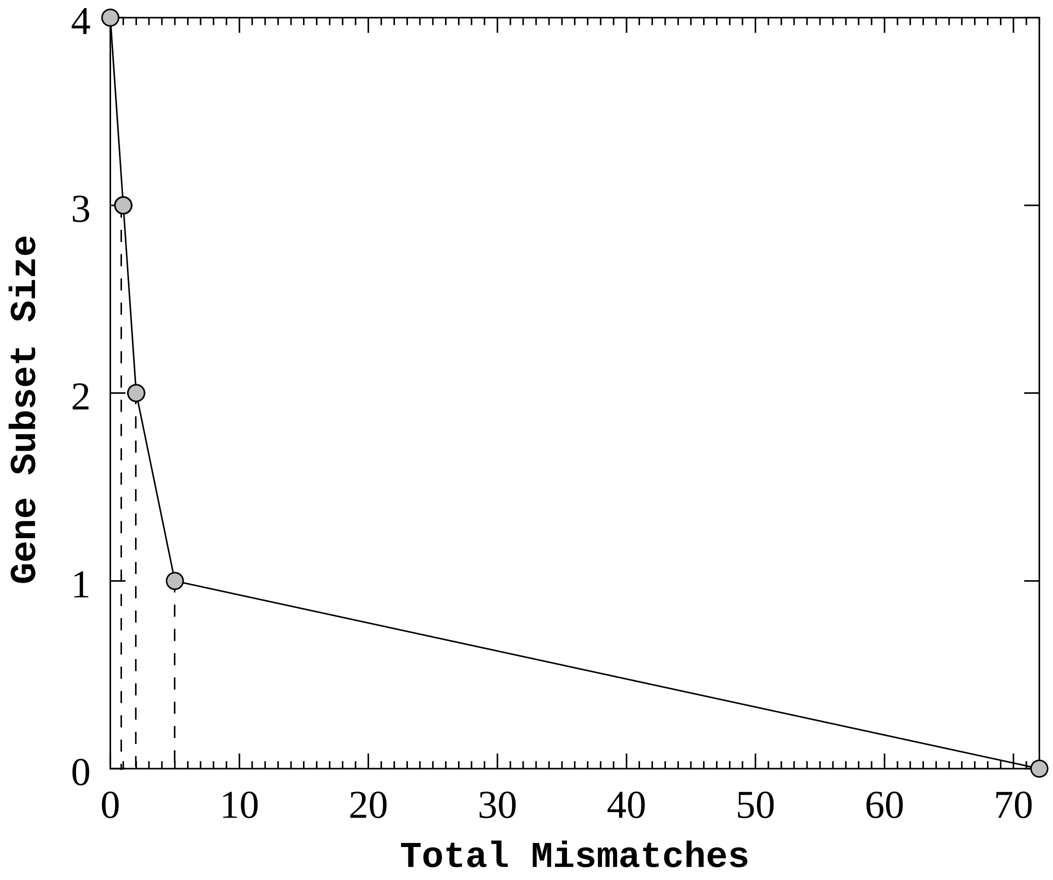

Fig. 1. Two-objective solutions obtained using NSGA-II for the Leukemia samples. Here, we have considered only two objectives (one is the minimization of gene subset size (f_1) and the another one is the minimization of total number of mismatches in both training and test samples ($f_2 + f_3$).

1.2 Minimization of Gene Subset Size with $\theta = 0$

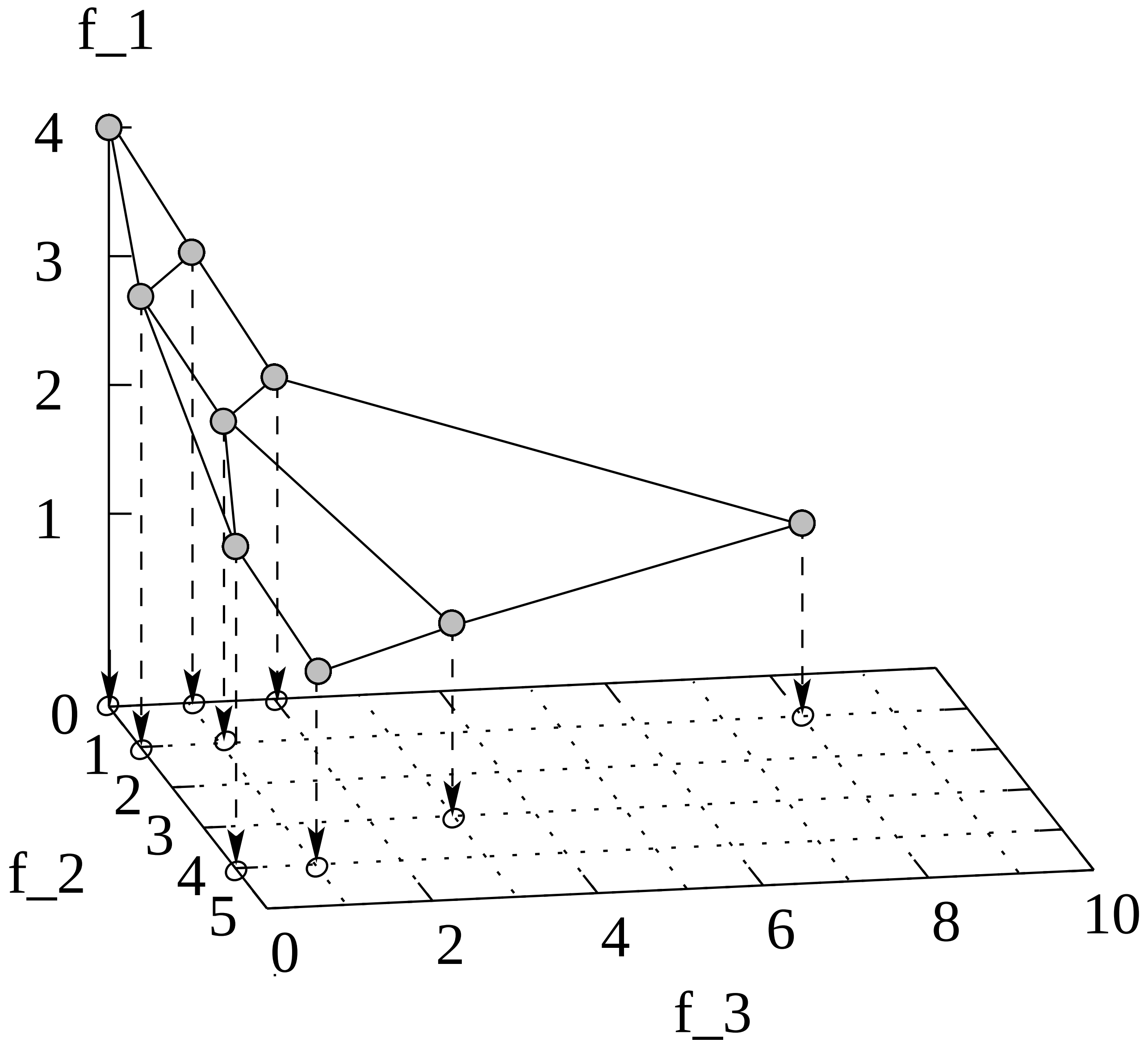

Fig. 2. Three-objective solutions obtained using NSGA-II for the Leukemia samples. Here, we have considered three objectives: (i) minimization of gene subset size (f_1), (ii) minimization of number of mismatches in training samples (f_2), and (iii) minimization of number of mismatches in test samples (f_3).

1.3 Maximization of Gene Subset Size with $\theta = 0$

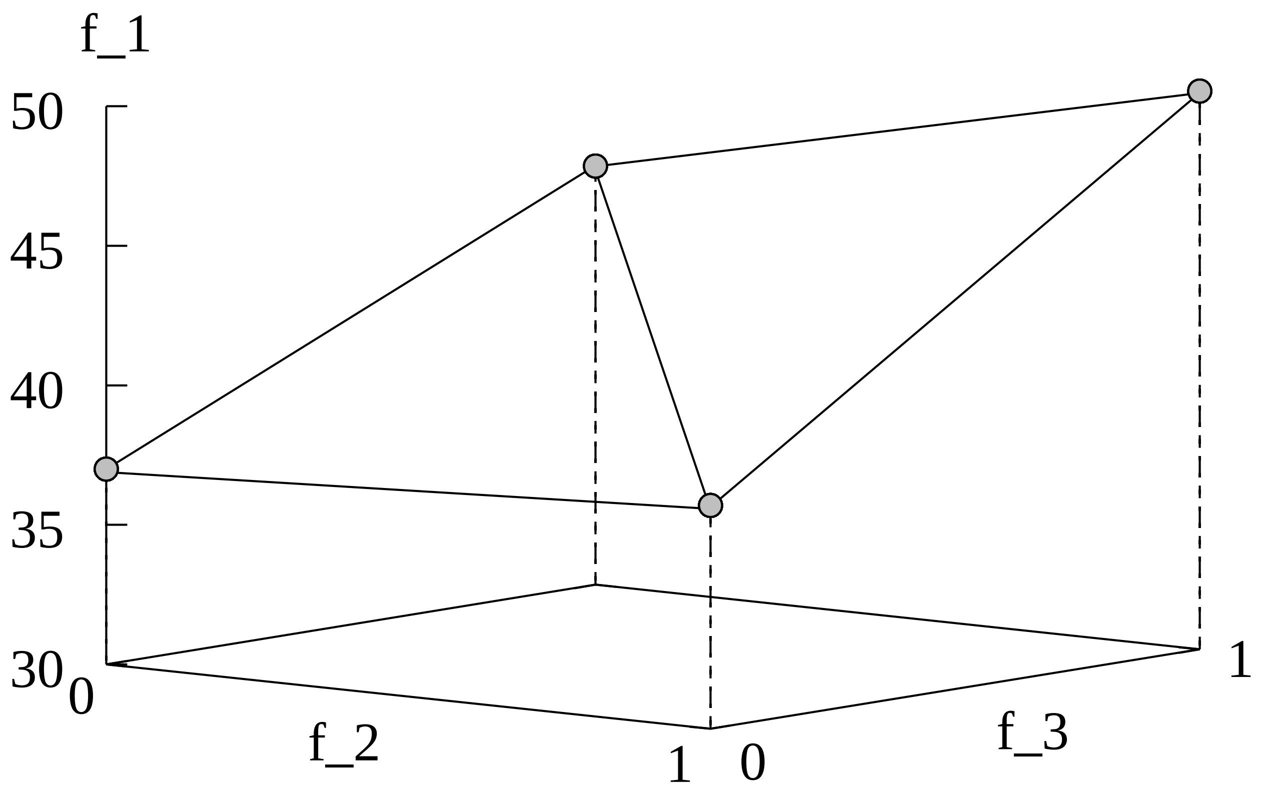

Fig. 3. Maximization of gene subset size for the Leukemia samples. Here, we have considered three objectives: (i) maximization of gene subset size (f_1), (ii) minimization of number of mismatches in training samples (f_2), and (iii) minimization of number of mismatches in test samples (f_3).

1.4 Modified Domination Criterion for Multiple Gene Subset Sizes

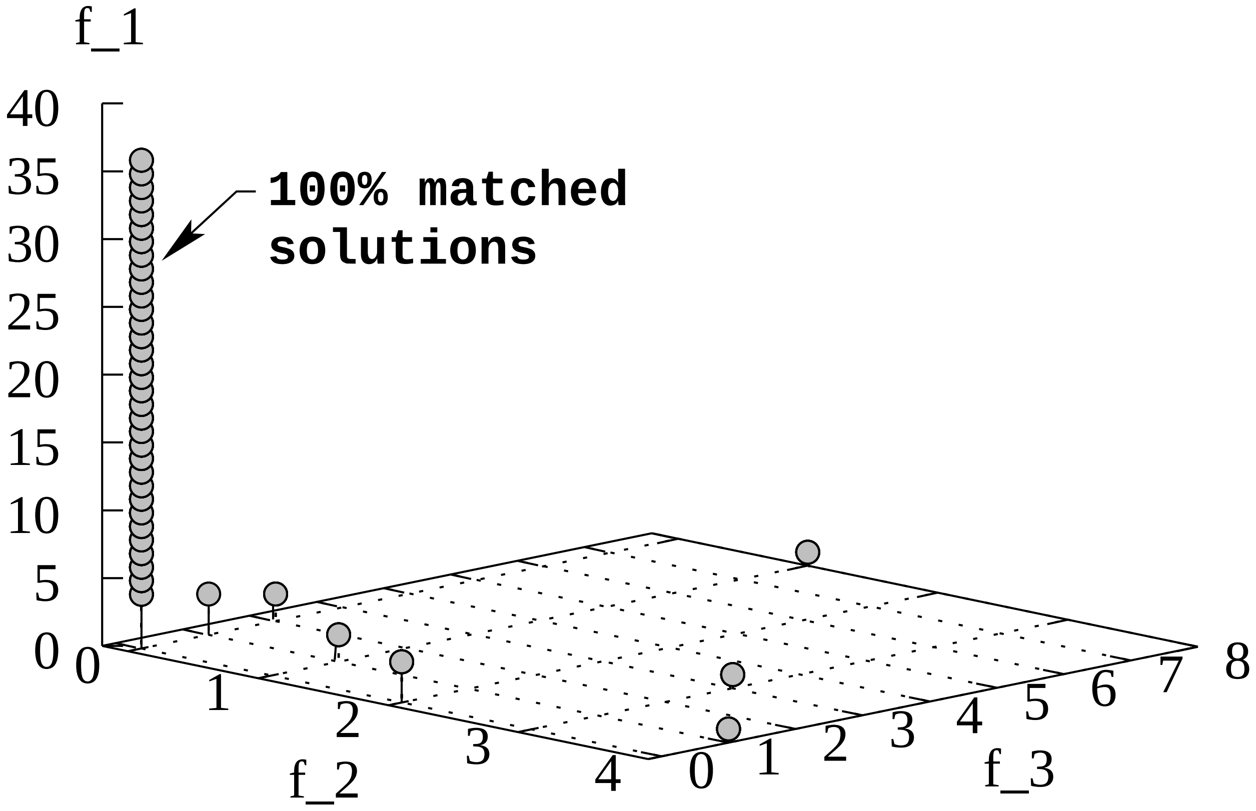

Fig. 4. NSGA-II solutions with modified domination criterion for the Leukemia samples. Here, we have considered three objectives: (i) maximization of gene subset size (f_1), (ii) minimization of number of mismatches in training samples (f_2), and (iii) minimization of number of mismatches in test samples (f_3).

1.5 Multi-Modal NSGA-II for 50-gene Leukemia Data Set with $\theta = 0$

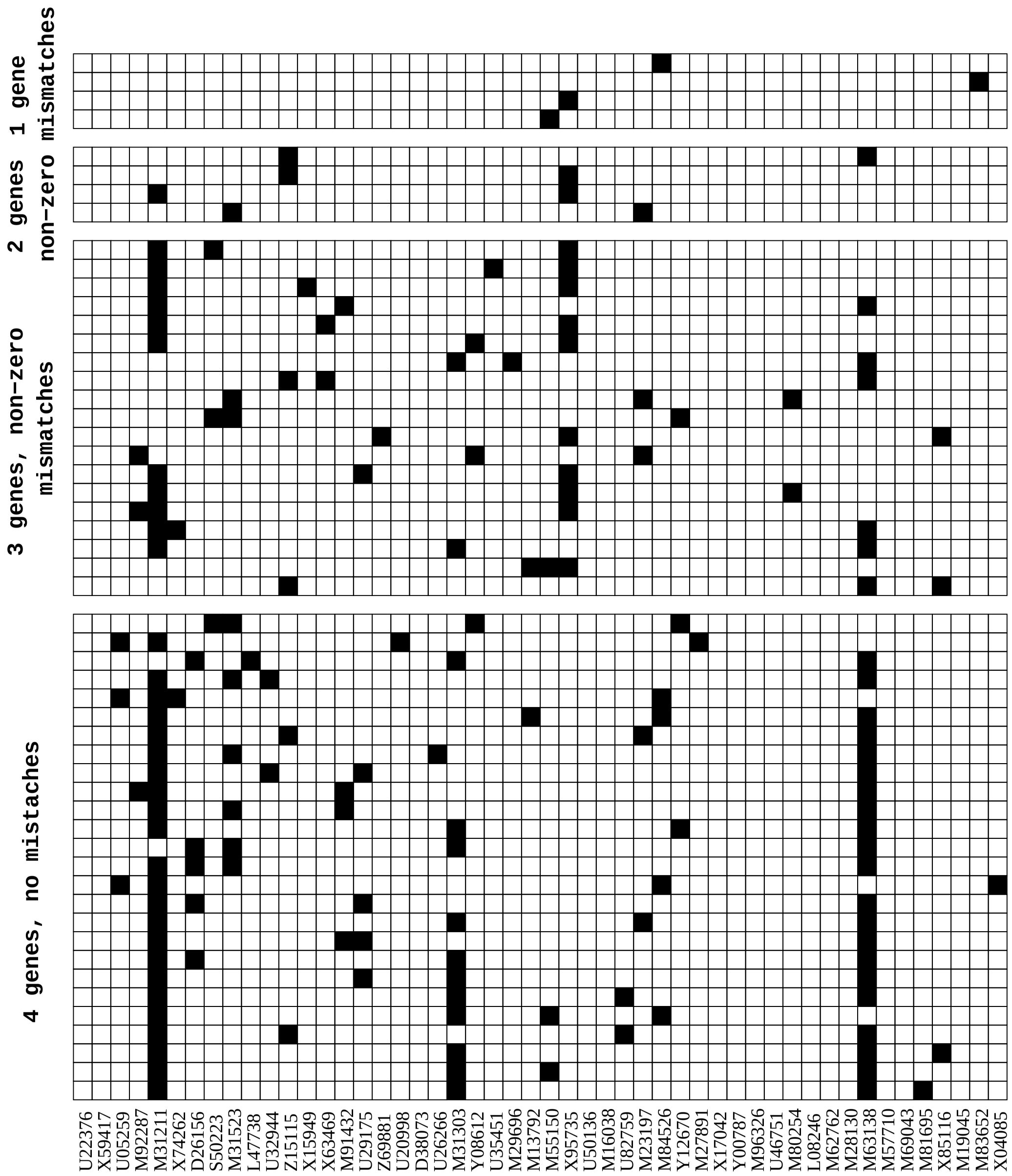

Fig. 5. Multi-modal solutions obtained using NSGA-II for the Leukemia samples without any threshold on prediction strength. Here, each row corresponds one classifier and the shaded box represnts that the gene is included in the classifier.

1.6 Multi-Modal NSGA-II for 50-gene Leukemia Data Set with $\theta = 30\%$

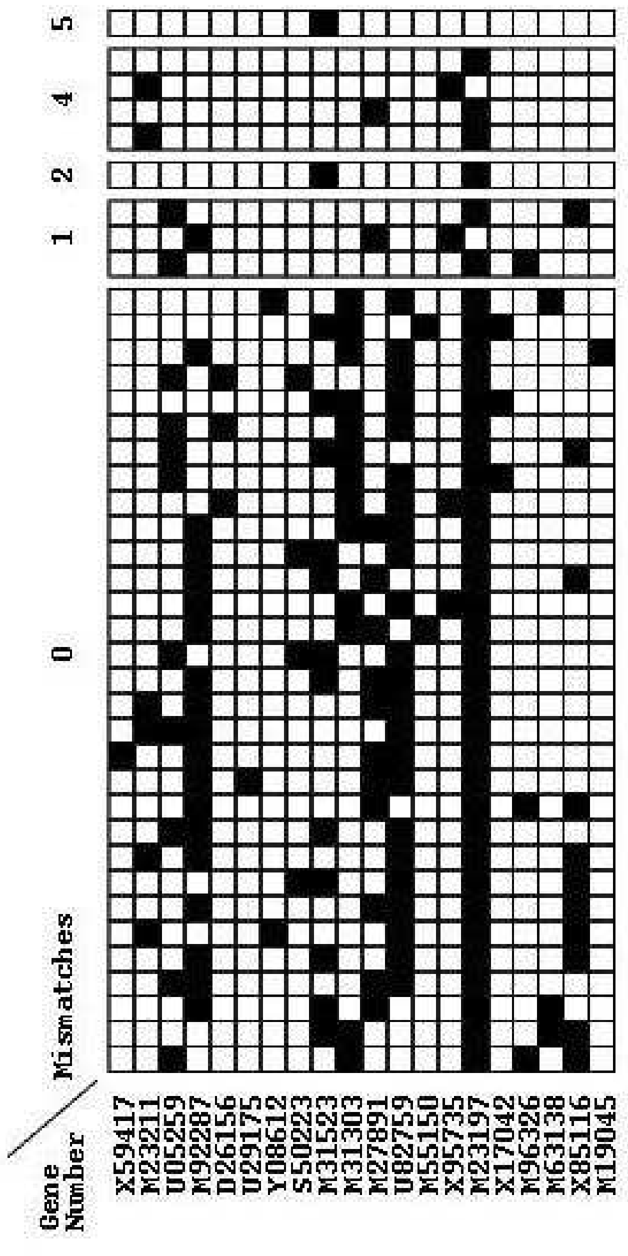

Fig. 6. Multi-modal solutions obtained using NSGA-II for the Leukemia samples with $\theta = 30\%$ threshold on prediction strength. Here, each row corresponds one classifier and the shaded box represnts that the gene is included in the classifier.