Table 3
Comparison of cross-validation accuracies on various combination methods for CNS tumor data set (%)

	Inputs			
	1	2	3	4
BFCS with PART	65.6 ± 11.1	70.9 ± 14.3	74.5 ± 9.7	773[a] ± 8.8
BFCS with S2N	67.5 ± 11.5	67.4 ± 13.2	71.0 ± 9.7	71.1 ± 9.5
BFCS with *U*-test	67.5 ± 11.5	66.3 ± 12.8	71.2 ± 9.2	71.3 ± 9.1
BFCS without screening	67.5 ± 11.5	66.3 ± 12.8	71.2 ± 9.2	71.3 ± 9.1
SVM with PART	65.1 ± 14.9	65.6 ± 12.9	65.3 ± 12.9	65.8 ± 14.2
SVM with S2N	68.0 ± 11.6	66.8 ± 9.0	69.5 ± 8.1	68.3 ± 10.3
SVM with *U*-test	67.6 ± 11.9	66.3 ± 10.1	68.0 ± 8.8	65.7 ± 8.1
SVM without screening	67.6 ± 11.9	65.9 ± 9.6	68.2 ± 8.6	66.3 ± 10.0
FNN-SWEEP with PART	65.1 ± 11.3	66.5 ± 10.3	65.5 ± 12.6	62.2 ± 13.1
FNN-SWEEP with S2N	67.2 ± 12.6	62.9 ± 11.9	60.9 ± 10.4	59.1 ± 12.9
FNN-SWEEP with *U*-test	67.0 ± 12.6	62.5 ± 11.1	60.4 ± 10.1	59.1 ± 14.4
FNN-SWEEP without screening	67.0 ± 12.6	62.7 ± 10.6	60.3 ± 11.6	58.7 ± 11.9
*k*NN with PART	60.3 ± 11.6	59.3 ± 10.5	58.9 ± 12.2	59.8 ± 11.4
*k*NN with S2N	59.5 ± 11.9	57.2 ± 10.8	55.6 ± 10.9	55.0 ± 10.6
*k*NN with *U*-test	59.5 ± 10.9	58.6 ± 11.5	58.0 ± 9.9	57.1 ± 11.1
*k*NN without screening	58.0 ± 12.6	56.6 ± 11.7	57.5 ± 9.7	57.5 ± 9.0
MRA with PART	65.2 ± 11.2	64.2 ± 11.1	61.8 ± 14.6	55.2 ± 11.8
MRA with S2N	67.2 ± 11.9	63.3 ± 12.7	63.0 ± 10.9	56.9 ± 9.6
MRA with *U*-test	67.2 ± 11.9	61.8 ± 11.6	60.1 ± 10.4	55.1 ± 12.1
MRA without screening	67.2 ± 11.9	62.7 ± 11.3	59.1 ± 10.3	54.3 ± 13.6
WV with PART	61.7 ± 14.3	63.9 ± 12.9	60.9 ± 13.0	64.6 ± 12.0
WV with S2N	63.3 ± 13.8	63.3 ± 12.1	62.6 ± 12.1	63.1 ± 11.3
WV with *U*-test	66.1 ± 11.4	62.6 ± 9.3	62.3 ± 10.4	63.2 ± 10.1
WV without screening	66.1 ± 11.4	62.6 ± 11.4	63.0 ± 11.3	63.6 ± 9.6

The average blinded accuracies and their S.D.s were calculated from 10 combination models constructed by PIM.
[a] The highest accuracy.

was compared on the basis of the accuracy by using a blinded data set that was not used for modeling. By using 10 independent class predictor models, the average accuracy for blinded data set was calculated for the CNS and leukemia data sets.

The results of leukemia data are shown in Table 1. The result shows that average accuracy of the PART-BFCS models is the highest as shown in Table 1. In this experiment, top 10 independent class predictor models were constructed by PIM (parameter increasing method) [17] for each condition and data set. And the numbers of construction of high performance model (100% or 97.1% accuracy) were counted for each method as shown in Table 2. Four models among 10 models of five-input show 97.1% or more accuracy for PART-BFCS method. Next, the results for CNS data are shown in Table 3. The inputs used in the predictors were gradually increased from the one-input model to four-input model. As shown in Table 3, the PART-BFCS method clearly showed high performance when compared with the other methods in all input models with the exception of one-input model. The accuracy of the PART-BFCS method gradually increased and eventually, it reached 77.3% in the four-input models. On the other hand, SVM, FNN-SWEEP, *k*NN, MRA, and WV with various filtering showed an accuracy of 55.1–68.3%, which was lower than that of PART-BFCS. Average accuracy of three-input SVM models with S2N was the highest except BFCS models (69.5%). By using *U*-test, however, we found that the accuracy of BFCS with PART was significantly ($P = 5.94 \times 10^{-4}$) higher than one of SVM with S2N. In the four-input models, PART-BFCS method could constructed the most models that showed accuracies were 86.7% (first highest) or 85.7% (second highest), as shown in Table 2. These results could be explained by the facts that PART is the useful filtering method that could improve performances of simple models [5], BFCS is the modeling method in which the model is constructed by assembling simple models, such as one-input FNN. Otherwise, complex models are constructed by other modeling methods. Table 2 shows that the most high performance models were constructed by PART-BFCS method. Therefore, combination of PART and BFCS is the best one.

3.2. Evaluation of prediction results using improved RI_{BFCS}

PART-BFCS method can estimate assurance of results by calculating reliability index for BFCS (RI_{BFCS}). In the present study, we propose improved RI_{BFCS} (new RI_{BFCS}) by modifying equation of RI_{BFCS} (old RI_{BFCS}) for more practical cancer diagnosis. For acute leukemia and CNS data, both RI_{BFCS} of each patient in blinded data were calculated (Fig. 3). Fig. 3 shows distributions of correct and incorrect sample for old and new RI_{BFCS}. It is necessary that there are many incorrect samples in low RI_{BFCS} and many correct samples in high RI_{BFCS}. For old RI_{BFCS}, two distributions are not separated ($P = 0.169$, 0.311), as shown in Fig. 3A and B. On the other hand, they

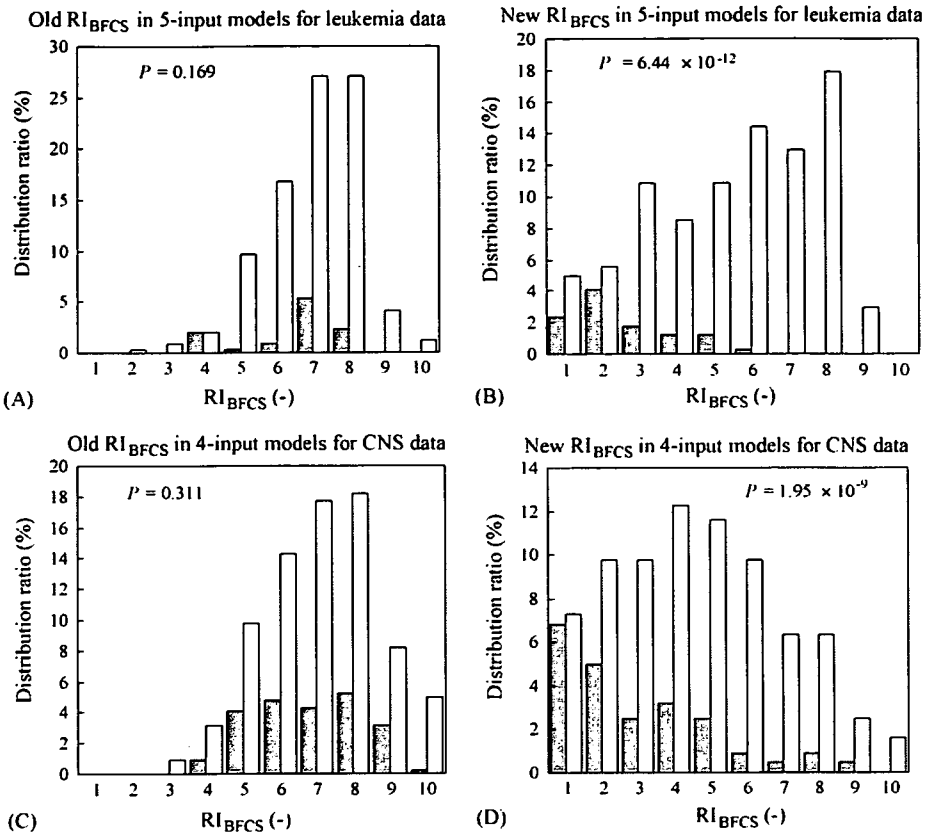

Fig. 3. Comparison of old and new RI_{BFCS}. White bars and gray bars indicate the distribution ratio of correct sample and incorrect samples, respectively. The *P*-values were calculated by Mann–Whitney test and indicate the difference in RI_{BFCS} distribution between the correct and incorrect samples.

are clearly separated ($P = 6.44 \times 10^{-12}$, 1.95×10^{-9}) for new RI_{BFCS}, as shown in Fig. 3B and D. Based on this new index, the discriminated group with over 90% prediction accuracy was separated from the others. For example, the patients who had new $RI_{BFCS} > 5$ corresponded to 48.5% of all patients for leukemia data, and an accuracy of 99.4% was achieved. And, the patients who had new $RI_{BFCS} > 5$ corresponded to 29.3% of all patients for CNS data, and a accuracy of 90.7% was achieved. This result implies new RI_{BFCS} more practical than old one. Old RI is mean distance from boundary line for each gene. BFCS is one of voting methods by assembling simple methods. Improved RI is modified by adding each signs of simple models in the BFCS model. Thus, improved RI is superior to old RI.

3.3. Comparison of selected genes with known prognostic marker genes

We investigated the presence of previously reported prognostic marker genes among the genes selected in the 10 constructed combinations of four-input PART-BFCS models. There were total of 40 genes in 10 models. Some genes were selected several times. In the case of PART-BFCS, 14 genes among 40 genes are independent, as shown in Table 4. Three genes among these 14 genes were reported to be prognostic markers for cancer: The *CCND1* gene was reported by Tan et al. [20] to be a high-risk marker gene. *CCND1* plays an important role in regulating the progress of the cell division during the G1 phase of the cell cycle. Overexpression of *CCND1* correlates with sensitivity to cisplatin [21]. The *LIF* gene was reported by Park et al. [22] to be a low-risk marker gene. *LIF* induces growth arrest and differentiation of cells. The *USP4* (*UNPH*) gene was reported by Frederick et al. [23] to be a low-risk marker gene. These observations accurately matched with low or high gene expression of the above-mentioned three marker genes, as shown in Table 4. These findings suggest that the PART-BFCS method may be used to identify new marker genes.

3.4. Comparison of genes used in PART-BFCS predictors and other predictors for CNS data

We firstly compared FNN-SWEEP and BFCS to investigate numerical character of the genes selected by PART-BFCS. Both FNN-SWEEP and BFCS are based on FNN. The one-gene predictors were constructed for each gene from second input to fourth input in the two methods. And then, average modeling accuracy of one-gene predictors for 10 combinations, was calculated (Table 5). The BFCS genes used as one-gene predictors showed clearly higher accuracy than FNN-SWEEP ones, as shown in Table 5. The average modeling accuracies of the genes from second to fourth were 83.3%, 77.3% and 79.0% for BFCS, and 72.0%, 68.7% and 65.7% for FNN-SWEEP, respectively. The PIM method was used in the FNN-SWEEP. This method is

Table 4
The genes used in PART-BFCS class predictor for one set of CNS tumor data set

Accession number	Gene name	Descriptions	Times of selection	Average intensity of surviving patients[a]	Average intensity of dead patients[b]	Threshold of model[c]	Prognostic markers
U20657	USP4	Ubiquitin specific protease 4 (proto-oncogene)	9	796	127	391	[d]
X59798	CCND1	Cyclin D1 (PRAD1: parathyroid adenomatosis 1)	6	0	2176	330	[e]
M73547	C5orf18	Chromosome 5 open reading frame 18	6	1439	275	797	
AB000460	C4orf8	Chromosome 4 open reading frame 8	4	2429	1094	1605	
L33243	PKD1	Polycystic kidney disease 1 (autosomal dominant)	4	1498	228	815	
X13967	LIF	Leukemia inhibitory factor (cholinergic differentiation factor)	2	464	5	206	[f]
L10333	RTN1	Reticulon 1	2	4483	821	1747	
D30756	M17S2	Membrane component, chromosome 17, surface marker 2	1	591	59	236	
D83018	NELL2	NEL-like 2 (chicken)	1	2710	851	1416	
HG2238-HT2321	NUMA1	Nuclear mitotic apparatus protein 1, alt. splice form 2	1	2833	1197	1721	
J04046	CALM3	Calmodulin 3 (phosphorylase kinase, delta)	1	3287	1022	1753	
S76475	NTRK3	Neurotrophic tyrosine kinase, receptor, type 3 (TrkC)	1	2002	96	687	
U25849	ACPI	Acid phosphatase 1, soluble	1	206	1082	602	
Y09616	CES2	Carboxylesterase 2 (intestine, liver)	1	2894	1065	1706	

[a] The average intensity of gene expression in the patients predicted as survivors.
[b] The average intensity of gene expression in the patients predicted as dead.
[c] The threshold of gene expression in the weak learner model.
[d] The marker gene reported by Frederick et al. [23] as a low-risk marker.
[e] The marker gene reported by Tan et al. [20] as a high-risk marker.
[f] The marker gene reported by Park et al. [22] as a low-risk marker.

very useful to select input combination that shows high accuracy by combining low accuracy inputs. But, the application of PIM to high dimensional data, such as microarray data, may cause overfitting. On the other hand, the boosting used in the BFCS is the method that can construct high-accuracy predictor by combining one-gene predictors. Thus, low-accuracy one-gene predictors are hardly selected. It may be for this reason that BFCS showed high performance.

Next, BFCS were compared with PART-BFCS. Average of gene expression for the 40 genes in 10 combinations of four-input models, was calculated for each class (survivors or dead). And average standard deviation (S.D.) of lower gene expression class for 40 genes is shown in Table 6. Table 6 shows that the S.D. of PART-BFCS was lower than one of BFCS. The values of S.D.s were 0.57 for BFCS and 0.39 for PART-BFCS. This tendency is corresponding to the fact previously reported by us [5]. The genes with low S.D. in lower class may show high generalization performance.

3.5. IF-THEN rules extracted from PART-BFCS model

After modeling, the IF-THEN rules for CNS prognosis were obtained from the model of highest blind accuracy among the 10 combinations. The model includes the *CCND1* gene and *USP4* (*UNPH*) gene as known prognostic markers. The IF-THEN rules have been obtained as a matrix that is classified based on expression level of such selected genes (Fig. 4). Using this matrix, simple and precise rules were obtained as follows. The simplest rule is that patients with high expression of *CCND1* gene are likely to exhibit poor prognosis. Six patients showed high expression of *CCND1* gene and five of them were actually dead patients, which corresponds to 28% (5/18) of all the dead

Table 5
Average modeling accuracy of one-input models between BFCS and FNN-SWEEP (%)

	Order of selection		
	Second	Third	Fourth
BFCS	83.3	77.3	79.0
FNN-SWEEP	72.0	68.7	65.7

In this experiment, 10 combinations of four-input models were constructed. Three genes from second to fourth in four-input were selected as combination of genes for each method. The modeling accuracies when these three genes were used alone as one-gene predictors, were calculated for 10 combinations.

Table 6
Average S.D. of gene expression in lower class between BFCS and PART-BFCS

Methods	The S.D.s of lower class
BFCS	0.56
PART-BFCS	0.39

Average of gene expression for the 40 genes in 10 combinations of four-input models were calculated for each class (survivors or dead). And then, average standard deviation (S.D.) of lower gene expression class for 40 genes was calculated.

				CCND1			
				Low		High	
				PKD1			
				Low	High	Low	High
USP4 (Unph)	Low	*C5orf18*	Low	**3(M), 4(B), 7(M), 8(M), 9(B), 10(M), 12(M), 13(M), 17(M)**	**11(B), 16(M)**	**6(B), 18(M)**	**5(M)**
			High	**2(B)**, *42(B), 55(B)*	*38(M), 44(M), 47(M), 50(M), 57(B)*		
	High		Low	*19(M)*	*40(M), 49(M)*	**1(M)**	**15(M)**
			High	*51(B), 54(M)*	**14(B)** *20(M), 21(M), 37(B), 39(M), 41(M), 43(M), 45(M), 46(B), 48(B), 52(M), 53(B), 58(M), 59(M)*		*56(M)*

Fig. 4. IF-THEN rules in the top two model of PART-BFCS. Since the expression level of each gene can be divided into either high or low groups using fuzzy logic, this model comprised 16 ($=2^4$) fuzzy rules. Numbers in each matrix are identical to the patient numbers previously described by Pomeroy et al. [7]. Numbers in bold type and italic type indicate the poor and good prognosis patients, respectively. Patient numbers are placed in the matrix according to the expression levels of each patient. Patient numbers in the circle represent incorrect classification by the PART-BFCS. (B) indicates sample in blinded data set. (M) indicates sample in modeling data set.

patients. Next simple rule is that patients with low expression of *CCND1* gene and low expression of *USP4* (*UNPH*) gene are likely to exhibit poor prognosis. Nineteen patients showed low expression of *USP4* (*UNPH*) gene and 12 of them were actually dead patients, which corresponds to 92% (12/13) of dead patients showing low expression of the *CCND1* gene. Nineteen patients showed high expression of *USP4* (*UNPH*) gene and low expression of *CCND1* gene, and 18 of them (95%) were actually surviving patients, which corresponds to 69% (18/26) of all the surviving patients. It was found that surviving or dead patients were clustered at specific parts of the matrix. The following rule was also found: patients were likely to exhibit a poor prognosis when the *USP4* (*UNPH*) expression was low and *C5orf18* expression was low. It was also found that on this matrix, two patients showing poor prognosis were incorrectly predicted as showing good prognosis. This may be due to the inability for complete removal of their tumors by CNS surgery.

4. Conclusions

In the present study, we investigated combinations of various filter and wrapper approaches, and found that combination method of PART and BFCS (a kind of boosting) is significantly superior to other methods with regard to high prediction accuracy for construction of class predictor from gene expression data. This method could select some marker genes related to cancer outcome. In addition, we proposed improved RI_{BFCS} of PART-BFCS. Based on this new index, the discriminated group with over 90% prediction accuracy was separated from the others. It is necessary that there are about 90% or more prediction accuracy in the practical diagnosis application. These results suggest that the PART-BFCS method has a high potential to function as a new method of marker gene selection for the diagnosis of patients, using high dimensional data such as DNA microarray, mass spectrometry (MS), and two-dimensional polyacrylamide gel electrophoresis (2D-PAGE).

Acknowledgment

This work was supported by the Hori Information Science Promotion Foundation for financial support.

References

[1] T. Ando, M. Suguro, T. Hanai, T. Kobayashi, H. Honda, M. Seto, Fuzzy neural network applied to gene expression profiling for predicting the prognosis of diffuse large B-cell lymphoma, Jpn. J. Cancer Res. 93 (2002) 1207–1212.
[2] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (2002) 389–422.
[3] H. Takahashi, H. Honda, A new reliable cancer diagnosis method using boosted fuzzy classifier with SWEEP operator method, J. Chem. Eng. Jpn. 38 (2005) 763–773.
[4] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, E.S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537.
[5] H. Takahashi, T. Kobayashi, H. Honda, Construction of robust prognostic predictors by using projective adaptive resonance theory as a gene filtering method, Bioinformatics 21 (2005) 179–186.
[6] D.J. Park, P.T. Vuong, S. de Vos, D. Douter, H.P. Koeffler, Comparative analysis of genes regulated by PML/RAR alpha and PLZF/RAR alpha in response to retinoic acid using oligonucleotide arrays, Blood 102 (2003) 3727–3736.
[7] S.L. Pomeroy, P. Tamayo, M. Gaasenbeek, L.M. Sturla, M. Angelo, M.E. McLaughlin, J.Y. Kim, L.C. Goumnerova, P.M. Black, C. Lau, J.C. Allen, D. Zagzag, J.M. Olson, T. Curran, C. Wetmore, J.A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D.N. Louis, J.P. Mesirov, E.S. Lander, T.R. Golub, Prediction of central nervous system embryonal tumour outcome based on gene expression, Nature 415 (2002) 436–442.
[8] R.E. Schapire, The strength of weak learnability, Mach. Learn. 5 (1990) 197–227.
[9] Y. Freund, R. Schapire, A decision-theoretic generalization of online learning and an application to boosting, J. Comput. Syst. Sci. 55 (1997) 119–139.
[10] Y. Freund, Adaptive version of the boost by majority algorithm, Mach. Learn. 43 (2000) 293–318.
[11] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, Ann. Stat. 28 (2000) 337–407.

[12] S. Horikawa, T. Furuhashi, Y. Uchikawa, On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm, IEEE Trans. Neural Networ. 3 (1992) 801–806.
[13] Y. Huang, Y. Li, Prediction of protein subcellular locations using fuzzy k-NN method, Bioinformatics 20 (2004) 21–28.
[14] V.N. Vapnik, A. Chervonenkis, A note on one class of perceptrons, Automat. Rem. Control 25 (1964) 821–837.
[15] T. Joachims, Making Large-scale SVM Learning Practical, MIT Press, Cambridge, 1999.
[16] H. Noguchi, T. Hanai, W. Takahashi, T. Ichii, M. Tanikawa, S. Masuoka, H. Honda, T. Kobayashi, Model construction for quality of beer and brewing process using FNN, Kagaku Kougaku Ronbunshu 25 (1999) 695–701.
[17] H. Noguchi, T. Hanai, H. Honda, L.C. Harrison, T. Kobayashi, Fuzzy neural network-based prediction of the motif for MHC class II binding peptides, J. Biosci. Bioeng. 92 (2001) 227–231.
[18] Y. Cao, J. Wu, Projective ART for clustering data sets in high dimensional spaces, Neural Networ. 15 (2002) 105–120.
[19] Y. Cao, J. Wu, Dynamics of projective adaptive resonance theory model: the foundation of PART algorithm, IEEE Trans. Neural Networ. 15 (2004) 245–260.
[20] P.G. Tan, Z. Xing, Z.Q. Li, Expression of cyclin D1 in brain gliomas and its significance, Ai Zheng 23 (2004) 63–65.
[21] J. Akervall, D.M. Kurnit, M. Adams, S. Zhu, S.G. Fisher, C.R. Bradford, T.E. Carey, Overexpression of cyclin D1 correlates with sensitivity to cis-platin in squamous cell carcinoma cell lines of the head and neck, Acta Otolaryngol. 124 (2004) 851–857.
[22] J.I. Park, C.J. Strock, D.W. Ball, B.D. Nelkin, The Ras/Raf/MEK/extracellular signal-regulated kinase pathway induces autocrine–paracrine growth inhibition via the leukemia inhibitory factor/JAK/STAT pathway, Mol. Cell. Biol. 23 (2003) 543–554.
[23] A. Frederick, M. Rolfe, M.I. Chiu, The human UNP locus at 3p21.31 encodes two tissue-selective, cytoplasmic isoforms with deubiquitinating activity that have reduced expression in small cell lung carcinoma cell lines, Oncogene 16 (1998) 153–165.