Supplemental materials: Performance-weighted-voting model: an ensemble machine learning method for cancer type classification using whole-exome sequencing mutation

Yawei Li[1], Yuan Luo[1*]

[1] Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA
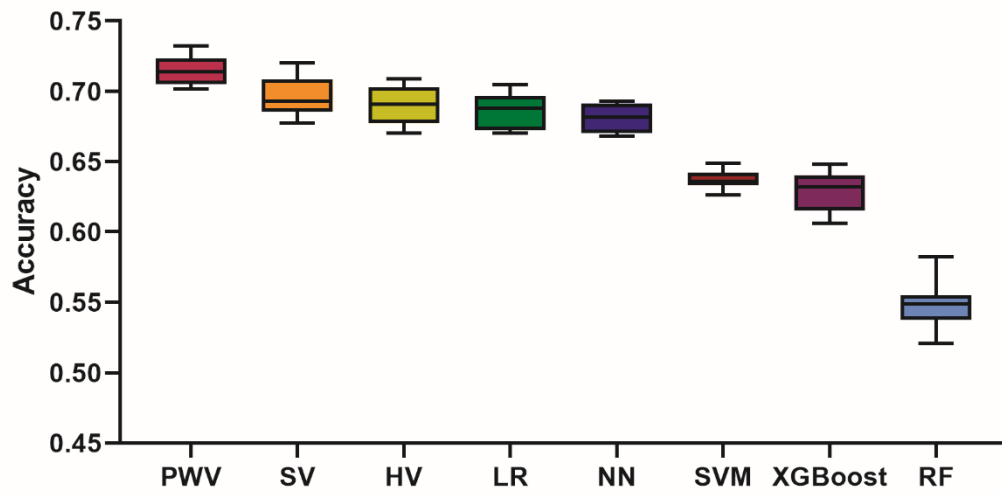
[*] Correspondence: yuan.luo@northwestern.edu

Figure S1. The predictive performance of the performance-weighted-voting model compared with the other models. Hard-voting model (HV, yellow box), soft-voting model (SV, orange box) and performance-weighted-voting model (PWV, violet red box) were introduced compared with the five weak classifiers. Each model was trained and predicted 10 times independently.

Table S1: The precision, recall and F1-score of the eight models.

		BLCA	BRCA	GBM	HNSC	KIRC	LGG	LIHC	LUAD	LUSC	PRAD	SKCM	STAD	THCA	UCEC
Logistic Regression	precision	72%[a] ± 5%[b]	61% ± 3%	60% ± 4%	56% ± 5%	85% ± 3%	88% ± 2%	71% ± 4%	74% ± 4%	73% ± 5%	46% ± 5%	94% ± 2%	66% ± 6%	70% ± 5%	79% ± 4%
	recall	61% ± 4%	66% ± 3%	63% ± 5%	58% ± 5%	69% ± 6%	82% ± 3%	58% ± 6%	59% ± 4%	70% ± 4%	70% ± 5%	79% ± 3%	49% ± 6%	92% ± 1%	76% ± 5%
	F1-score	66% ± 3%	63% ± 2%	61% ± 4%	57% ± 4%	76% ± 4%	85% ± 2%	64% ± 4%	65% ± 3%	72% ± 4%	56% ± 3%	86% ± 2%	56% ± 4%	79% ± 3%	77% ± 3%
SVM	precision	64% ± 2%	44% ± 3%	68% ± 6%	52% ± 3%	89% ± 3%	94% ± 2%	70% ± 8%	78% ± 7%	65% ± 5%	41% ± 4%	98% ± 2%	67% ± 8%	76% ± 5%	86% ± 5%
	recall	54% ± 3%	66% ± 5%	48% ± 6%	50% ± 4%	61% ± 6%	79% ± 3%	46% ± 5%	44% ± 5%	71% ± 4%	77% ± 6%	76% ± 3%	40% ± 5%	94% ± 2%	68% ± 5%
	F1-score	58% ± 2%	52% ± 3%	56% ± 6%	51% ± 2%	72% ± 5%	86% ± 2%	55% ± 5%	56% ± 4%	68% ± 4%	53% ± 3%	85% ± 2%	49% ± 4%	84% ± 4%	76% ± 4%
Random Forest	precision	65% ± 6%	71% ± 4%	41% ± 6%	64% ± 4%	79% ± 5%	74% ± 8%	64% ± 5%	73% ± 5%	45% ± 4%	36% ± 4%	68% ± 7%	53% ± 12%	34% ± 6%	80% ± 7%
	recall	54% ± 7%	45% ± 3%	47% ± 6%	29% ± 5%	69% ± 7%	82% ± 2%	39% ± 5%	33% ± 6%	83% ± 4%	29% ± 8%	76% ± 3%	10% ± 4%	99% ± 1%	67% ± 3%
	F1-score	59% ± 6%	55% ± 3%	43% ± 5%	39% ± 5%	73% ± 6%	78% ± 5%	48% ± 5%	45% ± 6%	58% ± 3%	32% ± 5%	71% ± 5%	17% ± 6%	51% ± 6%	73% ± 5%
XGBoost	precision	68% ± 6%	56% ± 2%	46% ± 5%	55% ± 4%	82% ± 6%	87% ± 3%	65% ± 5%	61% ± 5%	59% ± 5%	39% ± 5%	86% ± 2%	50% ± 7%	74% ± 4%	80% ± 4%
	recall	58% ± 6%	67% ± 3%	41% ± 6%	51% ± 5%	69% ± 6%	85% ± 3%	52% ± 6%	52% ± 5%	61% ± 5%	67% ± 4%	76% ± 3%	32% ± 5%	77% ± 3%	73% ± 4%
	F1-score	62% ± 5%	61% ± 2%	43% ± 5%	52% ± 4%	75% ± 6%	86% ± 2%	57% ± 5%	56% ± 3%	60% ± 4%	49% ± 5%	81% ± 2%	39% ± 4%	75% ± 3%	76% ± 2%
Neural Network	precision	67% ± 4%	64% ± 3%	57% ± 3%	57% ± 5%	85% ± 3%	90% ± 3%	67% ± 5%	67% ± 4%	70% ± 4%	45% ± 4%	93% ± 2%	60% ± 5%	76% ± 5%	80% ± 5%
	recall	60% ± 5%	63% ± 3%	60% ± 4%	54% ± 3%	70% ± 6%	82% ± 3%	58% ± 4%	58% ± 4%	73% ± 4%	72% ± 4%	81% ± 2%	52% ± 5%	90% ± 2%	74% ± 6%
	F1-score	63% ± 4%	64% ± 2%	58% ± 3%	55% ± 3%	76% ± 4%	86% ± 2%	62% ± 3%	62% ± 2%	71% ± 3%	55% ± 3%	86% ± 1%	56% ± 4%	83% ± 3%	77% ± 3%
Hard-Voting	precision	72% ± 5%	61% ± 5%	59% ± 5%	58% ± 5%	88% ± 4%	90% ± 2%	75% ± 5%	74% ± 4%	67% ± 7%	45% ± 4%	94% ± 3%	71% ± 7%	72% ± 5%	86% ± 5%
	recall	63% ± 5%	67% ± 3%	60% ± 7%	56% ± 4%	69% ± 5%	82% ± 3%	59% ± 6%	55% ± 5%	76% ± 4%	74% ± 6%	79% ± 2%	45% ± 6%	93% ± 1%	77% ± 4%
	F1-score	67% ± 4%	64% ± 2%	59% ± 5%	57% ± 3%	78% ± 4%	86% ± 2%	66% ± 4%	63% ± 3%	71% ± 4%	56% ± 3%	86% ± 2%	55% ± 5%	81% ± 3%	81% ± 3%
Soft-Voting	precision	71% ± 4%	61% ± 4%	59% ± 5%	58% ± 5%	88% ± 3%	89% ± 2%	74% ± 5%	73% ± 6%	69% ± 6%	46% ± 4%	94% ± 3%	67% ± 6%	73% ± 6%	86% ± 5%
	recall	63% ± 5%	68% ± 2%	59% ± 6%	56% ± 4%	70% ± 6%	83% ± 3%	59% ± 5%	57% ± 6%	75% ± 5%	73% ± 4%	80% ± 2%	46% ± 6%	93% ± 2%	76% ± 4%
	F1-score	67% ± 4%	64% ± 2%	59% ± 5%	57% ± 3%	78% ± 4%	86% ± 2%	65% ± 4%	64% ± 2%	72% ± 4%	56% ± 3%	86% ± 2%	54% ± 4%	82% ± 4%	81% ± 3%
Performance-Weighted-Voting	precision	72% ± 4%	69% ± 2%	61% ± 4%	58% ± 4%	88% ± 2%	91% ± 3%	62% ± 4%	73% ± 5%	74% ± 5%	48% ± 4%	94% ± 2%	60% ± 6%	80% ± 3%	85% ± 4%
	recall	65% ± 5%	62% ± 2%	66% ± 7%	59% ± 4%	72% ± 7%	83% ± 3%	69% ± 4%	60% ± 4%	71% ± 3%	78% ± 5%	81% ± 2%	57% ± 7%	93% ± 2%	79% ± 4%
	F1-score	68% ± 3%	66% ± 2%	63% ± 5%	60% ± 3%	79% ± 5%	87% ± 2%	65% ± 2%	66% ± 3%	72% ± 2%	60% ± 3%	87% ± 2%	58% ± 4%	86% ± 2%	82% ± 3%

[a] The average number of 10 repeats

[b] The standard deviation of 10 repeats

Table S2: 52 misclassification cases that received the same predictions from five classifiers

Case ID	Cancer Type	Performance-weighted-voting
TCGA-FD-A3B7	BLCA	BRCA
TCGA-XF-AAMH	BLCA	LIHC
TCGA-A7-A0CH	BRCA	THCA
TCGA-A2-A25F	BRCA	THCA
TCGA-76-6193	GBM	BRCA
TCGA-26-1442	GBM	LGG
TCGA-06-5417	GBM	LGG
TCGA-32-4208	GBM	LGG
TCGA-14-4157	GBM	LGG
TCGA-CR-7368	HNSC	BLCA
TCGA-MT-A7BN	HNSC	GBM
TCGA-CR-7371	HNSC	LUSC
TCGA-DQ-5629	HNSC	LUSC
TCGA-CR-7370	HNSC	LUSC
TCGA-CN-6020	HNSC	LUSC
TCGA-CR-7398	HNSC	LUSC
TCGA-CV-A45W	HNSC	LUSC
TCGA-CV-5443	HNSC	PRAD
TCGA-HD-A634	HNSC	PRAD
TCGA-BP-5010	KIRC	GBM
TCGA-BP-5199	KIRC	UCEC
TCGA-VM-A8C9	LGG	BRCA
TCGA-HT-8110	LGG	GBM
TCGA-TM-A7C3	LGG	GBM
TCGA-QH-A6CX	LGG	GBM
TCGA-DU-7292	LGG	GBM
TCGA-WJ-A86L	LIHC	HNSC
TCGA-XR-A8TG	LIHC	HNSC
TCGA-LG-A9QC	LIHC	KIRC
TCGA-DD-A4NA	LIHC	LGG
TCGA-97-8177	LUAD	GBM
TCGA-55-A57B	LUAD	LGG
TCGA-L9-A743	LUAD	LUSC
TCGA-55-8301	LUAD	LUSC
TCGA-69-8255	LUAD	LUSC
TCGA-22-4605	LUSC	HNSC
TCGA-63-6202	LUSC	LUAD
TCGA-HC-A9TH	PRAD	BRCA
TCGA-J9-A8CL	PRAD	BRCA
TCGA-EJ-7125	PRAD	LGG
TCGA-KK-A6E7	PRAD	LGG
TCGA-EJ-A8FU	PRAD	THCA
TCGA-FW-A3I3	SKCM	GBM
TCGA-FS-A1ZU	SKCM	PRAD
TCGA-CG-5722	STAD	BRCA
TCGA-D7-6518	STAD	BRCA
TCGA-BR-A4J9	STAD	BRCA
TCGA-BR-8592	STAD	BRCA
TCGA-D7-8573	STAD	LIHC
TCGA-VQ-A8PJ	STAD	LUSC
TCGA-D1-A1NX	UCEC	BRCA
TCGA-B5-A5OE	UCEC	HNSC