Electronic Supplementary Information

for

Bayesian graph convolutional network for reliable prediction of molecular properties with uncertainty quantification

Seongok Ryu[a], Yongchan Kwon[b] and Woo Youn Kim[a,c]

[a]Department of Chemistry, KAIST, 291 Daehak-Ro, Yuseong-Gu, Daejeon, 34141, Republic of Korea

[b]Department of statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

[c]KI for Artificial Intelligence, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea

1. Training details

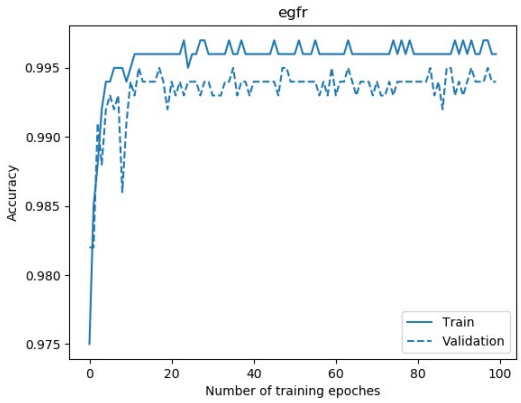
(1) The models in section 4.1

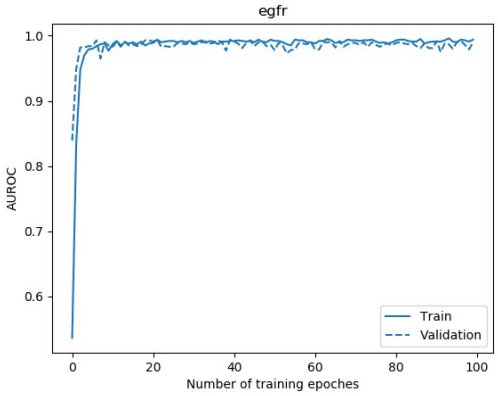
1) DUD-E dataset

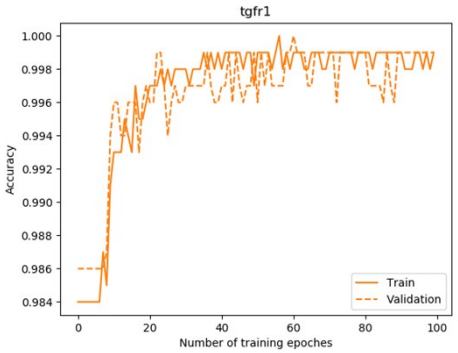
* Number of training data and time for training

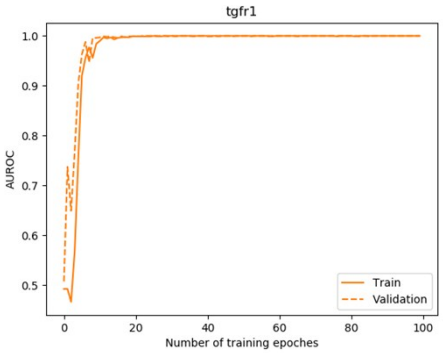
	Number of samples			Time for training (s)
	Train	Validation	Test	
egfr	25627	2847	7118	9887
tgfr1	6217	690	1726	2605
vgfr2	18259	2028	5071	7234
braf	7274	808	2020	3319
abl1	7872	874	2186	3581

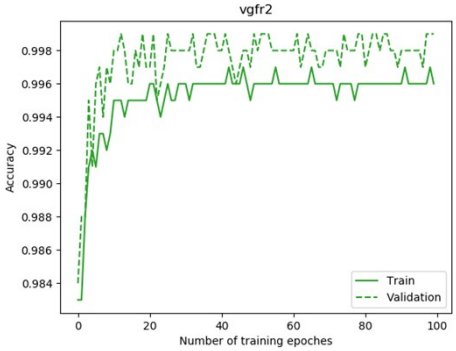
* Accuracy & AUROC curve

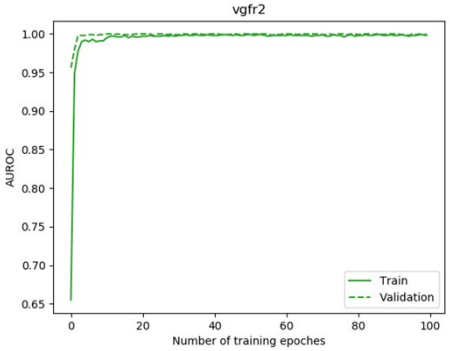


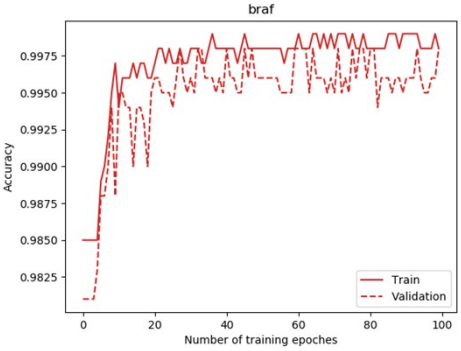


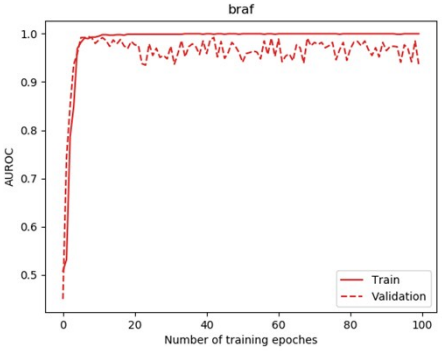


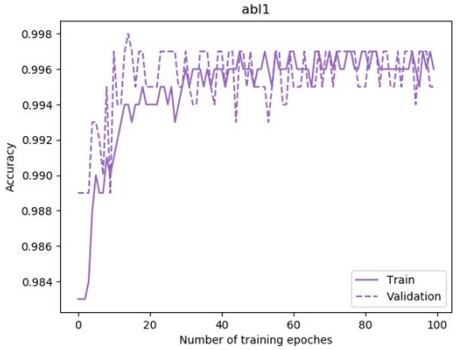


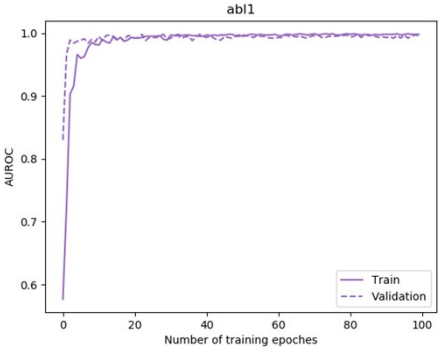

vgfr2
Accuracy
Number of training epoches
Train
Validation


vgfr2
AUROC
Number of training epoches
Train
Validation


braf
Accuracy
Number of training epoches
Train
Validation


braf
AUROC
Number of training epoches
Train
Validation


abl1
Accuracy
Number of training epoches
Train
Validation

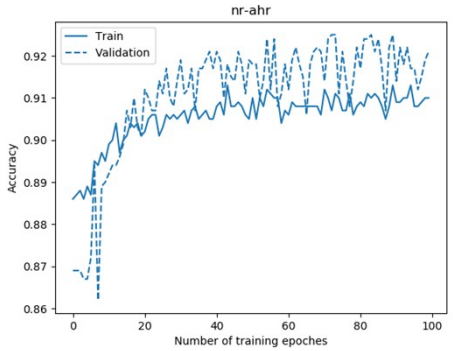

abl1
AUROC
Number of training epoches
Train
Validation

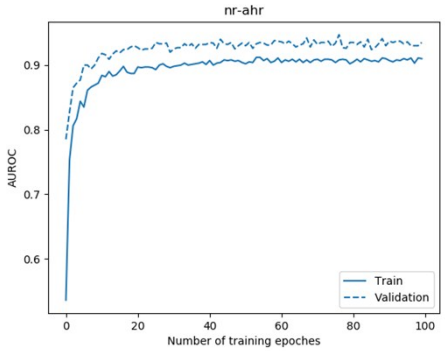
2) Tox-21 dataset

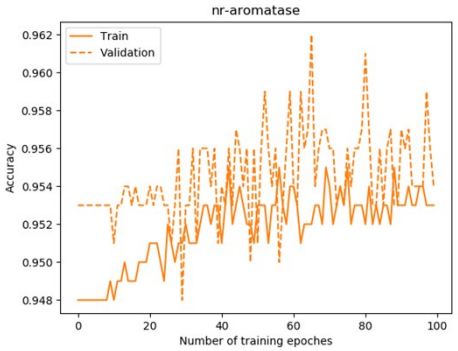
* Number of training data and time for training

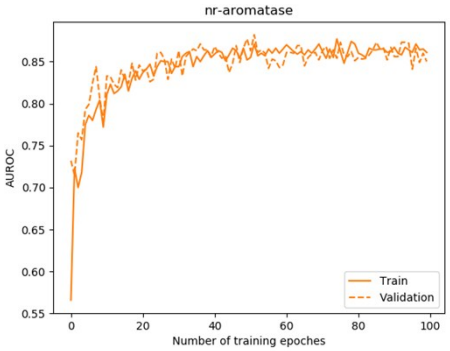
	Number of samples			Time for training (s)
	Train	Validation	Test	
nr-ahr	6481	720	1800	2680
nr-aromatase	5706	634	1585	2358
nr-ppar-gamma	6489	721	1802	2644
sr-mmp	5801	644	1611	2439
sr-p53	6813	756	1892	2508

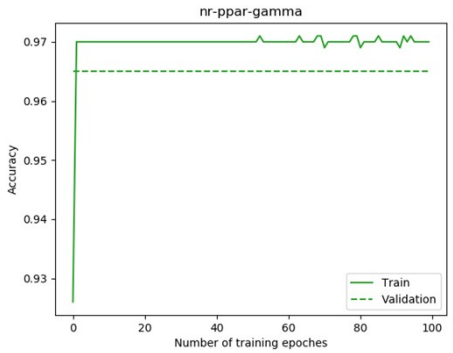
* Accuracy & AUROC curve



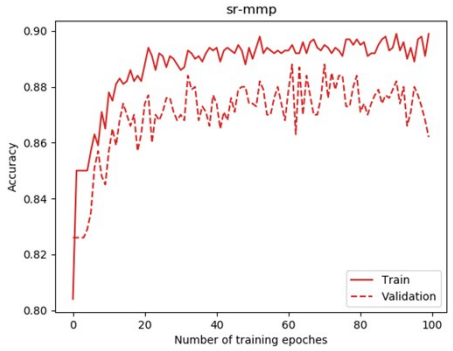


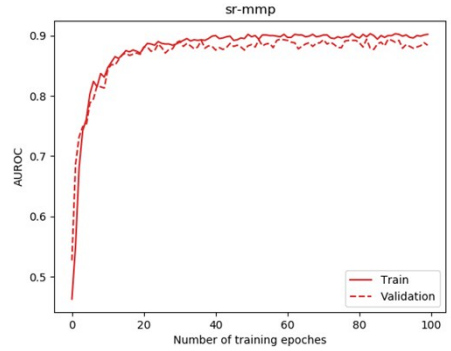


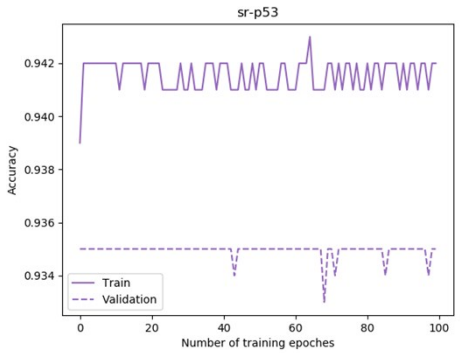


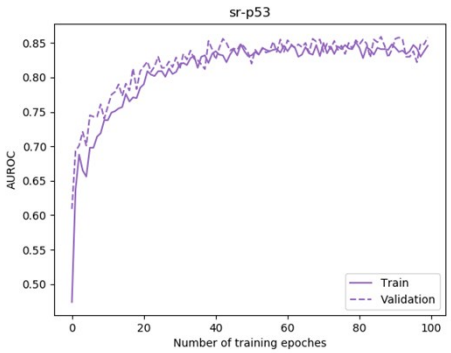

nr-ppar-gamma
Accuracy
Number of training epoches
Train
Validation


nr-ppar-gamma
AUROC
Number of training epoches
Train
Validation


sr-mmp
Accuracy
Number of training epoches
Train
Validation


sr-mmp
AUROC
Number of training epoches
Train
Validation


sr-p53
Accuracy
Number of training epoches
Train
Validation

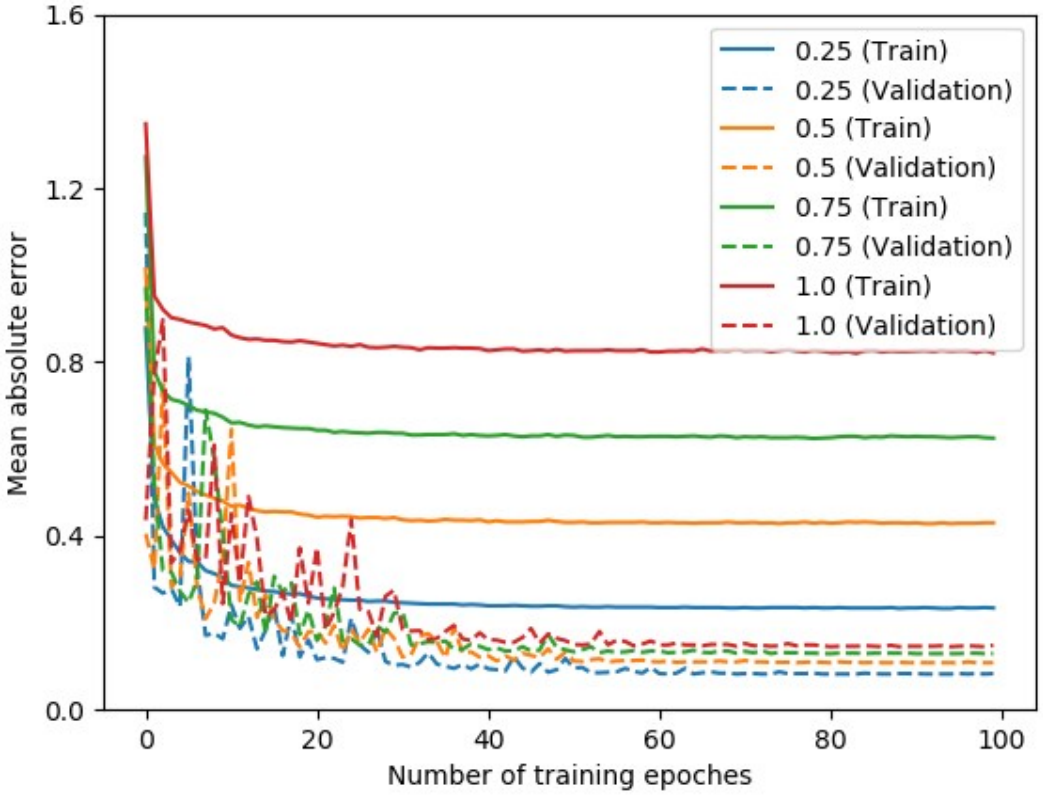

sr-p53
AUROC
Number of training epoches
Train
Validation

(2) The models in section 4.3

*** Number of training data and time for training**

	Number of samples			Time for training (s)
	Train	Validation	Test	
$\sigma^2 = 0.25, 0.5, 0.75, 1.0$	97287	10809	27023	$\sim 4 \times 10^4$

*** Mean absolute error curve**



Note that we adjusted noise only for the training samples. (Therefore, the mean absolute error of training curves are shifted.)

(3) The model in section 4.4

*** Number of training data and time for training**

	Number of samples			Time for training (s)
	Train	Validation	Test	
-	21585	2398	5995	8433

*** Mean absolute error curve**