Supplementary to DeepSol

Raghvendra Mall
Qatar Computing Research Institute,
Hamad Bin Khalifa University, Doha, Qatar

1 Parameter Settings

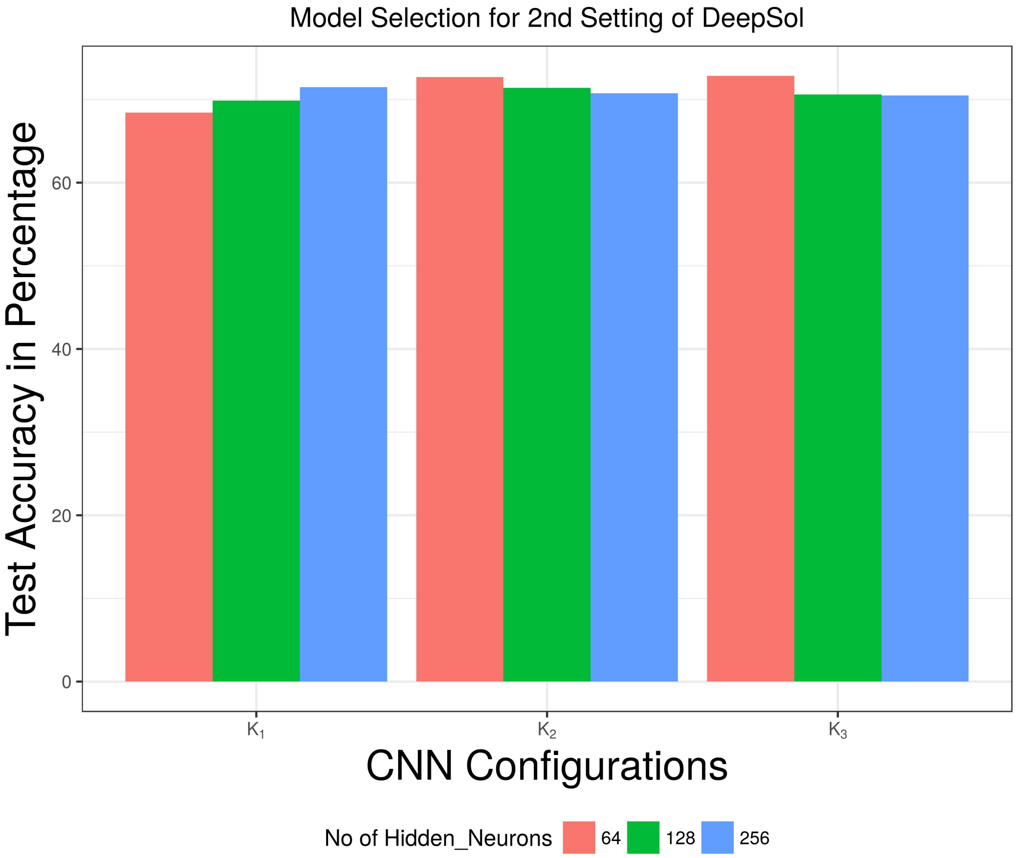

Figure 1: *Here* K_1*,* K_2 *and* K_3 *represent the sets s.t.* $K_1 = \{(3,64),(7,64),(11,128)\}$*,* $K_2 = \{(3,64),(5,64),(7,64),(11,128),(13,128),(15,128)\}$ *and* $K_3 = \{(2,64),(3,64),\ldots,(14,128),(15,128)\}$*. For example,* K_1 *consists of filter sizes* $f_k \in \{3,7,11\}$ *which are used to extract local contexts given the window size of "biological word". Subsequent to the multi-convolutional layer, we performed global max-pooling to select the maximum value from each feature map which are then concatenated together to generate the local contextual feature vector. From Figure 1, we observe that CNN configuration* K_3 *in combination with* $\mathbf{f}_c = 64$ *neurons has best predictive accuracy when compared with other parameter settings in case of* 1^{st} *experimental setting. This best DeepSol model is hereby referred as DeepSol S1.*

Model	**CNN Config**	**Layers**	**f_c Neurons**	**Test Accuracy** (Valid Accuracy)	**Precision**(Per class)	**Recall**(Per class)	**F-Score**(Per class)
Model 1	$2:3:\ldots:15$	CNN:1,FC:1	64	0.76(0.79)	0.78(0.71,0.85)	0.76(0.89,0.64)	0.76(0.79,0.73)
Model 2	$2:3:\ldots:15$	CNN:1,FC:1	128	0.76(0.79)	0.77(0.71,0.82)	0.76(0.86,0.65)	0.75(0.78,0.73)
Model 3	$2:3:\ldots:15$	CNN:1,FC:1	256	0.75(0.79)	0.77(0.70,0.85)	0.75(0.89,0.61)	0.75(0.78,0.71)
Model 4	$2:3:\ldots:15$	CNN:1,FC:2	(1,256),(2,128)	0.76(0.79)	0.77(0.72,0.83)	0.76(0.86,0.66)	0.76(0.78,0.73)
Model 5	$2:3:\ldots:15$	CNN:1,FC:2	(1,128),(2,64)	0.75(0.79)	0.77(0.70,0.84)	0.75(0.88,0.63)	0.75(0.78,0.72)
Model 6	$2:3:\ldots:15$	CNN:1,FC:2	(1,256),(2,64)	0.75(0.80)	0.77(0.71,0.84)	0.75(0.88,0.63)	0.75(0.78,0.72)
* Model 7	$3:5:\ldots:15$	CNN:1,FC:1	64	**0.77**(0.79)	**0.78**(0.72,0.84)	**0.77**(0.88,0.66)	**0.76**(0.79,0.74)
Model 8	$3:5:\ldots:15$	CNN:1,FC:1	128	0.76(0.79)	0.77(0.71,0.83)	0.76(0.86,0.65)	0.75(0.78,0.73)
Model 9	$3:5:\ldots:15$	CNN:1,FC:1	256	0.75(0.79)	0.77(0.70,0.84)	0.75(0.88,0.62)	0.75(0.78,0.71)
Model 10	$3:5:\ldots:15$	CNN:1,FC:2	(1,256),(2,64)	0.76(0.79)	0.77(0.72,0.82)	0.76(0.85,0.66)	0.76 (0.78,0.73)
Model 11	$3:5:\ldots:15$	CNN:1,FC:2	(1,128),(2,64)	0.75(0.79)	0.77(0.70,0.84)	0.75(0.88,0.62)	0.75(0.78,0.72)
Model 12	$3:5:\ldots:15$	CNN:1,FC:2	(1,256),(2,128)	0.76(0.79)	0.77(0.71,0.84)	0.76(0.87,0.64)	0.76(0.78,0.73)

Table 1: *Comprehensive comparison of DeepSol models for different values of parameters in the* 2^{nd} *setting. Here CNN:1 stands for* 1 *convolutional layer, FC:1 stands for* 1 *fully connected layer, (1,256),(2,64) stand for* 1^{st} *fully connected layer has 256 neurons and* 2^{nd} *fully connected layer has 64 neurons. Here* $*$ *represents the parameter setting for the most accurate DeepSol model on the independent test set which is further referred as DeepSol S2.*

Model	**CNN Config**	**Bio FC Layers, Neurons**	**Layers**	**f_c neurons**	**Test Accuracy (Valid Accuracy)**	**Precision (Per class)**	**Recall (Per class)**	**F-Score (Per class)**
Model 1	$2:3:\ldots:15$	1,64	CNN:1,FC:1	64	0.75(0.80)	0.77(0.70,0.85)	0.75(0.89,0.61)	0.75(0.78,0.71)
Model 2	$2:3:\ldots:15$	1,64	CNN:1,FC:1	128	0.75(0.79)	0.77(0.70,0.84)	0.75(0.89,0.61)	0.74(0.78,0.71)
Model 3	$2:3:\ldots:15$	1,64	CNN:1,FC:1	256	0.76(0.80)	0.78(0.71,0.84)	0.76(0.88,0.64)	0.76(0.79,0.73)
Model 4	$2:3:\ldots:15$	1,128	CNN:1,FC:1	64	0.76(0.80)	0.77(0.71,0.84)	0.76(0.87,0.64)	0.75(0.78,0.72)
Model 5	$2:3:\ldots:15$	1,128	CNN:1,FC:1	128	0.76(0.79)	0.77(0.71,0.82)	0.76(0.86,0.66)	0.75(0.78,0.73)
*Model 6	$2:3:\ldots:15$	1,256	CNN:1,FC:1	64	**0.77**(0.80)	**0.77**(0.73,0.81)	**0.76**(0.84,0.69)	**0.76**(0.78,0.75)
Model 7	$2:3:\ldots:15$	1,256	CNN:1,FC:1	128	0.76(0.79)	0.77(0.71,0.83)	0.76(0.86,0.65)	0.75(0.78,0.73)
Model 8	$2:3:\ldots:15$	1,128	CNN:1,FC:1	256	0.76(0.79)	0.77(0.70,0.84)	0.76(0.89,0.62)	0.75(0.78,0.72)
Model 9	$2:3:\ldots:15$	1,256	CNN:1,FC:1	256	0.76(0.79)	0.77(0.71,0.84)	0.76(0.88,0.63)	0.75(0.78,0.72)
Model 10	$3:5:\ldots:15$	1,64	CNN:1,FC:1	64	0.75(0.79)	0.77(0.70,0.84)	0.75(0.88,0.63)	0.75(0.78,0.72)
Model 11	$3:5:\ldots:15$	1,64	CNN:1,FC:1	128	0.76(0.80)	0.77(0.71,0.83)	0.76(0.87,0.64)	0.75(0.78,0.72)
Model 12	$3:5:\ldots:15$	1,64	CNN:1,FC:1	256	0.76(0.80)	0.77(0.71,0.83)	0.75(0.87,0.64)	0.75(0.78,0.72)
Model 13	$3:5:\ldots:15$	1,128	CNN:1,FC:1	64	0.76(0.79)	0.77(0.70,0.84)	0.75(0.88,0.63)	0.75(0.78,0.72)
Model 14	$3:5:\ldots:15$	1,128	CNN:1,FC:1	128	0.75(0.79)	0.77(0.70,0.85)	0.75(0.89,0.62)	0.75(0.78,0.72)
Model 15	$3:5:\ldots:15$	1,128	CNN:1,FC:1	256	0.76(0.79)	0.77(0.71,0.84)	0.76(0.87,0.64)	0.75(0.78,0.73)
Model 16	$3:5:\ldots:15$	1,256	CNN:1,FC:1	64	0.76(0.79)	0.77(0.71,0.84)	0.76(0.88,0.63)	0.75(0.78,0.72)

Table 2: *Comprehensive comparison of DeepSol models for different values of parameters in the* 3^{rd} *setting. Here CNN:1 stands for* 1 *convolutional layer, FC:1 stands for fully connected layer and* $*$ *represents the parameter setting for the most accurate DeepSol model on the independent test set which is further referred as DeepSol S3. Here Bio FC layers and neurons represent the fully connected layer on top of* 57 *additional features and number of hidden neurons for this layer.*

2 Multi-layered Multi-CNN Filter based DeepSol Models

We built additional architectures using multi-layered multi-filtered convolutional features and also with supplementary biological features. The performance of models obtained from just using multi-layered multi-filtered convolutional features from the raw protein sequence is depicted in Table 3. For models with additional biological features, the results are shown in Table 4.

References

[1] Pawel Smialowski, Gero Doose, Phillipp Torkler, Stefanie Kaufmann, and Dmitrij Frishman. PROSO II - a new method for protein solubility prediction. *FEBS Journal*, 279(12):2192–2200, jun 2012.

Model	**CNN Config**	**Layers**	**f_c Neurons**	**Test Accuracy** (Valid Accuracy)	**Precision**(Per class)	**Recall**(Per class)	**F-Score**(Per class)
Model 1	$(1,3:7:11)$, $(2,3:7:11)$, $(3,3:7:11)$	CNN:3,FC:1	256	0.68(0.72)	0.72(0.62,0.83)	0.68(0.91,0.44)	0.66(0.74,0.58)
Model 2	$(1,3:5:\ldots:13:15)$, $(2,3:7:11)$	CNN:2,FC:1	128	0.69(0.73)	0.73(0.63,0.84)	0.69(0.91,0.47)	0.67(0.75,0.60)
Model 3	$(1,3:5:\ldots:13:15)$, $(2,3:7:11:15)$	CNN:2,FC:1	256	0.69(0.74)	0.73(0.64,0.81)	0.69(0.88,0.50)	0.68(0.74,0.62)
Model 4	$(1,3:5:\ldots:13:15)$, $(2,3:7:11:15)$	CNN:2,FC:1	128	0.70(0.74)	0.71(0.67,0.76)	0.70(0.81,0.59)	0.70(0.73,0.67)
Model 5	$(1,3:5:\ldots:13:15)$, $(2,3:7:11:15)$	CNN:2,FC:1	64	0.71(0.73)	0.73(0.66,0.79)	0.71(0.85,0.57)	0.70(0.75,0.66)

Table 3: *Comparison of DeepSol models for different values of parameters in the* 1^{st} *setting. Here CNN:2 stands for* 2 *convolutional layers, FC:1 stands for fully connected layer,* $(1,3:5:\ldots,13:15)$ *represents* 1^{st} *convolution layer with filter size* $f_k \in \{3,5,\ldots,13,15\}$ *and* $(2,3:7:11:15)$ *represents* 2^{nd} *convolution layer with filter size* $f_k \in \{3,7,11,15\}$*. We observe that predictive performance of these models are lower than DeepSol S2 (see Table 1).*

Model	**CNN Config**	**Bio FC Layers, Neurons**	**Layers**	**f_c neurons**	**Test Accuracy (Valid Accuracy)**	**Precision (Per class)**	**Recall (Per class)**	**F-Score (Per class)**
Model 1	$2:3:\ldots:14:15$	1,64	CNN:1,FC:2	(1,128),(2,64)	0.76(0.79)	0.77(0.71,0.83)	0.76(0.87,0.65)	0.76(0.78,0.73)
Model 2	$2:3:\ldots:14:15$	1,64	CNN:1,FC:2	(1,256),(2,128)	0.76(0.79)	0.77(0.72,0.83)	0.76(0.86,0.66)	0.76(0.78,0.73)
Model 3	$2:3:\ldots:14:15$	1,64	CNN:1,FC:2	(1,256),(2,64)	0.75(0.80)	0.77(0.70,0.84)	0.75(0.88,0.63)	0.75(0.78,0.72)
Model 4	$2:3:\ldots:14:15$	2,(1,64),(2,64)	CNN:1,FC:1	64	0.75(0.79)	0.76(0.71,0.82)	0.75(0.86,0.65)	0.75(0.78,0.72)
* Model 5	$2:3:\ldots:14:15$	2,(1,128),(2,64)	CNN:1,FC:1	64	0.76(0.79)	0.78(0.71,0.85)	0.76(0.89,0.63)	0.76(0.79,0.72)
Model 6	$2:3:\ldots:14:15$	2,(1,256),(2,64)	CNN:1,FC:1	64	0.75(0.79)	0.77(0.69,0.85)	0.75(0.90,0.60)	0.74(0.78,0.71)
Model 7	$2:3:\ldots:14:15$	2,(1,64),(2,64)	CNN:1,FC:2	(1,64),(2,64)	0.75(0.79)	0.76(0.70,0.83)	0.75(0.88,0.62)	0.74(0.78,0.71)
Model 8	$2:3:\ldots:14:15$	2,(1,128),(2,64)	CNN:1,FC:2	(1,64),(2,64)	0.76(0.79)	0.77(0.72,0.83)	0.76(0.87,0.65)	0.76(0.78,0.73)
Model 9	$2:3:\ldots:14:15$	2,(1,256),(2,64)	CNN:1,FC:2	(1,64),(2,64)	0.75(0.79)	0.76(0.72,0.80)	0.75(0.83,0.68)	0.75(0.77,0.73)
Model 10	$2:3:\ldots:14:15$	3,(1,256),(2,128),(3,64)	CNN:1,FC:2	(1,64),(2,64)	0.75(0.79)	0.77(0.70,0.84)	0.75(0.88,0.62)	0.74(0.78,0.71)

Table 4: *Comprehensive comparison of additional DeepSol models for different values of parameters in the* 3^{rd} *setting. Here CNN:1 stands for* 1 *convolutional layer, FC:1 stands for fully connected layer,* $(1,128)$ *stands for* 1^{st} *layer with* 128 *neurons and* $*$ *represents the parameter setting for the best model among these models. Here Bio FC layers and neurons represent the fully connected layer on top of* 57 *additional features and number of hidden neurons for this layer.*

Methods	**Accuracy**	**MCC**	**Selectivity (Soluble)**	**Selectivity (Insoluble)**	**Sensitivity (Soluble)**	**Sensitivity (Insoluble)**	**Gain (Soluble)**	**Gain (Insoluble)**
PROSO II	0.71	0.42	0.69	0.73	**0.75**	0.67	1.39	1.46
DeepSol S1	0.76	0.51	0.69	0.82	0.77	0.75	1.38	1.64s
DeepSol S2	0.80	0.58	**0.80**	0.80	0.69	**0.88**	1.59	1.6
DeepSol S3	**0.82**	**0.62**	**0.80**	**0.83**	**0.75**	0.86	**1.60**	**1.65**

Table 5: *Comparison of the mean prediction performance of DeepSol S1, S2 and S3 with that of PROSO II using* 10*-fold cross-validation on the training set comprising* 69,420 *protein sequences. PROSO II performance values obtained from [1]. The best results are highlighted in bold.*

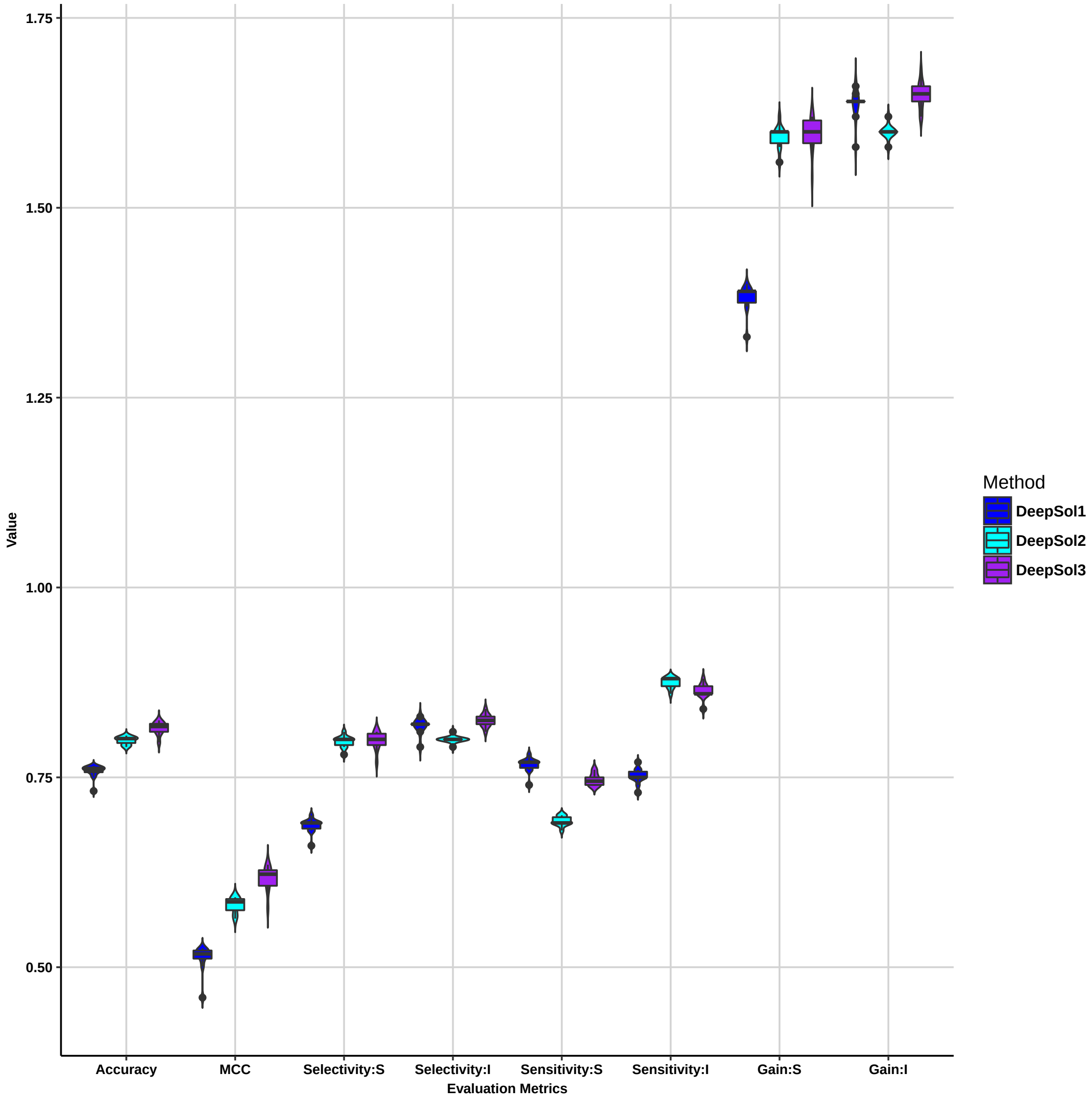

Figure 2: *Comparison of DeepSol models w.r.t. various evaluation metrics when performing* 10*-fold cross-validation on the training set. Here the annotation ":S" is used for soluble class and the annotation ":I" is used for insoluble class along the x-axis. The boxplots highlight that the variance in the results of the DeepSol models are pretty low for various evaluation metrics. DeepSol S3 clearly outperforms DeepSol S1 and DeepSol S2 when performing* 10*-fold cross-validation w.r.t. evaluation metrics like Accuracy, MCC, Selectivity and Gain.*