1. **F-measure for various classifiers**

Figure S1 shows the average F-measure for various classifiers on the test data.



Figure S1. The average F-measure for various classifiers on the test data.

1. **LASSO**

We used mean squared error (MSE) as an objective function and then conducted 5-fold cross-validation to estimate the shrinkage parameter. The following Figure S3 showed the relationship between the shrinkage parameter λ and the cross validated MSE.

In Figure S2, we reported two values for λ. First, λ = 3.55e-04 (the right vertical line, green) minimizes the cross-validated MSE. Second, λ = 0.0134 (the left vertical line, blue) is the largest shrinkage value, whose cross-validated MSE is within one standard error of the minimum MSE. In our revised manuscript, we set λ as 0.0134 to conduct the feature importance analysis.

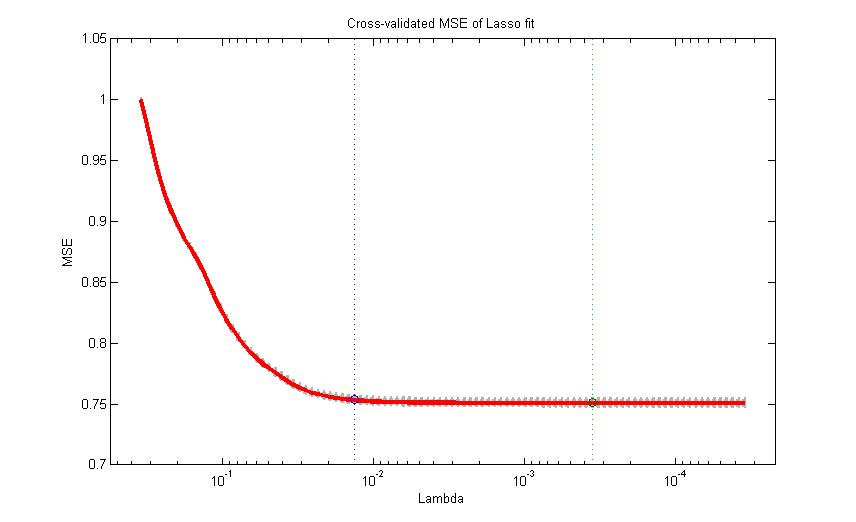


Figure S2. The relationship between the shrinkage parameter λ and the cross validated MSE.

Feature normalization would be critical to evaluate the feature importance using the LASSO coefficients.

In Table S1, we compared the feature rankings based on 3 different types of feature normalization. These 3 types of feature normalization are as follows. First, “Non-Norm” in second column is using the raw data without any feature normalization. Second, “Norm[0, 1]” in third column indicates that we linearly normalize the feature values to the range [0, 1]. Third, “Norm\_µ0σ1” means that the normalized features have a mean 0 and a standard deviation 1.

Table S1. The LASSO coefficients from different types of feature normalization.

|  |  |  |  |
| --- | --- | --- | --- |
| **Features** | **Non-Norm** | **Norm[0, 1]** | **Norm\_µ0σ1** |
| FSweight | -2.0992 | -1.9375 | -0.4117 |
| GEcorrelation | 0 | 0 | 0 |
| GOsim | 0 | 0 | 0 |
| C2Ssim | 4.5767 | 4.3879 | 0.8975 |
| GOsimMF | -0.2585 | -0.2684 | -0.0507 |
| GOsimCC | 0.6896 | 0.6905 | 0.1352 |
| ComReal | 0 | 0 | 0 |
| ComPredict | 0 | 0 | 0 |
| DegreeA | 2.5776 | 1.9504 | 0.5055 |
| DegreeB | 2.2923 | 2.3214 | 0.4495 |
| ParalogA | -0.8597 | -0.8725 | -0.1686 |
| ParalogB | -0.7402 | -0.755 | -0.1452 |
| ParalogAB | 0.1902 | 0.2089 | 0.0373 |
| Pathway | 0 | 0 | 0 |
| DomainA | 0.4319 | 0.4427 | 0.0847 |
| DomainB | 0.8329 | 0.8452 | 0.1633 |
| DomainAB | 0 | 0 | 0 |

It is very interesting that the above feature rankings based on the 3 different types of feature normalization are highly correlated. For example, the rankings based on “Non-Norm” and “Norm\_µ0σ1” are exactly the same. Meanwhile, the spearman correlation between “Non-Norm” and “Norm[0, 1]” is 0. 997 with a p-value 1.33e-18.

1. **Weighting schemes for MetaSL**

If we directly use the AUC scores of individual classifiers as their weights, and the AUC of MetaSL is 0.865. Next, we show the results using two other weighting schemes for MetaSL.

* 1. Linear Transform

We also tried different scaling for the weights of classiﬁers in Equation 1 , i.e., *wi = α\*(xi −min)/(max−min) + (1 – α).* Here, *xi* is the AUC of *ith* classifier and *wi* is its weight. In addition, *max* and *min* are the maximum and minimum AUC over all individual classifiers. For *α* = 0.5, 0.6, 0.7, 0.8 and 0.9, the average AUC values for our MetaSL are 0.869, 0.869, 0.869, 0.870 and 0.871, respectively. These results show that the performance of MetaSL is robust to diﬀerent scaling for the weights of classifiers.

(1)

3.2 Exponential Transform

We use the exponential transform in the equation 2. Here, *n* is the number of classifiers, *wi* is the original AUC score (No Transform) or weight from the linear transform *(xi−min)/(max−min)*, and *ewi* is the weight from exponential transform. As shown in Table S3, the AUC scores of MetaSL for these two cases are 0.865 and 0.870, respectively.

(2)

Table S2 summarizes the AUC of MetaSL using various methods for weighting the individual classifiers. We can observe that Linear transform (α = 0.9 or 1), or Exponential transform together with linear transform will achieve higher AUC than without any transform for weights.

Table S2. AUC of MetaSL using various scaling for weights

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Linear transform**  *wi = α\*(xi −min)/(max−min) + (1 – α)* | | | | | | **No**  **Transform** | **Exponential Transform** | |
| *α* = 0.5 | *α* = 0.6 | *α* = 0.7 | *α* = 0.8 | *α* = 0.9 | *α* = 1 |  | Linear | No |
| **AUC** | 0.869 | 0.869 | 0.869 | 0.870 | 0.871 | 0.871 | 0.865 | 0.870 | 0.865 |