**Meta-DPI, a Metamethod for Predicting Protein Interfaces**

***M. Walder1, E.Edelstein1, S. Lazarev1, M. Carroll1, R.Viswanathan1,\*,E. Fajaro2, and A.Fiser2***

*1Department of Chemistry, Yeshiva College, Yeshiva University, New York, NY 10033*

*2Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, NY 10461*

**Introduction:**

Proteins have a diverse range of functions that are fundamental to the processes of life, and they perform those functions through interactions with other proteins, DNA, RNA, or small molecules [1] They catalyze biochemical reactions, function as hormones that maintain homeostasis, transport molecules around the body and through cell membranes, and are an essential part of the immune system’s response to fight infection. An organism’s interactome conveys a systems-level lens of the network of all protein-protein interactions (PPIs) that occur within it. Knowledge of a protein’s interaction partners is key to understanding its function and for precise characterization of its broader role in the interactome. To elucidate the molecular mechanism by which a protein interacts with its partners, the interfacial residues must be determined [2]. Additionally, the central role that PPIs play in the progression of many disease states makes the determination of interfacial residues a critical step in the process of developing a drug that targets a PPI.

The primary experimental methods for determining the structure of protein complexes and their interfacial residues are X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy. Due to the low-throughput and costly nature of these experimental approaches, computational prediction methods are employed to streamline the process of identifying the interfacial residues of PPIs. One class of computational methods are intrinsic-based approaches that train machine learning algorithms on a dataset of experimentally determined complex structures to create a model that relates sequence and structural features with the likelihood for residues to be at the interface. Sequence features include hydrophobicity, amino acid interface propensity, physico-chemical properties, and evolutionary conservation, and structural features include secondary structure, solvent-accessible surface area, and geometric shape. While intrinsic-based methods have been steadily enhanced over the past 20 years, their future improvement is limited because further combination of existing features and classifiers has little impact on performance. The other class of computational methods are template-based approaches, which identify a query protein’s homologues or structural neighbors with a known complex structure and map the homologues’ or neighbor’s interface residues onto the query protein. The drawback of template-based methods is that their effectiveness is highly contingent on the query protein having homologues or structural neighbors that have had their complex structure determined [1].

To overcome the limitations of intrinsic and template-based methods, metamethods that integrate orthogonal predictors can be developed to enhance prediction performance. Meta-PPISP is one such metamethod that combined the predictors cons-PPISP [3], Promate [4], and PINUP [5] through linear regression analysis [6]. The construction of meta-PPISP was not ideal because it only combined complementary intrinsic-based approaches, and it did not combine a template-based approach with an intrinsic-based approach. Additionally, it employed linear regression analysis for method combination, instead of using logistic regression analysis, which is more effective for discrete categorical data like a residue’s interfacial score. Here, we describe the development of meta-DPI, a metamethod that integrates orthologous approaches, PredUs 2.0 (template-based) [2] and ISPRED4 (intrinsic-based) [7], from the two predictor classes with the recently developed docking-based approach DockPred [8]. The models employed for method combination in the two versions of meta-DPI were logistic regression and random forest, which are more effective than linear regression for discrete categorical data like interfacial value of amino acid residues (interface or non-interface).

**Materials and Methods:**

We have used a total of 233 protein structures, taken from both the Docking Benchmark and NOX[9] database, with known complex structures available from the Protein Data Bank (PDB). The interface residues have been determined by using the CSU program to find the contacts between structural units.[10]. Any hetero atom of a residue in a protein that is within 4.0 A of any other hetero atom of the partner protein in the complex is considered an interface residue. This same dataset has been used previously to study the effectiveness of a docking based method to identify protein interfaces.

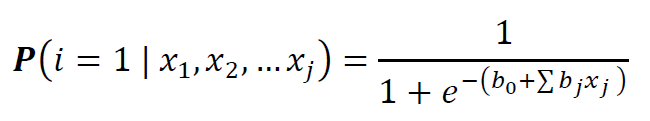
**Interface Prediction Method:**

In our approach, the meta method, meta-DPI, combines three orthogonal methods, one template based method, PredUs2.0[2], one docking based method, DOCKPRED[8], and one intrinsic method, ISPRED4[7]. Meta-DPI combines these three interface classifiers by using a logistic regression method as well as random forest method.

**Meta-DPI by Logistic Regression:**

Using the three different classifiers, DOCKPRED, PRedUS2.0 and ISPRED4, a normalized score between 0 and 1 is calculated for each residue for every protein in the database of 233 proteins. This normalized score is the likelihood of the residue being at the interface as determined by each of the three classifiers. The prediction scores obtained from these three classifiers are then combined using logistic regression to provide the prediction scores for the meta-method, meta-DPI. A logistic regression model was chosen for the development of meta-DPI since there are only two possibilities for each residue, interface (1) or non-interface (0) and hence it is not a continuous variable. We, therefore, expect the logistic regression model to fit the discrete categorical data better than a linear regression model.

Using xi (where i=1 for DOCKPRED, i = 2 for PredUs2.0 and i = 3 for ISPRED4), the interface likelihood values obtained by the three classifiers, logistic regression method uses the function,



Where ***P***(i=1|x1,x2,…..xj) is the probability that i = 1, that is, a given residue is an interface residue, given the values of xi. In order to find the best fitting parameters, b0, b1, b2, and b3, a subset of the database was used as the training set (as described below in the Cross Validation section) and the coefficients bi (i=0,1,2,3) were obtained by Maximum Likelihood.

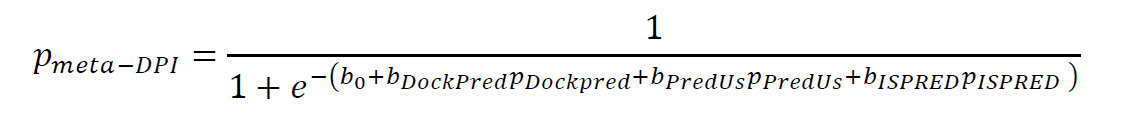
**Meta-DPI by Random Forest (RF):**

The normalized scores obtained from the three classifiers were also combined using the random forest method. At each node of the tree, the classifiers and the cutoff values, for each classifier and each level of the tree, are chosen to optimize the results. The parameters representing the ensemble of trees in the forest, the maximum number of levels for each tree, and a tree pruning parameter, α, which chooses the subtree that minimizes the cost complexity measure, were all optimized to find the best fitting model with the three classifiers. The values for the optimized parameters are shown in Table 1 and these were optimized to yield the best values for AUC-ROC, calculated as described below. For all the RF calculations, the optimal values of 100 trees, 10 levels and a pruning parameter of 2.5X10-5 were used. The random forest model was trained using a subset of the database as the training set, as described below in the Cross-Validation section.

**Leave One Out and K-fold Cross Validation:**

In order to develop the parameters for the meta method, a training set is chosen. We first used a Leave One Out Cross-Validation (LOOC) where all but one protein from the database was used as the training set and the parameters derived were used to predict the results for the one protein that was left out of the training set. This process was repeated by leaving a different protein out of the training set each time.

For K-fold cross validation, the database was randomly divided into K sets, 5 sets for five-fold cross validation and 10 sets for 10-fold cross validation. If the total number of proteins in the database is not exactly divisible by 5 or 10, the additional proteins were added to some of the K-sets. Using 4 of the 5 sets as training (or 9 of the 10 sets for 10-fold cross-validation), the parameters bi, in the logistic regression model, were calculated and then these parameters were used to make the prediction on the fifth (tenth) set. The process was repeated by choosing a different set of 4 (or 9) until the probability of a residue being at the interface is predicted for all the proteins in the complete data set. By applying the parameters obtained from the training, a value for pmeta-DPI, between 0 and 1, was calculated for every residue in each of these proteins in the database using:



A similar LOOC and K-fold cross validation was used for training the random forest model. Once the trees are trained using the training set, the random forest model classifies the residues in the test set to one of the terminal nodes (leaves) in each tree in the forest. Based on the results from the training set, the probability of being an interface residue is calculated for each terminal node in each tree. For example, if a terminal node in a tree in the random forest contains a total of 100 residues from the training set, and if 85 of those residues are non-interface and 15 are interface, then the probability of being an interface residue for that terminal node is 0.15. A similar probability value is calculated for each terminal node for every tree in the random forest. For the test set, the probability of being an interface residue is calculated as the average probability of all the terminal nodes in the forest into which the test residue is classified.

**Single Threshold Evaluation Metrics of meta-DPI:**

The meta-DPI method, either through logistic regression or random forest, yields a value for the probability of being an interface residue, p, for each residue of every protein in the database. In order to then classify each residue as interface or non-interface based on the value of p, one approach is to select the residues with the top N values of p as interface. This approach chooses a single threshold value, N, for each query protein. More recently[2], a dynamic cutoff method was proposed to find N for each query protein according to the following equation:

where R is the number of surface exposed residues for each query protein. Using this threshold value, the elements of the Confusion matrix, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) can be determined. F1 and MCC scores are then calculated using the equations below:

Confusion Matrix:

Precision =

Recall =

F1 Score =

MCC =

**Threshold Free Evaluation Metrics - AUC under ROC and PR Curves:**

Receiver Operator Characteristic (ROC) curves were calculated by plotting the true positive rate (TPR) vs the false positive rate (FPR) for different threshold values for p, ranging from 0 to 1 at intervals of 0.01. The area under the curve (AUC-ROC) was calculated.

A plot of Precision vs Recall (PR) and the area under the curve (AUC-PR) is considered a better measure of performance, particularly for unbalanced dataset. This is the case while identifying interface residues since the number of interface residues is significantly smaller than the non-interface residues.

**Results and Discussion:**

We explored the possibility that a meta method, meta-DPI, that combined three orthogonal methods(classifiers) that rely on different structural and sequence properties of a protein would be a better predictor of interface residues. This would enable a better predictor method based on the currently existing methods. We used three methods, one of which is template based, PredUS, that relies on structural similarities of the query protein and the availability of complex structures for analogues of the query protein, another that is intrinsic based, ISPRED4, that relies on the properties of the query protein, and a third, DOCKPRED, that relies on docking the query protein to a series of non-cognate ligands and is not based on information on structurally similar proteins. DOCKPRED has the advantage that it does not rely on the information of structurally similar proteins with known interface residues, Results from these two classifiers were then combined with another predictor, ISPRED4, that relies on the sequence of residues in the protein and properties of these residues, like hydrophobicity, solvent accessibility, interface propensity, etc.

We assess the performance of the meta-DPI method by comparing the F1-scores and MCC values obtained by using a dynamic threshold (see Methods) for each query in the database. These results are shown in Table 2. The error bars correspond to a 95% confidence interval based on a normal distribution. The values in Table 2 for Logistic regression and RF are based on the LOOC for the training. The F1 scores range from 0.347 to 0.367 for the individual classifiers. For the meta-DPI method, the F1 scores are higher, yielding a value of 0.486 for logistic regression and 0.426 for the random forest method. As seen from Table 2, meta methods are a significant improvement over the individual methods. While the F1 score summarizes the performance of a method in generating the positive class, the MCC score is large only if the method performs well in predicting both the positive and negative classes. So, the MCC score is more informative than the F1 score on the performance of these methods. The MCC scores range from 0.292 to 0.314 for the individual methods while the meta methods yield higher scores of 0.449 and 0.381. We have calculated the error bars for both the F1 score and MCC at the 95% confidence interval. We have also used the non-parametric and distribution free Kolmogorov-Smirnov test at the 95% level to show that the differences in F1 score and MCC obtained by the different methods are significant. These results are shown in Table 3. All the p-values calculated using this test are below 0.05 thereby indicating that all the differences in F1 and MCC values calculated using the different methods are statistically significant. By both the F1 and MCC measure, the meta methods perform significantly better compared to the three individual methods.

The ROC curves for the three individual methods are compared with the meta-method in Figure 1. The area under the ROC curve is shown in Table 4. The AUC-ROC for the three individual methods range from 0.693 to 0.810 while that for the meta methods are 0.918 and 0.874 using LOOC. By this measure as well, the meta methods perform much better than the individual methods. The PR curve are a better assessment of performance, especially for unbalanced sets where the number of positive class (number of interface residues) is much smaller than the negative class (non-interface residues). For a random unskilled classifier, the area under the PR curve would equal . The PR curves for the individual methods are compared with the meta methods in Figure 2. The AUC-PR values as shown in Table 4 range from 0.244 to 0.333 for the individual methods while for the meta methods it is 0583 and 0.393 using LOOC. The PR curves also yield a much better performance for the meta methods compared to the individual methods.

In order to assess the statistical significance of the differences in the AUC-ROC values for the different methods, we used the StAR[11] software. The StAR software uses the chi-squared distribution to test the null hypothesis that there is no difference between the AUC-ROC curves originating from the different methods. The difference between any two methods is assessed at a significance level of 0.05. The results from the StAR statistical analysis is shown in Table 4. The elements above the diagonal are the differences in the AUC-ROC values between the different methods. The elements below the diagonal list the p-values that list the significance of these difference in the AUC values. The p- values in the Table 5 are all smaller than 0.05, except for the differences between DOCKPRED and ISPRED, indicating that the differences in AUC values between all the other methods are statistically significant. The meta-DPI, therefore, performs better than any of the individual methods by all the statistical measures used.

We tested the effect of the size of the training set on the performance of the meta-DPI methods. The results in Tables 2 and 4 are based on the LOOC. The results for F1 and MCC scores from five-fold and ten-fold cross-validation are shown in Table 6. For the five-fold cross-validation, we have five sets of F1 and MCC values calculated. Each set of F1 and MCC scores correspond to 44 proteins from our database which constitute the test set while the remaining query proteins in the database form the training set. The average (and standard deviation) for F1 and MCC scores over the five sets is 0.417±0.039, 0.373±0.037, respectively, for the logistic regression meta method and 0.422±0.044, 0.378±0.049, respectively, for the random forest meta method. For the ten-fold cross-validation, the average values for F1 and MCC are 0.418±0.044 and 0.375±0.046 respectively for logistic regression and 0.425±0.046 and 0.383±0.053 respectively for random forest. Using the five-fold cross-validation, the estimated errors for the AUC-ROC are 0.856±0.026 and 0.872±0.25 respectively for logistic regression and random forest. Using five-fold cross-validation, the estimated errors for the AUC-PR are 0.389±0.059 and 0.395±0.059 respectively for logistic regression and random forest meta methods.

The results for a 5-fold and 10-fold cross-validation methods for the F1, MCC, AUC-ROC and AUC-PR along with the standard deviations between the k-sets are summarized in Table 7. Comparing the results from Tables 2 and 4 with Table 7, the LOOC with a much larger training set certainly performs better than the 5-fold and 10-fold cross-validation methods. But the LOOC is time consuming since the calculation needs to be repeated as many times as there are query proteins. A five-fold cross-validation uses 44 query proteins in a test set while the other query proteins from the database are used as the training set. A ten-fold cross-validation uses 22 query proteins in the test set while the other query proteins from the database constitute the training set. The results from Table 6 show that the results for F1, MCC, AUC-ROC and AUC-PR are almost identical for both the 5-fold and 10-fold cross-validation. So, a 5-fold cross-validation yields reliable results with a significant saving in computational effort.

**Comparison with other meta-methods:**

Other meta methods have recently been developed and tested. Meta-PPISP[6] uses three individual methods that are based on sequence conservation and other attributes such as solvent accessibility, secondary structure, sequence profiles. The three methods are combined using linear regression. We compare the results of our meta-methods with those from meta-PPISP for the same database. The results are shown in Table 8. Another meta method, VORFFIP, recently developed by Segura et al[12] uses a random forest method to combine four sets of residue features, structural, sequence conservation, crystallographic B-factor, and energy terms. We compare results obtained by meta-DPI with VORFFIP for the same database of query proteins and the results are included in Table 8. We compare the significance of the differences in F1 and MCC scores obtained by the three meta-methods using the non-parametric and distribution free Kolmogorov-Smirnov test at the 95% level. The calculated p-values are shown in Table 9. All the p-values are less than 0.05 indicating that the differences between these methods are statistically significant. Table 10 shows the results of StAR statistical analysis to assess the significance of the AUC-ROC values between the four different meta-methods at a 95% confidence level. As seen in Table 10, the differences in AUC-ROC values between all of these methods is statistically significant. Figure 3 and 4 compare the ROC and PR curves for meta-DPI with meta-PPISP and VORFFIP. The AUC values for ROC and PR are compared in Table 8.

The current meta method performs much better than meta-PPISP. While VORFFIP performs very close to the current method, the results from meta-DPI using logistic regression with the three classifiers proposed in the current meta method yields significantly better results by all statistical four statistical measures, F1, MCC, AUC-ROC and AUC-PR, used to assess the performance of the methods.

**References:**

1. Esmaielbeiki, R., et al., *Progress and challenges in predicting protein interfaces.* Brief Bioinform, 2016. **17**(1): p. 117-31.

2. Zhang, Q.C., et al., *PredUs: a web server for predicting protein interfaces using structural neighbors.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W283-7.

3. Chen, H. and H.X. Zhou, *Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data.* Proteins, 2005. **61**(1): p. 21-35.

4. Neuvirth, H., R. Raz, and G. Schreiber, *ProMate: a structure based prediction program to identify the location of protein-protein binding sites.* J Mol Biol, 2004. **338**(1): p. 181-99.

5. Liang, S., et al., *Protein binding site prediction using an empirical scoring function.* Nucleic Acids Res, 2006. **34**(13): p. 3698-707.

6. Qin, S. and H.X. Zhou, *meta-PPISP: a meta web server for protein-protein interaction site prediction.* Bioinformatics, 2007. **23**(24): p. 3386-7.

7. Savojardo, C., et al., *ISPRED4: interaction sites PREDiction in protein structures with a refining grammar model.* Bioinformatics, 2017. **33**(11): p. 1656-1663.

8. Viswanathan, R., et al., *Protein-protein binding supersites.* PLoS Comput Biol, 2019. **15**(1): p. e1006704.

9. Zhu, H., et al., *NOXclass: prediction of protein-protein interaction types.* BMC Bioinformatics, 2006. **7**: p. 27.

10. Sobolev, V., et al., *Automated analysis of interatomic contacts in proteins.* Bioinformatics, 1999. **15**(4): p. 327-32.

11. Vergara, I.A., et al., *StAR: a simple tool for the statistical comparison of ROC curves.* BMC Bioinformatics, 2008. **9**: p. 265.

12. Segura, J., P.F. Jones, and N. Fernandez-Fuentes, *Improving the prediction of protein binding sites by combining heterogeneous data and Voronoi diagrams.* BMC Bioinformatics, 2011. **12**: p. 352.

**Table 1: Optimized Random Forest Parameters**

|  |  |  |  |
| --- | --- | --- | --- |
| **Number of Trees** | **Number of Levels** | **Alpha** | **AUC-ROC** |
| **10** | **5** | **0** | **0.867** |
| **50** | **5** | **0** | **0.868** |
| **100** | **5** | **0** | **0.870** |
| **200** | **5** | **0** | **0.870** |
| **100** | **10** | **0** | **0.873** |
| **100** | **15** | **0** | **0.853** |
| **100** | **25** | **0** | **0.793** |
| **100** | **10** | **2.5X10-5** | **0.873** |
| **100** | **10** | **5.0X10-5** | **0.873** |
| **100** | **10** | **7.5X10-5** | **0.872** |
| **100** | **10** | **1.0X10-4** | **0.869** |
| **100** | **10** | **1.5X10-4** | **0.869** |
| **100** | **10** | **2.0X10-4** | **0.686** |

**Table 2: Comparison of F1 and MCC scores Using LOOC**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Predus** | **Ispred** | **Dockpred** | **Logreg** | **RF** |
|  |  |  |  |  |  |
| **F1** | **0.345±0.022** | **0.372±0.019** | **0.367±0.027** | **0.485±0.019** | **0.423±0.021** |
| **MCC** | **0.293±0.025** | **0.325±0.021** | **0.318±0.031** | **0.462±0.021** | **0.386±0.022** |

**Table 3: Kolmogorov-Smirnov p-values at 95% significance for F1 and MCC scores**

|  |  |  |  |
| --- | --- | --- | --- |
| **F-score** |  |  |  |
|  | **PredUs** | **Ispred** | **Dockpred** |
| **Logreg** | **3.84E-13** | **7.09E-12** | **2.77E-09** |
| **RF** | **6.15E-05** | **7.13E-04** | **1.11E-02** |
|  |  |  |  |
| **MCC-score** |  |  |  |
| **Logreg** | **5.52E-21** | **3.85E-18** | **1.77E-14** |
| **RF** | **5.12E-09** | **9.51E-08** | **8.27E-07** |

**Table 4: Comparison of AUC-ROC and AUC-PR Using LOOC**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Predus** | **Ispred** | **Dockpred** | **Logreg** | **RF** |
| **AUC-ROC** | **0.693** | **0.810** | **0.801** | **0.918** | **0.874** |
| **AUC-PR** | **0.244** | **0.333** | **0.26** | **0.583** | **0.393** |

**Table 5: Statistical Significance of AUC-ROC Values**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **logreg** | **RF** | **ispred** | **dockpred** | **predus** |
| **logreg** | **N.A.** | **0.046** | **0.109** | **0.117** | **0.223** |
| **RF** | **< 0.05** | **N.A.** | **0.063** | **0.071** | **0.177** |
| **ispred** | **< 0.05** | **< 0.05** | **N.A.** | **0.007** | **0.113** |
| **dockpred** | **< 0.05** | **< 0.05** | **0.108** | **N.A.** | **0.106** |
| **predus** | **< 0.05** | **< 0.05** | **< 0.05** | **< 0.05** | **N.A.** |

**Table 6: F1 and MCC scores from 5-fold and 10-fold cross-validation**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Set #** | **Logreg-F1 score** | **Logerg-MCC score** | **Logreg-F1 score** | **Logreg-MCC score** | **RF-F1 score** | **RF-MCC score** | **RF-F1 score** | **RF-MCC score** |
| **1** | **0.474** | **0.415** | **0.483** | |  | | --- | | **0.426** | | **0.478** | **0.420** | **0.508** | **0.455** |
| **2** | **0.366** | **0.316** | **0.465** | **0.404** | **0.356** | **0.304** | **0.454** | **0.392** |
| **3** | **0.432** | **0.376** | **0.380** | **0.332** | **0.426** | **0.369** | **0.371** | **0.321** |
| **4** | **0.409** | **0.364** | **0.350** | **0.298** | **0.417** | **0.373** | **0.353** | **0.302** |
| **5** | **0.405** | **0.394** | **0.477** | **0.431** | **0.433** | **0.427** | **0.452** | **0.402** |
| **6** |  |  | **0.398** | **0.337** |  |  | **0.398** | **0.337** |
| **7** |  |  | **0.400** | **0.353** |  |  | **0.423** | **0.380** |
| **8** |  |  | **0.405** | **0.362** |  |  | **0.409** | **0.366** |
| **9** |  |  | **0.429** | **0.421** |  |  | **0.466** | **0.464** |
| **10** |  |  | **0.396** | **0.384** |  |  | **0.417** | **0.408** |

**Table 7: Comparison of 5-fold and 10-fold cross validation**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **F1** | **MCC** | **AUC-ROC** | **AUC-PR** |
| **5-fold** |  |  |  |  |  |
| **Logreg** |  | **0.417±0.036** | **0.373±0.033** | **0.856±0.023** | **0.389±0.053** |
| **RF** |  | **0.422±0.039** | **0.379±0.044** | **0.872±0.022** | **0.395±0.053** |
|  |  |  |  |  |  |
| **10-fold** |  |  |  |  |  |
| **Logreg** |  | **0.418±0.042** | **0.374±0.045** | **0.854±0.026** | **0.390±0.059** |
| **RF** |  | **0.425±0.044** | **0.383±0.053** | **0.873±0.028** | **0.399±0.061** |

**Table 8: Comparison of results with other meta methods**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **F1** | **MCC** | **AUC-ROC** | **AUC-PR** |
|  |  |  |  |  |
| **Logreg** | **0.486** | **0.449** | **0.918** | **0.583** |
| **RF** | **0.426** | **0.381** | **0.874** | **0.393** |
| **Meta-PPISP** | **0.332** | **0.275** | **0.794** | **0.251** |
| **VORFFIP** | **0.457** | **0.416** | **0.888** | **0.483** |

**Table 9: Kolmogorov-Smirnov p-values at 95% significance for F1 and MCC scores**

|  |  |  |
| --- | --- | --- |
| **F-score** |  |  |
|  | **Meta-PPISP** | **VORFFIP** |
| **Logreg** | **7.98E-15** | **7.13E-04** |
| **RF** | **1.59E-05** | **5.91E-03** |
|  |  |  |
| **MCC-score** |  |  |
| **Logreg** | **1.02E-22** | **2.52E-05** |
| **RF** | **7.89E-10** | **2.02E-02** |

**Table 10: Statistical Significance of AUC-ROC Values**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **logreg** | **RF** | **VORFFIP** | **Meta-PPISP** |
| **logreg** | **N.A.** | **0.047** | **0.030** | **0.121** |
| **RF** | **< 0.05** | **N.A.** | **0.016** | **0.074** |
| **VORFFIP** | **< 0.05** | **< 0.05** | **N.A.** | **0.090** |
| **Meta-PPISP** | **<0.05** | **<0.05** | **<0.05** | **NA** |

**Figure Captions:**

Figure 1: Comparison of ROC curves for the three independent methods with the meta methods.



Figure 2: Comparison of PR curves for the three independent methods with the meta methods.



Figure 3: Comparison of ROC curves for the meta methods with meta-PPISP and VORFIPP.



Figure 4: Comparison of PR curves for the meta methods with meta-PPISP and VORFIPP.



Figure 1:

Figure 2:

Figure 3:

Figure 4: