

# Meta-Signer: Metagenomic Signature Identifier based on Rank Aggregation of Features

Derek Reiman<sup>1</sup>, Ahmed Metwally<sup>2</sup>, Jun Sun<sup>3</sup>, and Yang Dai<sup>1</sup>

<sup>1</sup>Department of Bioengineering, University of Illinois at Chicago, Chicago, USA 60612

<sup>2</sup>Department of Genetics, Stanford University, Stanford, USA 94305

<sup>3</sup>Department of Medicine, University of Illinois at Chicago, Chicago, USA 60612

## Extended Data

For evaluation, we compared the features ranked using BioSigner and the PERMANOVA test. We ranked features of the PRISM dataset described in the main manuscript using the different methods when considering the binary case (IBD vs Healthy). For Biosigner, we varied the *pvalN* parameter using 0.5, 0.1, and 0.2 to change the levels of significance. We used the top 30 taxa from each method, except for Biosigner, which identified less than 30 taxa. We then trained machine learning models on PRISM dataset using the features selected from each method and evaluated predictions on the external test set. The results are shown in Table 1.

**Table 1.** An example of a simple table with caption.

|                     |           | Meta-Signer | Permanova   | BioSigner (0.05) | BioSigner (0.10) | BioSigner (0.20) |
|---------------------|-----------|-------------|-------------|------------------|------------------|------------------|
| RF                  | AUC       | <b>0.83</b> | 0.80        | 0.66             | 0.65             | 0.72             |
|                     | MCC       | <b>0.41</b> | 0.35        | 0.07             | 0.22             | 0.20             |
|                     | Precision | <b>0.75</b> | 0.71        | 0.59             | 0.69             | 0.64             |
|                     | Recall    | <b>0.75</b> | 0.72        | 0.65             | 0.69             | 0.66             |
|                     | F1        | <b>0.73</b> | 0.71        | 0.58             | 0.63             | 0.65             |
| SVM                 | AUC       | 0.62        | <b>0.69</b> | 0.64             | 0.57             | 0.60             |
|                     | MCC       | 0.08        | <b>0.25</b> | -0.08            | -0.09            | 0.09             |
|                     | Precision | 0.59        | <b>0.79</b> | 0.50             | 0.43             | 0.61             |
|                     | Recall    | 0.63        | <b>0.69</b> | 0.62             | 0.65             | 0.66             |
|                     | F1        | <b>0.60</b> | 0.59        | 0.53             | 0.52             | 0.57             |
| Logistic Regression | AUC       | <b>0.69</b> | <b>0.69</b> | 0.58             | 0.53             | 0.62             |
|                     | MCC       | 0.20        | <b>0.30</b> | 0.09             | 0.04             | 0.02             |
|                     | Precision | 0.64        | <b>0.69</b> | 0.61             | 0.58             | 0.56             |
|                     | Recall    | 0.65        | <b>0.71</b> | 0.66             | 0.65             | 0.60             |
|                     | F1        | <b>0.64</b> | 0.63        | 0.57             | 0.56             | 0.57             |
| MLPNN               | AUC       | 0.67        | <b>0.71</b> | 0.53             | 0.62             | 0.69             |
|                     | MCC       | 0.15        | <b>0.18</b> | 0.00             | 0.02             | 0.08             |
|                     | Precision | 0.62        | <b>0.65</b> | 0.55             | 0.56             | 0.59             |
|                     | Recall    | 0.65        | <b>0.68</b> | 0.65             | 0.60             | 0.63             |
|                     | F1        | <b>0.63</b> | <b>0.63</b> | 0.54             | 0.57             | 0.60             |

In addition we compare Meta-Signer to PERMANOVA when using the three classes for the PRISM dataset. Biosigner was excluded because it was unable to perform the analysis with three classes. The results are shown in Table 2.

---

**Table 2.** An example of a simple table with caption.

|            |           | Meta-Signer | Permanova   |
|------------|-----------|-------------|-------------|
|            | AUC       | <b>0.73</b> | 0.71        |
|            | MCC       | <b>0.19</b> | 0.17        |
| RF         | Precision | 0.36        | <b>0.47</b> |
|            | Recall    | <b>0.45</b> | <b>0.45</b> |
|            | F1        | 0.36        | 0.43        |
|            | AUC       | <b>0.75</b> | 0.68        |
|            | MCC       | <b>0.36</b> | 0.12        |
| SVM        | Precision | <b>0.60</b> | 0.41        |
|            | Recall    | <b>0.57</b> | 0.42        |
|            | F1        | <b>0.57</b> | 0.40        |
|            | AUC       | <b>0.73</b> | 0.66        |
| Logistic   | MCC       | <b>0.26</b> | 0.19        |
| Regression | Precision | <b>0.52</b> | 0.47        |
|            | Recall    | <b>0.51</b> | 0.46        |
|            | F1        | <b>0.50</b> | 0.46        |
|            | AUC       | <b>0.71</b> | 0.62        |
|            | MCC       | <b>0.19</b> | 0.17        |
| MLPNN      | Precision | 0.47        | <b>0.48</b> |
|            | Recall    | <b>0.46</b> | 0.45        |
|            | F1        | <b>0.46</b> | 0.44        |