# Ensemble Model for Promiscuity Prediction

We have designed an ensemble model to predict enzyme promiscuity class. Our ensemble framework comprises of 48 base learners which provide their outputs to a meta learner that predicts enzyme promiscuity. Although all of our base learners are trained using the same learning algorithm (Support Vector Classifier \cite), the heterogeneity among them is governed by the 47 different feature extraction techniques (listed in Table 1) used to encode the set of enzyme sequences. Each of the 47 different feature extraction techniques creates a unique feature vector representation of enzymes which serve as input to a designated base learner. The base learner trained on the set of encoded enzyme sequences yields the predicted promiscuity class of a given enzyme sequence as an output. The outputs of the k-best base learners are passed on to the meta learner that uses a majority voting scheme to predict the enzyme promiscuity category. The k-best base learners are selected on the basis of their performance on validation set. The workflow of our ensemble model is presented in Figure 1.

Table 1: Feature extraction type, name, software used to deploy and literature from which it was adopted are shown.

|  |  |  |  |
| --- | --- | --- | --- |
| Type | Name | Software Used | Literature Reference |
| Kernel | Spectrum Kernel | KeBABS | [1] |
| Kernel | Mismatch Kernel | KeBABS | [2] |
| Kernel | Gappy Pair Kernel | KeBABS | [3] |
| N-gram | Kmer | Numpy | [4] |
| N-gram | GAA-kmer | Numpy | NA |
| Physicochemical | AAC | iFeature | [5] |
| Physicochemical | CKSAAP | iFeature | [6] |
| Physicochemical | TPC | iFeature | [5] |
| Physicochemical | DPC | iFeature | [7] |
| Physicochemical | DDE | iFeature | [7] |
| Physicochemical | GAAC | iFeature | [8] |
| Physicochemical | CKSAAGP | iFeature | [9] |
| Physicochemical | GTPC | iFeature | [9] |
| Physicochemical | GDPC | iFeature | [9] |
| Physicochemical | Moran | iFeature | [10] |
| Physicochemical | Geary | iFeature | [11] |
| Physicochemical | NMBroto | iFeature | [12] |
| Physicochemical | CTDC | iFeature | [13]–[15] |
| Physicochemical | CTDT | iFeature | [13]–[15] |
| Physicochemical | CTDD | iFeature | [13]–[15] |
| Physicochemical | CTriad | iFeature | [16] |
| Physicochemical | KSCTriad | iFeature | [9] |
| Physicochemical | SOCNumber | iFeature | [9] |
| Physicochemical | QSOrder | iFeature | [9] |
| Physicochemical | PAAC | iFeature | [17] |
| Physicochemical | APAAC | iFeature | [17] |
| PSSM-based | AAC-PSSM | POSSUM | [18] |
| PSSM-based | DPC-PSSM | POSSUM | [18] |
| PSSM-based | AADP-PSSM | POSSUM | [18] |
| PSSM-based | PSSM-AC | POSSUM | [19] |
| PSSM-based | PSSM-CC | POSSUM | [20] |
| PSSM-based | RPSSM | POSSUM | [21] |
| PSSM-based | Tri-gram | POSSUM | [22] |
| PSSM-based | EDP | POSSUM | [23] |
| PSSM-based | EEDP | POSSUM | [23] |
| PSSM-based | MEDP | POSSUM | [23] |
| PSSM-based | TPC-PSSM | POSSUM | [24] |
| PSSM-based | AATP | POSSUM | [24] |
| PSSM-based | k-separated bigrams | POSSUM | [25] |
| PSSM-based | D-FPSSM | POSSUM | [26] |
| PSSM-based | S-FPSSM | POSSUM | [26] |
| PSSM-based | PSE-PSSM | POSSUM | [27] |
| PSSM-based | DP-PSSM | POSSUM | [28] |
| PSSM-based | PSSM-Composition | POSSUM | [29] |
| PSSM-based | Smoothed-PSSM | POSSUM | [30] |
| PSSM-based | AB-PSSM | POSSUM | [31] |
| PSSM-based | RPM-PSSM | POSSUM | [31] |

# Ensemble Model Results

Diagram

Description automatically generated

Figure 1: Workflow of the ensemble model. Different feature representations create separate models, and the final model output is the majority vote of the predictions made by each individual model.