

Figure 1. Diagram of the model construction used for downstream analysis. Positional coordinates of TAD boundaries were obtained from Rao et al from contact matrices of X different cell lines. The linear genome was binned according to the resolution of the respective HiC experiment (5kb for GM12878; 10kb for all others). The response vector Y used for classification was determined by whether or not a genomic bin contained a TAD boundary or not. The positional coordinates of each functional genomic element, obtained from ENCODE, was used to define the feature space of the models.

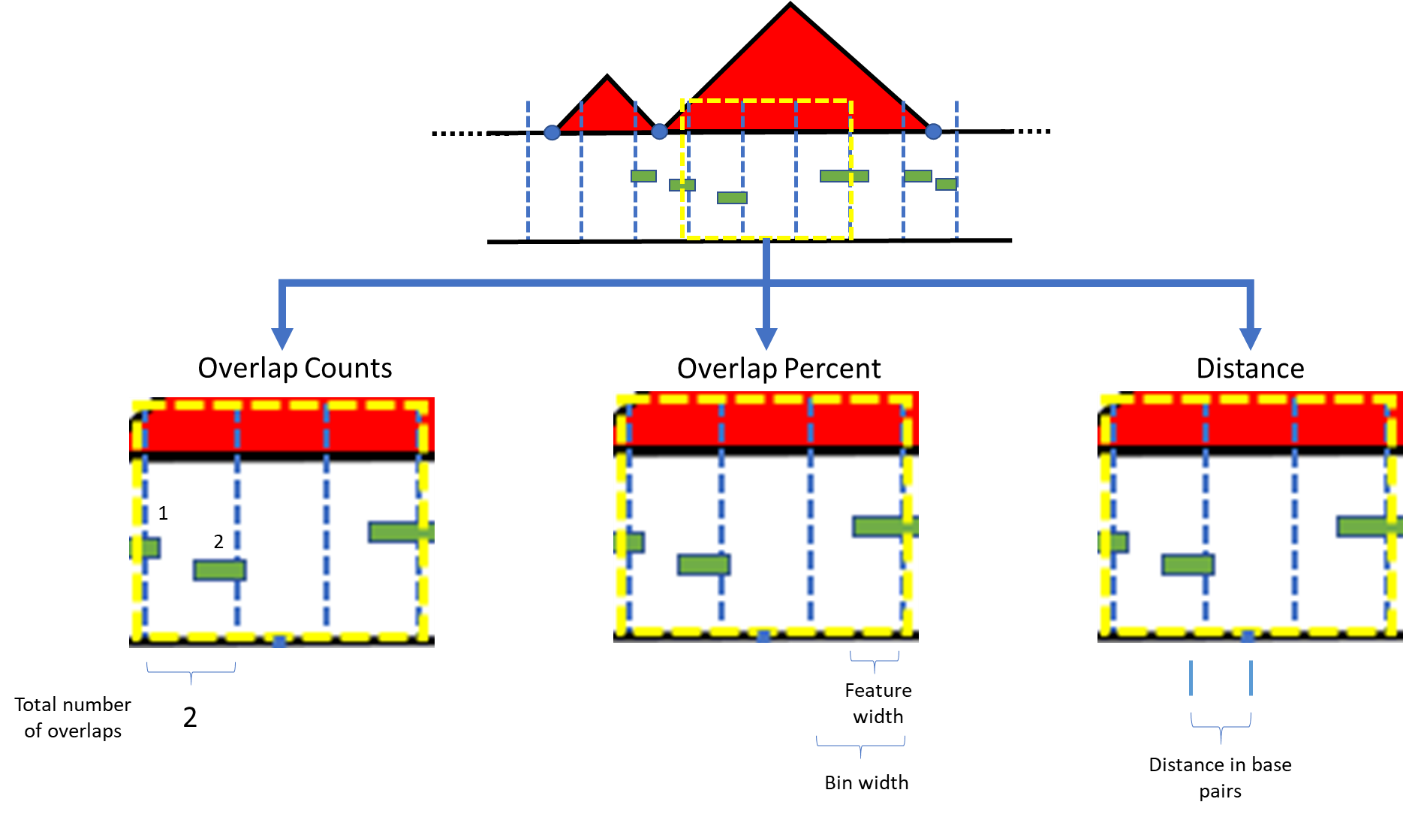


Figure 2. Diagram of the 3 predictor types considered when assessing the relationship between TAD boundaries and functional genomic elements. Each predictor type was used as the feature space in downstream analyses for predicting which genomic elements were associated with the formation of TAD boundaries. (A) The overlap count (OC) predictors were calculated by considering the total number of elemental regions that overlapped with each genomic bin. For bin (I), there is only one overlapped region with that bin. Therefore, the ith position in the feature vector would be 1. (B) The overlap percent (OP) predictors were calculated by dividing the sum of all feature widths within a bin and dividing by the total bin width (either 5 or 10kb depending on the cell line ). An example for bin (III) is provided. (C) The distance predictors were calculated by measuring the distance (in base pairs) from the center of each genomic bin to the center of the nearest elemental region of interest. An example for bin (II) is provided.

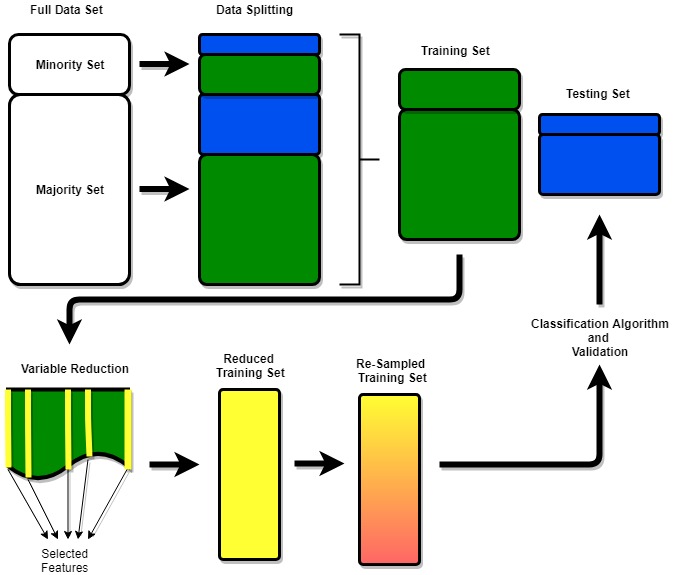


Figure 3. A diagram of the modelling pipeline. The process was the same for each combination of predictor type, variable reduction technique, and resampling method considered. The data was split into a 7:3 training set to testing set ratio. The variable reduction technique of choice was performed. The training set was then reduced and the classification algorithm of choice was performed. Each model was validated on the same testing set.

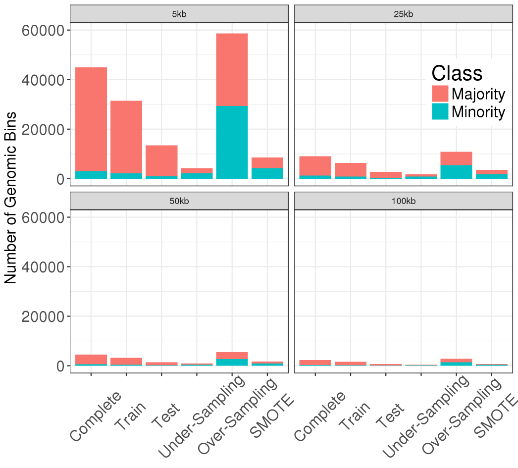


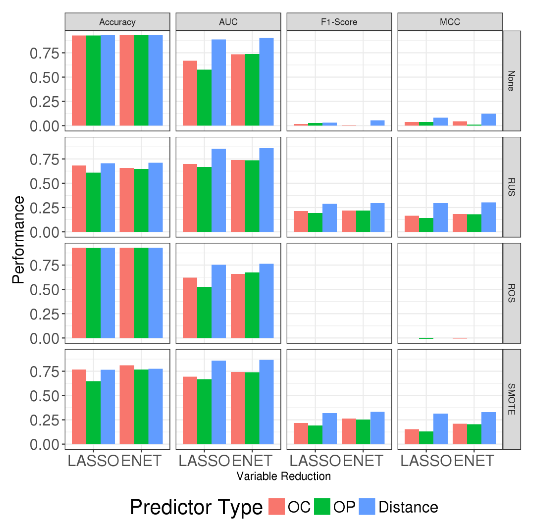
Figure 4.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **5kb Resolution** | | | | | | | |
| **Genomic Bin** | **Complete** | **Training** | **Testing** | **No Balancing** | **Random Under-Sampling** | **Random Over-Sampling** | **SMOTE** |
| With TAD boundary | 3083 | 2140 | 943 | 2140 | 2140 | 29323 | 4280 |
| Without TAD boundary | 41865 | 29323 | 12542 | 29323 | 2140 | 29323 | 4280 |
| Total | 44948 | 31463 | 13485 | 31463 | 4280 | 58646 | 8560 |

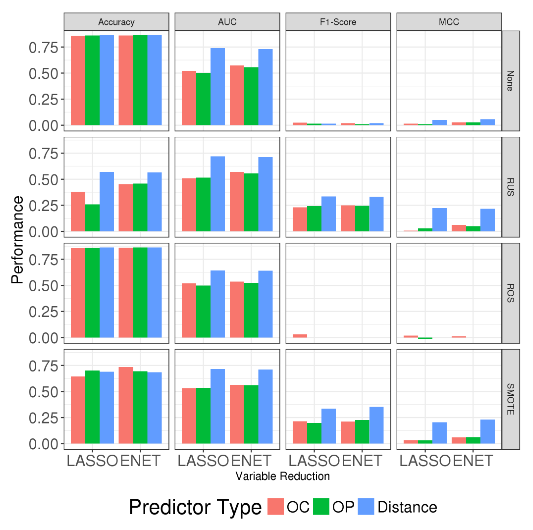
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **25kb Resolution** | | | | | | | |
| **Genomic Bin** | **Complete** | **Training** | **Testing** | **No Balancing** | **Random Under-Sampling** | **Random Over-Sampling** | **SMOTE** |
| With TAD boundary | 1266 | 895 | 371 | 895 | 895 | 5434 | 1790 |
| Without TAD boundary | 7776 | 5434 | 2342 | 5434 | 895 | 5434 | 1790 |
| Total | 9042 | 6329 | 2713 | 6329 | 1790 | 10868 | 3580 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **50kb Resolution** | | | | | | | |
| **Genomic Bin** | **Complete** | **Training** | **Testing** | **No Balancing** | **Random Under-Sampling** | **Random Over-Sampling** | **SMOTE** |
| With TAD boundary | 621 | 431 | 190 | 431 | 431 | 2747 | 862 |
| Without TAD boundary | 3920 | 2747 | 1173 | 2747 | 431 | 2747 | 862 |
| Total | 4541 | 3178 | 1363 | 3178 | 862 | 5494 | 1724 |

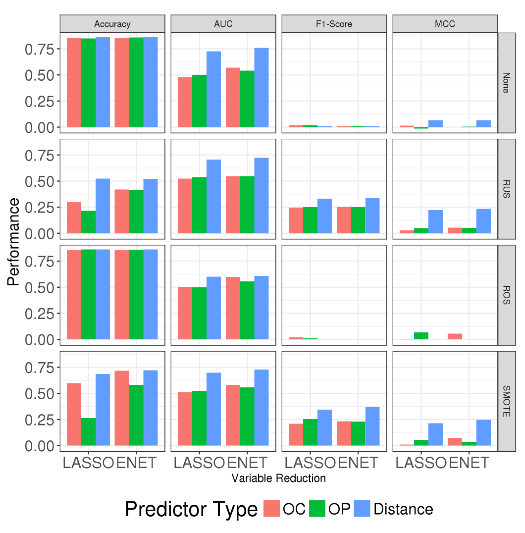
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **100kb Resolution** | | | | | | | |
| **Genomic Bin** | **Complete** | **Training** | **Testing** | **No Balancing** | **Random Under-Sampling** | **Random Over-Sampling** | **SMOTE** |
| With TAD boundary | 277 | 185 | 92 | 185 | 185 | 1410 | 370 |
| Without TAD boundary | 2002 | 1410 | 592 | 1410 | 185 | 1410 | 370 |
| Total | 2279 | 1595 | 684 | 1595 | 370 | 2820 | 740 |



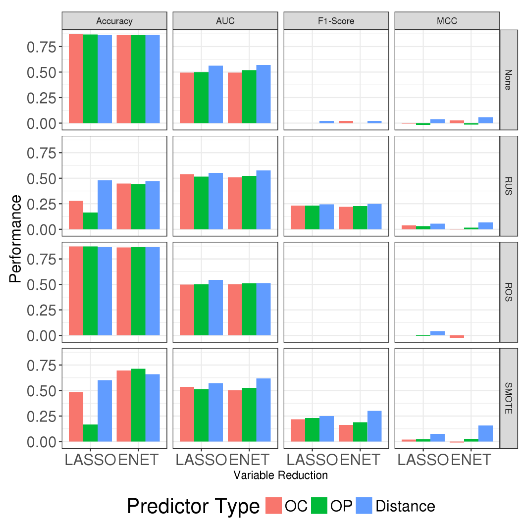
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | LASSO | | | Elastic-Net | | |
|  | Balancing Technique | OC | OP | Distance | OC | OP | Distance |
| Accuracy | None | 0.929 | 0.927 | 0.93 | 0.93 | 0.93 | 0.931 |
| RUS | 0.684 | 0.609 | 0.705 | 0.657 | 0.647 | 0.71 |
| ROS | 0.929 | 0.927 | 0.93 | 0.93 | 0.93 | 0.93 |
| SMOTE | 0.768 | 0.648 | 0.764 | 0.812 | 0.768 | 0.775 |
| AUC | None | 0.669 | 0.577 | 0.888 | 0.734 | 0.737 | 0.902 |
| RUS | 0.697 | 0.668 | 0.853 | 0.739 | 0.737 | 0.861 |
| ROS | 0.621 | 0.527 | 0.753 | 0.659 | 0.674 | 0.763 |
| SMOTE | 0.694 | 0.668 | 0.858 | 0.741 | 0.741 | 0.867 |
| F1-Score | None | 0.018 | 0.03 | 0.031 | 0.004 | 0.002 | 0.055 |
| RUS | 0.214 | 0.194 | 0.289 | 0.22 | 0.218 | 0.295 |
| ROS | NA | NA | NA | NA | NA | NA |
| SMOTE | 0.217 | 0.192 | 0.318 | 0.263 | 0.252 | 0.332 |
| MCC | None | 0.038 | 0.04 | 0.081 | 0.044 | 0.01 | 0.124 |
| RUS | 0.165 | 0.143 | 0.296 | 0.183 | 0.181 | 0.304 |
| ROS | -0.001 | -0.009 | NA | -0.005 | NA | NA |
| SMOTE | 0.153 | 0.132 | 0.313 | 0.209 | 0.205 | 0.331 |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | LASSO | | | Elastic-Net | | |
|  | Balancing Technique | OC | OP | Distance | OC | OP | Distance |
| Accuracy | None | 0.857 | 0.86 | 0.863 | 0.861 | 0.863 | 0.863 |
| RUS | 0.377 | 0.26 | 0.569 | 0.451 | 0.459 | 0.565 |
| ROS | 0.856 | 0.86 | 0.863 | 0.862 | 0.863 | 0.863 |
| SMOTE | 0.644 | 0.703 | 0.691 | 0.734 | 0.693 | 0.684 |
| AUC | None | 0.521 | 0.502 | 0.74 | 0.574 | 0.558 | 0.73 |
| RUS | 0.509 | 0.515 | 0.72 | 0.57 | 0.555 | 0.713 |
| ROS | 0.519 | 0.499 | 0.644 | 0.535 | 0.525 | 0.64 |
| SMOTE | 0.533 | 0.535 | 0.717 | 0.565 | 0.561 | 0.71 |
| F1-Score | None | 0.025 | 0.016 | 0.016 | 0.021 | 0.011 | 0.021 |
| RUS | 0.23 | 0.243 | 0.334 | 0.25 | 0.245 | 0.33 |
| ROS | 0.03 | NA | NA | NA | NA | NA |
| SMOTE | 0.215 | 0.197 | 0.336 | 0.212 | 0.227 | 0.352 |
| MCC | None | 0.016 | 0.011 | 0.05 | 0.028 | 0.027 | 0.057 |
| RUS | 0.008 | 0.029 | 0.224 | 0.061 | 0.051 | 0.217 |
| ROS | 0.019 | -0.013 | NA | 0.011 | NA | NA |
| SMOTE | 0.034 | 0.031 | 0.206 | 0.06 | 0.063 | 0.231 |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | LASSO | | | Elastic-Net | | |
|  | Balancing Technique | OC | OP | Distance | OC | OP | Distance |
| Accuracy | None | 0.856 | 0.849 | 0.861 | 0.857 | 0.858 | 0.861 |
| RUS | 0.299 | 0.214 | 0.524 | 0.42 | 0.417 | 0.522 |
| ROS | 0.855 | 0.861 | 0.861 | 0.86 | 0.86 | 0.861 |
| SMOTE | 0.599 | 0.263 | 0.686 | 0.718 | 0.58 | 0.721 |
| AUC | None | 0.479 | 0.5 | 0.726 | 0.57 | 0.541 | 0.759 |
| RUS | 0.523 | 0.537 | 0.706 | 0.548 | 0.547 | 0.722 |
| ROS | 0.503 | 0.5 | 0.601 | 0.598 | 0.557 | 0.609 |
| SMOTE | 0.513 | 0.524 | 0.698 | 0.582 | 0.557 | 0.728 |
| F1-Score | None | 0.02 | 0.019 | 0.01 | 0.01 | 0.01 | 0.01 |
| RUS | 0.245 | 0.251 | 0.331 | 0.252 | 0.251 | 0.336 |
| ROS | 0.019 | 0.01 | NA | NA | NA | NA |
| SMOTE | 0.208 | 0.253 | 0.344 | 0.232 | 0.229 | 0.369 |
| MCC | None | 0.015 | -0.014 | 0.067 | 0.001 | 0.005 | 0.067 |
| RUS | 0.028 | 0.048 | 0.222 | 0.054 | 0.051 | 0.234 |
| ROS | 0.006 | 0.067 | NA | 0.055 | NA | NA |
| SMOTE | 0.009 | 0.054 | 0.213 | 0.074 | 0.035 | 0.247 |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | LASSO | | | Elastic-Net | | |
|  | Balancing Technique | OC | OP | Distance | OC | OP | Distance |
| Accuracy | None | 0.871 | 0.869 | 0.864 | 0.863 | 0.864 | 0.865 |
| RUS | 0.279 | 0.165 | 0.48 | 0.448 | 0.444 | 0.471 |
| ROS | 0.871 | 0.871 | 0.865 | 0.862 | 0.864 | 0.864 |
| SMOTE | 0.485 | 0.167 | 0.602 | 0.696 | 0.712 | 0.659 |
| AUC | None | 0.496 | 0.501 | 0.563 | 0.494 | 0.519 | 0.57 |
| RUS | 0.539 | 0.515 | 0.552 | 0.509 | 0.523 | 0.576 |
| ROS | 0.497 | 0.5 | 0.544 | 0.501 | 0.509 | 0.513 |
| SMOTE | 0.535 | 0.513 | 0.572 | 0.502 | 0.526 | 0.618 |
| F1-Score | None | NA | NA | 0.021 | 0.021 | NA | 0.021 |
| RUS | 0.233 | 0.231 | 0.243 | 0.22 | 0.226 | 0.249 |
| ROS | NA | NA | NA | NA | NA | NA |
| SMOTE | 0.218 | 0.231 | 0.249 | 0.161 | 0.189 | 0.3 |
| MCC | None | -0.004 | -0.018 | 0.039 | 0.026 | -0.015 | 0.058 |
| RUS | 0.038 | 0.03 | 0.055 | 0.004 | 0.016 | 0.066 |
| ROS | NA | -0.004 | 0.043 | -0.024 | NA | NA |
| SMOTE | 0.02 | 0.025 | 0.076 | -0.01 | 0.028 | 0.158 |

