**Supplementary 9: Accuracies of Training and Testing Sets**

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1:** Model ensemble methods, sampling strategies, and corresponding accuracy metrics for the training set and testing sets (one internal and one external). | | | |
|  | **Final Model** | Model 1 | Model 2 |
| **Ensemble method** | **RF** | RF | glm |
| **Sampling** | **Cross validation with k-fold (5)** | Bootstrap 0.632 (with replacement) | Cross validation with k-fold (5) |
|  |  |  |  |
| **Training** |  |  |  |
| Accuracy | 0.92 | 0.91 | 0.82 |
| Precision | 1 | 0.86 | 0.65 |
| Recall | 0.71 | 0.79 | 0.79 |
| F1-Score | 0.83 | 0.83 | 0.72 |
|  |  |  |  |
| **Testing (Internal)** |  |  |  |
| Accuracy | 0.86 | 0.81 | 0.76 |
| Precision | 0.8 | 0.6 | 0.55 |
| Recall | 0.67 | 1 | 1 |
| F1-Score | 0.73 | 0.75 | 0.71 |
|  |  |  |  |
| **Testing (External)** |  |  |  |
| Accuracy | 0.69 | - | - |
| Precision | 0.71 | - | - |
| Recall | 0.82 | - | - |
| F1-Score | 0.76 | - | - |