**Appendix I: Individual Learners**

The following provides a brief description of the individual algorithms included in the ensemble super learner model.

1. **Elastic Net Binomial Regression:** Elastic Net Binomial Regression is a statistical method used for binary classification tasks. It combines the L1 and L2 regularization techniques (Lasso and Ridge regression) to address multicollinearity issues in the data. This method is particularly useful when dealing with high-dimensional datasets, where it simultaneously performs feature selection and regularization to improve the model's predictive accuracy.
2. **Ranger Random Forest:** Ranger is an implementation of the Random Forest algorithm, a versatile ensemble learning technique. It operates by constructing multiple decision trees during training and outputs the mode of the classes (for classification tasks) of the individual trees. Ranger enhances the traditional Random Forest by optimizing performance and memory usage, making it efficient for high-dimensional feature spaces.
3. **Support Vector Machine (SVM):** Support Vector Machine is a supervised machine learning algorithm used for both classification and regression tasks. SVM works by finding the optimal hyperplane that best divides a dataset into classes in a high-dimensional space. It is particularly effective in high-dimensional spaces and is capable of handling non-linear relationships through kernel specifications, making it a powerful tool for complex classification problems.
4. **Neural Network:** A neural network is a computational model inspired by the human brain, consisting of interconnected nodes (neurons) organized in layers. Each node connects to another and has an associated weight and threshold. If the output of any node exceeds a specified threshold value, then that node is activated and transmits a signal to the next layer of the network. Neural networks excel in capturing complex patterns and relationships in large and high-dimensional datasets.
5. **Extreme Gradient Boosting (XGBoost):** Extreme Gradient Boosting is an ensemble learning technique that combines the predictions of several weak learners (typically decision trees) to create a stronger and more robust model. XGBoost combines traditional decision tree methods with boosting by sequentially and iteratively building trees and correcting errors made by previous models until performance is optimized.
6. **Bayesian Generalized Linear Regression:** Bayesian Generalized Linear Regression is a statistical framework that combines generalized linear models with Bayesian inference. It extends traditional regression techniques by incorporating prior knowledge about the relationships between variables. By applying Bayesian methods, this approach provides a probabilistic interpretation of the model parameters, allowing for more nuanced uncertainty quantification and better handling of small sample sizes.
7. **Mean of Outcome Variable:** The sample mean of the outcome variable is used as a benchmark algorithm against which more complex algorithms are compared.