Improving the interpretability of random forest models of genetic association in the presence of non-additive interactions

Alena Orlenko, Jason H Moore

Background
• Non-additive interactions among genes are frequently associated with a number of phenotypes, including known complex diseases such as Alzheimer’s, diabetes, and cardiovascular disease.
• Detecting interactions requires careful selection of analytical methods, and some machine learning algorithms are unable or underpowered to detect or model feature interactions that exhibit non-additivity.
• The Random Forest (RF) method is often employed in these efforts due to its ability to detect and model non-additive interactions. RF has the built-in ability to estimate feature importance via a characteristic that allows the model to be interpreted with the order and effect size of the feature association with the outcome. This characteristic is very important for epidemiological and clinical studies where results of predictive modeling could be used to define the future direction of the research efforts.
• An alternative way to interpret the model is with a permutation feature importance metric which employs a permutation approach and with the Shapley additive explanations which considers any additive interactions among features and compared RF’s way interactions within the Glaucoma disease datasets. SNPs main effects, two-way and three-way interactions

Methods
• We used Heuristic Identification of Biological Architectures for simulating Complex Hierarchical Interactions (HIBACHI) software to simulate genetic datasets with non-additive epistatic interactions.
• The HIBACHI framework has the ability to consider any desirable biological concept in the form of mathematical expressions that define the genotype-phenotype relationship and evolve models that can be used to simulate data consistent with that relationship. We set up a simulation goal to maximize two or three-way interactions among features and compared RF’s feature importance metrics with the sensitivity analysis results of the simulated data that provided us with the ground truth information on the feature ranks.
• To examine the convergence of the RF’s feature importance metrics we used two real-world datasets with evidence for non-additive interactions (genome-wide association study of Alzheimer’s Disease and a genome-wide association study of Primary Open Angle Glaucoma).
• We used the visualization of the statistical interaction network (ViSEN) method to analyze and visualize SNP main effects, and two-way and three-way gene interactions among SNPs for real-world datasets via the mutual information and information gain terms.

Evaluation of the feature importances in real-world datasets with non-additive interactions
• The most powerful predictor of Alzheimer’s disease at this time is ApoE4 gene variation: one or two copies of ApoE is associated with an increased risk of disease onset. Some carriers of ApoE4 variation haven’t developed Alzheimer’s disease so it is very likely that other genetic factors are involved in disease’s pathophysiology. ViSEN entropy-based analysis revealed several strong pairwise genetic interactions, along with the known largest independent signal from the ApoE variant (n=49558) (Fig.1A).
• VISEN method selected non-additive interactions within the Glaucoma disease dataset: several strong pairwise interactions in addition to the independent main effect contribution from the SNP affiliated with retinal ganglion cells pathology (n=2157719) have been confirmed (Fig.1B). The HIBACHI framework allowed us to compare three feature importance metrics: RF’s permutation feature importance metric (PFI), BIC and SHAP, and each was compared after RF analysis of data derived from genome-wide association studies of Glaucoma and Alzheimer’s.
• The resulting feature ranking confirms the lack of consensus between the studied metrics (Fig.2 A, B).

Evaluation of feature importance metrics performance with simulated datasets
• Prediction uncertainty has been associated with RF predictions in the past and we attempted to reveal the true interpretation with the computational experiments driven by HIBACHI simulations.
• We set up a simulation goal to maximize two or three-way interactions among features and compared RF’s feature importance metrics with the sensitivity analysis results of the simulated data that provided us with the ground truth information about the feature ranks.
• In all HIBACHI experimental setups, which included such factors as the proportion of additive and non-additive interactions, sample size and interaction complexity, RF metrics produced the most precise feature ranking (Table 1, Fig. 3).
• Although BIC and SHAP metrics exhibited feature ranks for the large percentage of replicates with BIC failed to identify the majority of them, it correctly identified features that belong to the interactive pair or trio by putting them as a top pair or trio features correspondingly.
• We found that RF predictions were more precise than BIC and SHAP. While BIC and SHAP metrics can still be useful, when there is a need for an absolute precision, RF estimation method should be used.

Results
• Three feature importance metrics were considered, PFI, BIC and SHAP, and each was compared after RF analysis of data derived from genome-wide association studies of Glaucoma and Alzheimer’s.

Conclusion
• We performed a comparative analysis of feature importance metrics with the aim to improve Random Forest’s interpretability in datasets with complex interactions.
• By analyzing both real and simulated data, we established that the permutation feature importance metric provides more precise feature importance rank estimation in the presence of non-additive interactions, while applying cooperative game theory approach.


Funding: This work was supported by National Institutes of Health (USA) grants LM010998 and AI16794.