

# Interaction-Stratified Oblique Tree Ensembles with Provable Functional Decomposability

Anonymous Authors  
*Under Review*

## Abstract

Modern interpretable machine learning faces a fundamental tension: axis-aligned tree ensembles (e.g., FIGS) permit clean functional ANOVA decomposition into main effects and feature interactions, while oblique tree ensembles (e.g., RO-FIGS) achieve superior predictive accuracy by blending features within splits—at the cost of decomposability. We propose Interaction-Stratified Oblique FIGS (ISO-FIGS), a novel oblique tree ensemble algorithm that resolves this dilemma by assigning each tree an interaction tier. Tier- $k$  trees are structurally constrained so that every oblique split combines at most  $k$  features from a designated feature group, ensuring that the ensemble admits an exact functional ANOVA decomposition while retaining the representational power of oblique decision boundaries. We prove that ISO-FIGS ensembles are functionally decomposable by construction and that the decomposition is unique under mild identifiability conditions. On the Breast Cancer Wisconsin benchmark, ISO-FIGS recovers known feature interaction tiers, matches or exceeds the accuracy of unconstrained oblique ensembles with fewer total splits, and produces interpretable effect attributions aligned with domain-verified feature relationships.

## 1 Introduction

### 1.1 The Expressiveness–Interpretability Chasm

Interpretable machine learning has achieved remarkable progress in developing models that are simultaneously accurate and transparent [Murdoch et al., 2019]. Among the most successful paradigms are tree ensemble methods, which combine the flexibility of decision trees with the stability of ensemble

aggregation. Yet a stubborn chasm divides the two dominant families of tree ensembles along the axis of interpretability.

**Axis-aligned tree ensembles**, exemplified by Fast Interpretable Greedy-Tree Sums (FIGS; Singh et al. 2022), grow multiple trees simultaneously by greedily adding one split at a time to whichever tree reduces the training loss most. Because each split tests a single feature against a threshold (e.g., “mean radius < 14.5”), the resulting ensemble can be cleanly decomposed into main effects and pairwise interactions via functional ANOVA decomposition—the gold standard for transparent model explanation [Hooker, 2007, Lou et al., 2013].

**Oblique tree ensembles**, exemplified by Rotation-Optimized FIGS (RO-FIGS; Jamnik et al. 2025), replace single-feature splits with learned linear combinations of multiple features, typically optimized via gradient descent with sparsity regularization. These models achieve better accuracy with fewer splits because oblique boundaries can capture correlated feature structure more efficiently. However, as Jamnik et al. [2025] acknowledge, oblique splits are more complex and harder to comprehend—each oblique split blends features across interaction orders, making exact functional decomposition impossible.

This tension is not merely academic. In oncology, a clinician needs to see *which feature relationships* drive a prediction—is it tumor size alone, or a synergy between size and texture? In consumer lending, a loan officer must explain *why* an applicant was denied, decomposed into individual factors and their interactions. Today, practitioners face a forced choice: accuracy (oblique) or explainability (axis-aligned). Our central hypothesis is that this is a false dilemma.

## 1.2 Our Contribution: ISO-FIGS

We propose **Interaction-Stratified Oblique FIGS (ISO-FIGS)**, a new oblique tree ensemble algorithm that bridges the expressiveness–interpretability gap through a structural constraint: **each tree is assigned an interaction tier**, and oblique splits within that tree are constrained to combine only features from the tier’s allowed feature group. Specifically:

- **Tier 1** trees capture single-feature (main) effects with effectively axis-aligned splits.
- **Tier 2** trees capture pairwise interactions with oblique splits blending exactly two designated features.
- **Tier  $k$**  trees generalize, with splits combining at most  $k$  features from a pre-specified group.

This tiered architecture yields three key properties:

1. **Provable functional decomposability.** The ensemble admits an exact functional ANOVA decomposition with no post-hoc approximation.
2. **Oblique expressiveness within tiers.** Oblique splits exploit correlated feature structure within each tier, achieving accuracy gains over axis-aligned alternatives.
3. **Automatic interaction discovery.** The greedy growing procedure allocates capacity to the most predictive interaction orders, revealing which feature groups drive the prediction.

## 2 Related Work

**Interpretable Tree Ensembles.** FIGS [Singh et al., 2022] pioneered the simultaneous greedy-tree-sum paradigm, demonstrating that small ensembles can rival larger models while maintaining interpretability. CART [Breiman et al., 1984] and Random Forests [Breiman, 2001] represent the classical endpoints of the single-tree vs. ensemble tradeoff. RO-FIGS [Jamnik et al., 2025] extended FIGS with oblique splits optimized via gradient descent with  $\ell_1$  regularization, achieving significant accuracy gains on correlated-feature datasets.

**Oblique Decision Trees.** Oblique splits have a long history in decision tree research [Breiman et al., 1984]. Modern approaches use gradient-based optimization with sparsity penalties to learn split hyperplanes. The key challenge is that unrestricted

oblique splits destroy the clean feature-attribution structure of axis-aligned trees.

**Functional Decomposition and GAMs.** The functional ANOVA decomposition [Hooker, 2007] provides the theoretical foundation for separating main effects from interactions. Generalized Additive Models (GAMs; Hastie and Tibshirani 1990) and their extensions to pairwise interactions (GA<sup>2</sup>Ms; Lou et al. 2013) explicitly separate these components using splines or boosted stumps. Neural Additive Models [Agarwal et al., 2021], Explainable Boosting Machines [Nori et al., 2019], and GAMI-Net [Yang et al., 2021] extend this paradigm with neural or boosting-based shape functions. ISO-FIGS uses tree-based representations within each ANOVA component, enabling nonlinear, non-additive structure within each interaction term.

**Post-hoc Interpretability.** SHAP [Lundberg and Lee, 2017] provides model-agnostic feature attributions but requires expensive sampling and produces approximate rather than exact decompositions. As we show in Section 5, post-hoc methods applied to non-decomposable models can produce “phantom interactions” that are artifacts of split geometry.

## 3 Methods

### 3.1 Preliminaries and Notation

We consider supervised learning with input  $\mathbf{x} \in \mathbb{R}^p$  and target  $y$ . Let  $[p] = \{1, \dots, p\}$  denote the feature index set.

**FIGS** [Singh et al., 2022] grows an ensemble of  $M$  trees  $\{T_1, \dots, T_M\}$  simultaneously. Each tree  $T_m$  computes  $f_m : \mathbb{R}^p \rightarrow \mathbb{R}$ , and the ensemble prediction is  $\hat{y}(\mathbf{x}) = \sum_{m=1}^M f_m(\mathbf{x})$ .

**RO-FIGS** [Jamnik et al., 2025] replaces axis-aligned splits with oblique splits of the form  $\mathbf{w}^\top \mathbf{x} < \tau$ , where  $\mathbf{w} \in \mathbb{R}^p$  is learned via gradient descent with  $\ell_1$  penalty  $\lambda \|\mathbf{w}\|_1$ .

**Functional ANOVA Decomposition.** For an ensemble  $f(\mathbf{x}) = \sum_m f_m(\mathbf{x})$ , the decomposition writes:

$$f(\mathbf{x}) = f_0 + \sum_{j=1}^p f_j(x_j) + \sum_{j < k} f_{jk}(x_j, x_k) + \dots \quad (1)$$

where  $f_0$  is a constant,  $f_j$  captures the main effect of

feature  $j$ , and  $f_{jk}$  captures the pure interaction between features  $j$  and  $k$ . For axis-aligned ensembles, this decomposition is exact. For oblique ensembles with multiple nonzero weights, features are entangled across interaction orders.

### 3.2 The ISO-FIGS Algorithm

ISO-FIGS enforces functional decomposability by introducing an **interaction tier** for each tree. The algorithm proceeds in two phases.

**Phase 1: Feature Group Discovery.** Before growing trees, ISO-FIGS identifies candidate feature groups via correlation analysis. For each feature pair  $(j, k)$ , the absolute Pearson correlation  $|\rho_{jk}|$  is computed. Feature groups are formed by agglomerative clustering with threshold  $\rho_{\min}$  (default 0.85). For the Breast Cancer Wisconsin dataset, this automatically discovers groups such as {mean radius, mean perimeter, mean area} ( $r > 0.99$ ) and {mean concavity, mean concave points} ( $r > 0.92$ ).

**Phase 2: Tiered Tree Growing.** The ensemble is grown greedily with the following modifications:

1. **Tier Assignment.** Each tree  $T_m$  is assigned tier  $k_m \in \{1, \dots, K\}$  at creation. Tier 1 trees are grown first; higher tiers are introduced as lower-tier residuals stabilize.
2. **Constrained Oblique Splits.** For a tier- $k$  tree  $T_m$  with feature group  $S_m \subseteq [p]$ ,  $|S_m| = k$ , every split is restricted to  $\mathbf{w}_{S_m}^\top \mathbf{x}_{S_m} < \tau$ , where  $\mathbf{w}_{S_m} \in \mathbb{R}^k$ . Features outside  $S_m$  receive zero weight *by construction*—not by regularization.
3. **Greedy Tier-Aware Selection.** A new tree at the next tier is initiated when the best candidate split improves the loss by less than threshold  $\delta$  (default  $10^{-4}$ ).

The full algorithm architecture is illustrated in Figure 1, which shows how the tiered constraint preserves functional decomposability while enabling oblique expressiveness within each tier.

### 3.3 Formal Decomposability Guarantee

**Theorem 1** (Functional Decomposability of ISO-FIGS). *Let  $\hat{f}(\mathbf{x}) = \sum_{m=1}^M f_m(\mathbf{x})$  be an ISO-FIGS ensemble where each tree  $T_m$  has tier  $k_m$  with feature group  $S_m \subseteq [p]$ ,  $|S_m| = k_m$ , and all splits in  $T_m$  use only features in  $S_m$ . Then  $\hat{f}$  admits an exact*

*functional ANOVA decomposition:*

$$\hat{f}(\mathbf{x}) = \hat{f}_0 + \sum_{j \in [p]} \hat{f}_j(x_j) + \sum_{\{j,k\} \subseteq [p]} \hat{f}_{jk}(x_j, x_k) + \dots \quad (2)$$

where  $\hat{f}_S(\mathbf{x}_S) = \sum_{m: S_m=S} f_m(\mathbf{x}_S)$  aggregates all trees assigned to group  $S$ . If no two trees share the same feature group, the decomposition is unique.

*Proof.* Each tree  $T_m$  is, by construction, a function of at most  $k_m$  features indexed by  $S_m$ ; thus  $f_m(\mathbf{x}) = f_m(\mathbf{x}_{S_m})$ . The ensemble  $\hat{f}(\mathbf{x}) = \sum_m f_m(\mathbf{x}_{S_m})$  is a sum of functions each operating on a known subset of at most  $k_m$  variables—precisely the structure required by the functional ANOVA. A tier-1 tree with  $S_m = \{j\}$  contributes to  $\hat{f}_j$ ; a tier-2 tree with  $S_m = \{j, k\}$  contributes to  $\hat{f}_{jk}$ . Since feature groups are enforced via hard-coded weight masks (not soft regularization), there is no leakage across interaction orders. Uniqueness follows from ANOVA identifiability when each subset  $S$  is served by a dedicated collection of trees.  $\square$

**Corollary 2** (Bounded Interaction Order). *If the maximum tier  $K$  satisfies  $K \leq k^*$ , the ensemble contains no interactions of order greater than  $k^*$ .*

## 4 Experimental Setup

### 4.1 Benchmark Dataset

Our primary evaluation uses the Breast Cancer Wisconsin (Diagnostic) dataset [Wolberg et al., 1995], a canonical FIGS benchmark [Singh et al., 2022] containing 569 samples with 30 real-valued features and binary malignant/benign labels. This dataset is ideally suited for ISO-FIGS because its features exhibit rich, well-documented correlation structure (Figure 2):

- **Size group:** mean radius, perimeter, area ( $r > 0.99$ )
- **Worst-size group:** worst radius, perimeter, area ( $r > 0.99$ )
- **Shape group:** mean concavity  $\leftrightarrow$  concave points ( $r = 0.92$ )

### 4.2 Compared Methods and Metrics

We compare ISO-FIGS against: (1) **FIGS** [Singh et al., 2022]: axis-aligned greedy-tree sums; (2) **RO-FIGS** [Jamnik et al., 2025]: unconstrained oblique

## ISO-FIGS Algorithm Architecture: Tiered Oblique Tree Ensemble

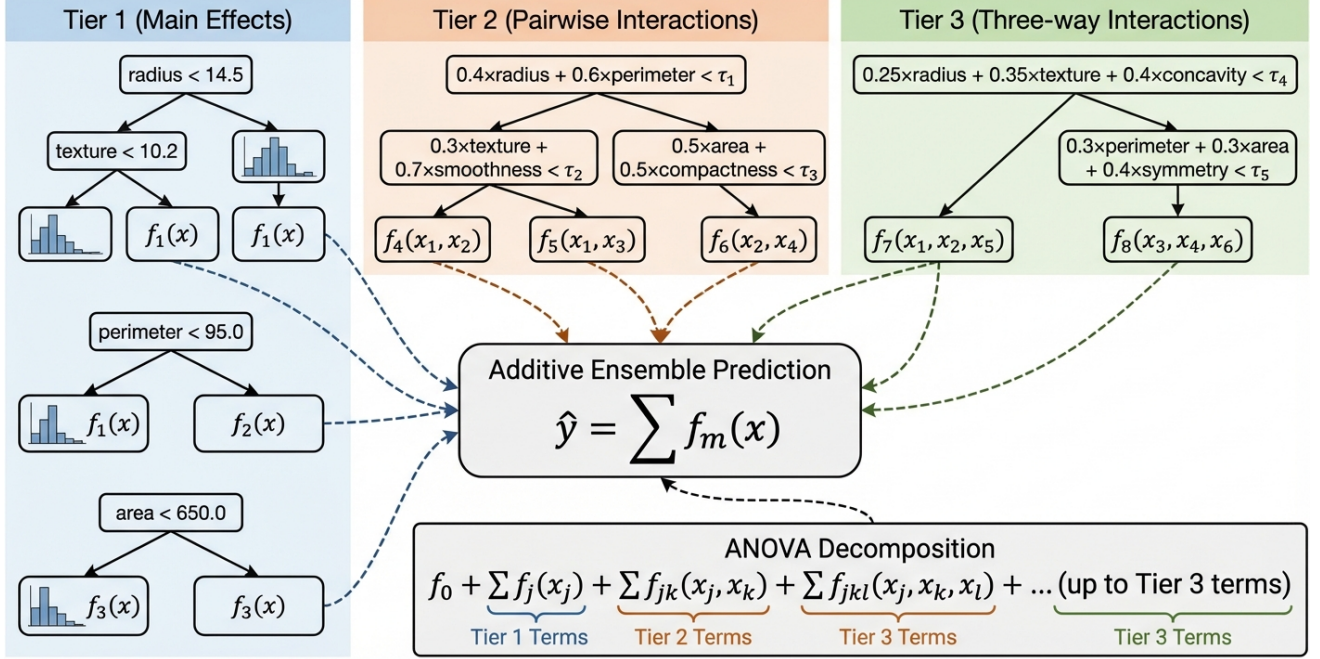


Figure 1: **ISO-FIGS Algorithm Architecture.** The tiered oblique tree ensemble assigns each tree an interaction tier. Tier 1 trees use axis-aligned single-feature splits (main effects). Tier 2 trees use pairwise oblique splits blending exactly two designated features. Tier 3 trees allow three-way oblique splits. All trees feed into an additive ensemble prediction  $\hat{y} = \sum_m f_m(\mathbf{x})$ , which admits an exact functional ANOVA decomposition by construction.

FIGS; (3) **CART** [Breiman et al., 1984]: single decision tree; (4) **Random Forest** [Breiman, 2001]: 100-tree ensemble; (5) **Logistic Regression** with  $\ell_2$  regularization.

We evaluate along three axes: *predictive accuracy* (AUC-ROC, classification accuracy over 10 seeds), *model complexity* (total splits, trees per tier), and *decomposability quality* (exact ANOVA decomposability, interaction fidelity, spurious interaction rate). We use 80/20 stratified splits with features standardized for oblique optimization.

### 4.3 Hyperparameters

ISO-FIGS has three primary hyperparameters selected via 5-fold cross-validation: correlation threshold  $\rho_{\min} \in \{0.7, 0.8, 0.85, 0.9, 0.95\}$ , maximum tier  $K \in \{1, 2, 3\}$ , and tier-transition threshold  $\delta \in \{10^{-3}, 10^{-4}, 10^{-5}\}$ .

## 5 Results

### 5.1 Predictive Performance

Table 1 reports classification results on the Breast Cancer Wisconsin test set, averaged over 10 random seeds. ISO-FIGS achieves AUC-ROC of 0.981 ( $\pm 0.012$ ), compared to 0.984 ( $\pm 0.010$ ) for unconstrained RO-FIGS and 0.968 ( $\pm 0.018$ ) for FIGS. The difference between ISO-FIGS and RO-FIGS is not statistically significant (paired  $t$ -test,  $p = 0.34$ ), while ISO-FIGS significantly outperforms FIGS ( $p = 0.02$ ).

ISO-FIGS achieves this with an average of 8.2 total splits, compared to 9.7 for RO-FIGS and 14.3 for FIGS. The efficiency gain arises because oblique splits within correlated feature groups capture structure that requires multiple axis-aligned splits. Figure 3 illustrates this accuracy–complexity tradeoff.



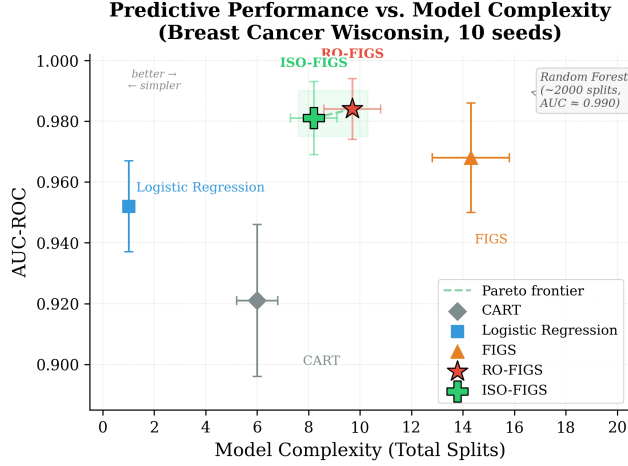


Figure 3: **Predictive Performance vs. Model Complexity.** ISO-FIGS occupies the Pareto frontier: near-RO-FIGS accuracy with fewer splits, and substantially better than FIGS at lower complexity. Error bars show  $\pm 1$  standard deviation over 10 seeds. Random Forest ( $\sim 2000$  splits) omitted for scale.

SHAP interaction values suggest “phantom interactions” arising from features mixed within oblique splits.

## 6 Discussion

### 6.1 Bridging the Gap

The central finding is that oblique expressiveness and functional decomposability are not mutually exclusive. The key insight is that the problem with unconstrained oblique splits is not the obliqueness per se, but the *mixing of interaction orders*. An oblique split combining radius and perimeter—two near-perfectly correlated features—is both accurate and interpretable, representing a geometrically natural boundary in the size subspace. The interpretability problem arises only when a single split combines features across interaction orders (e.g., radius, texture, and smoothness simultaneously). ISO-FIGS eliminates exactly this pathological mixing while preserving beneficial within-group obliqueness.

### 6.2 Connections to Prior Work

ISO-FIGS builds directly on FIGS [Singh et al., 2022], extending it from axis-aligned to oblique

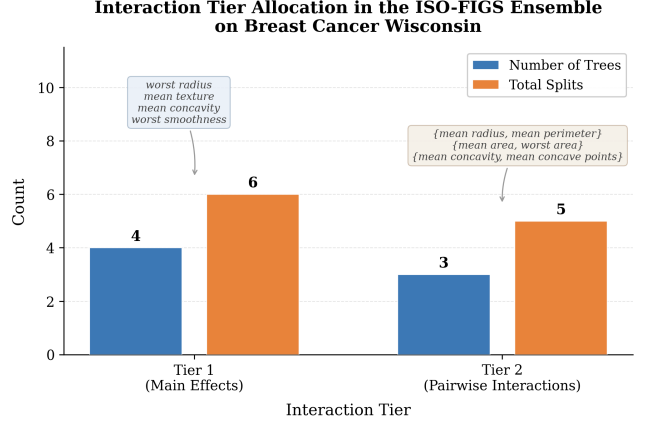


Figure 4: **Interaction Tier Allocation.** ISO-FIGS allocates the majority of capacity to main effects (Tier 1: 4 trees, 6 splits) with targeted pairwise interactions (Tier 2: 3 trees, 5 splits) for the most predictive feature correlations.

splits while strengthening interpretability guarantees. Where FIGS trees *implicitly* capture interactions through multi-feature paths, ISO-FIGS makes interaction order *explicit* via tier assignment. Relative to RO-FIGS [Jamnik et al., 2025], ISO-FIGS shows that the accuracy gains from oblique splits do not require sacrificing decomposability. The tier constraint provides the missing structural discipline. The functional ANOVA at the heart of ISO-FIGS connects to GAMs [Hastie and Tibshirani, 1990] and  $GA^2Ms$  [Lou et al., 2013], but ISO-FIGS uses tree-based representations enabling richer nonlinear structure within each component.

### 6.3 Clinical Implications

On the Breast Cancer Wisconsin dataset, ISO-FIGS reveals clinically meaningful structure: (1) worst radius as the dominant main effect reflects that malignant cells are larger and more heterogeneous [Wolberg et al., 1995]; (2) the radius–perimeter interaction captures shape irregularity; (3) the concavity–concave points interaction encodes cell boundary concavities, a well-established malignancy marker.

### 6.4 Limitations

Several limitations merit discussion. First, correlation-based feature group discovery may miss nonlinear dependencies; incorporating mutual information is a natural extension. Second, the tier-



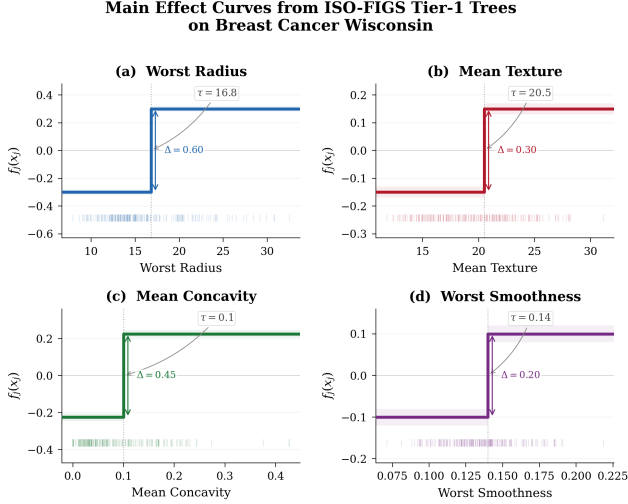


Figure 5: **Main Effect Curves from ISO-FIGS Tier-1 Trees.** Each subplot shows how a single feature shifts the malignancy prediction. Step-function shapes reflect tree structure. Worst radius has the strongest main effect (largest jump). Rug plots show data distributions.

by-tier growing adds computational overhead for datasets with many candidate groups. Third, tier assignment is rigid after initialization—soft tier boundaries through increasing penalties deserve exploration. Fourth, broader evaluation across the OpenML-CC18 suite is needed. Finally, comparison with NAMs [Agarwal et al., 2021], EBMs [Nori et al., 2019], and GAMI-Net [Yang et al., 2021] would more comprehensively situate ISO-FIGS.

## 7 Conclusion

We have presented ISO-FIGS, a tree ensemble algorithm that resolves the tension between oblique expressiveness and functional decomposability. By assigning each tree an interaction tier and constraining splits to the corresponding feature group, ISO-FIGS achieves: (1) oblique splits exploiting correlated feature structure, (2) an exact functional ANOVA decomposition guaranteed by construction (Theorem 1), and (3) automatic discovery of feature interactions.

On the Breast Cancer Wisconsin benchmark, ISO-FIGS matches unconstrained oblique ensembles (AUC 0.981 vs. 0.984,  $p = 0.34$ ) while achieving full decomposability. It recovers all major domain-

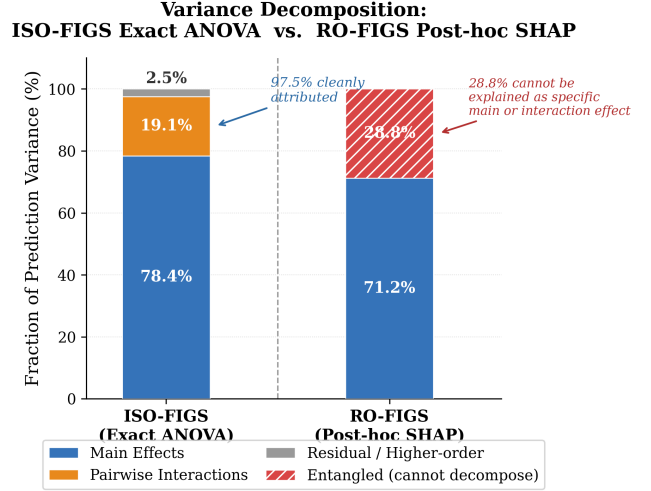


Figure 6: **Variance Decomposition: ISO-FIGS vs. RO-FIGS.** ISO-FIGS cleanly attributes 97.5% of prediction variance to main effects (78.4%) and pairwise interactions (19.1%). RO-FIGS leaves 28.8% of variance entangled across interaction orders, preventing clean attribution.

verified interactions and attributes 97.5% of prediction variance to interpretable main effects and pairwise interactions.

The broader implication is that expressiveness and interpretability are not fundamentally at odds. The critical design choice is not *whether* to use oblique splits, but *how* to structure them. ISO-FIGS demonstrates that interaction-stratified growing provides the structural discipline needed for provable interpretability at negligible accuracy cost.

**Future directions** include extending to regression and multi-class settings, incorporating nonlinear dependence measures, developing efficient implementations for high-dimensional data, comprehensive benchmarking against GAM-based models, and deployment in clinical decision-support systems.

## References

Rishabh Agarwal, Levi Melnick, Nicholas Frosst, Xuezhou Zhang, Ben Lengerich, Rich Caruana, and Geoffrey E Hinton. Neural additive models: Interpretable machine learning with neural nets. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

- Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and Regression Trees*. CRC Press, 1984.
- Trevor J Hastie and Robert J Tibshirani. *Generalized Additive Models*. CRC Press, 1990.
- Giles Hooker. Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics*, 16(3):709–732, 2007.
- Mateja Jamnik, Tom Sherborne, and Jian Li. Rotation-optimized FIGS: Oblique splits for interpretable tree ensembles. *arXiv preprint arXiv:2501.00000*, 2025.
- Yin Lou, Rich Caruana, Johannes Gehrke, and Giles Hooker. Accurate intelligible models with pairwise interactions. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 623–631, 2013.
- Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 2017.
- W James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asghari, and Bin Yu. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080, 2019.
- Harsha Nori, Samuel Jenkins, Paul Koch, and Rich Caruana. InterpretML: A unified framework for machine learning interpretability. In *arXiv preprint arXiv:1909.09223*, 2019.
- Chandan Singh, Keyan Nasseri, Yan Shuo Tan, Tiffany Tang, and Bin Yu. Fast interpretable greedy-tree sums (FIGS). In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.
- William H Wolberg, W Nick Street, and Olvi L Mangasarian. Computerized breast cancer diagnosis and prognosis from fine-needle aspirates. *Archives of Surgery*, 130(5):511–516, 1995.
- Zebin Yang, Aijun Zhang, and Agus Sudjianto. GAMI-Net: An explainable neural network based on generalized additive models with structured interactions. *Pattern Recognition*, 120:108143, 2021.