Artificial Intelligence Review (2025) 58:407
https://doi.org/10.1007/s10462-025-11339-y

Individual variable priority: a model-independent local gradient method for variable importance

Min Lu[1] · **Hemant Ishwaran**[1]

Accepted: 29 July 2025

Abstract
Traditional variable importance measures quantify overall feature contributions but often overlook individual-level heterogeneity. Several new procedures attempt to address this limitation but remain model dependent and may introduce biases. We propose individual variable priority (iVarPro), an extension of the Variable Priority (VarPro) framework, which uses rule-based, data-driven partitioning to estimate the gradient of the conditional mean function. By focusing on gradients, iVarPro captures how small perturbations in a variable influence an individual's outcome, providing a more precise and interpretable measure of importance. To demonstrate its advantages, we conducted simulations and analyzed a real-world survival dataset. Our results show that iVarPro more accurately captures the true functional relationship by flexibly leveraging local samples.

Keywords Conditional expectation · Local gradient (partial derivative) · Individual variable importance · Release region · Variable selection

1 Introduction

Understanding the individual importance of variables is a crucial yet underexplored aspect of statistical modeling and machine learning. Traditional methods quantify variable effects at the population level but often fail to capture heterogeneity across individuals. In many applications, particularly those requiring personalized decisions, it is insufficient to determine whether a variable is important on average; rather, we must understand how it influences specific predictions. For example, a machine learning model may predict a patient's risk of heart disease, identifying well-known clinical factors such as age, cholesterol, and smoking status as key predictors. Yet, for a specific patient, their elevated risk may be driven primarily by a biomarker, even if that biomarker has low average importance across the

✉ Hemant Ishwaran
hishwaran@miami.edu

Min Lu
m.lu6@umiami.edu

[1] Division of Biostatistics, Miller School of Medicine, University of Miami, Miami, USA

Published online: 05 November 2025

population. Such individual-level effects provide finer interpretability and are essential for tailoring treatment strategies and enabling precision decision-making.

A widely used approach for assessing variable importance is SHAP (Shapley Additive Explanations) (Lundberg and Lee 2017), which attributes feature contributions to model predictions using cooperative game theory. Unlike traditional population-level methods, SHAP provides both population- and individual-level assessments, making it a flexible tool for understanding local feature effects. Nevertheless, its interpretation remains model-dependent, as it relies on the model's internal feature interactions and the choice of background distribution.

Another approach, permutation importance, was originally introduced for random forests by Breiman (2001). It measures the change in prediction error when a variable is randomly permuted, providing insight into a feature's contribution to predictive performance. Although typically viewed as a population-level measure, permutation importance can also be adapted for individual-level assessments. In random forests, individual variable importance is determined by averaging the change in out-of-bag (OOB) prediction error for each instance across trees. Still, a key limitation of permutation importance is that it is inherently model-dependent and can introduce biases, particularly in the presence of correlated features (Strobl et al. 2007).

To further enhance individual importance estimation, methods such as LIME (Local Interpretable Model-Agnostic Explanations) (Ribeiro et al. 2016) and other localized techniques (Aas et al. 2021) have been developed. However, LIME perturbs feature values synthetically to approximate local model behavior, which can produce unrealistic data points and misrepresent true model dynamics. Case-Specific Random Forests (CSRF) (Xu et al. 2016) estimate individual variable importance using a leave-one-out forest approach, improving on permutation methods by assessing importance through prediction accuracy changes with local feature space partitions. Another method recently introduced by Dai et al. (2024) quantifies individual importance through mean squared error and local functional derivatives of the prediction function. Additionally, Winn-Nuñez et al. (2024) proposed a unified, prediction-based approach for local and global variable importance in nonlinear regression models, broadening applicability to complex modeling frameworks.

Despite their focus on individual-level interpretability, these methods share common limitations: they are largely model-dependent and can introduce biases. More broadly, prediction-based approaches may lack robustness across applications, as their reliability depends on the model's structure, feature dependencies, and, in some cases, the prediction error metric used to define importance.

Recent developments in feature effect estimation highlight the growing demand for local interpretability in applied machine learning. Methods such as Individual Conditional Expectation (ICE) curves (Goldstein et al. 2015) and Accumulated Local Effects (ALE) curves (Apley and Zhu 2020) allow practitioners to visualize the marginal impact of individual features on a prediction by varying that feature while holding others fixed or averaging out their effects. These approaches are particularly relevant in model explanation and counterfactual analysis, where understanding heterogeneity in predictions is critical. While our focus is not on counterfactual reasoning per se, the conceptual overlap is significant: both ICE/ALE aim to understand a feature's influence locally.

In previous work (Lu and Ishwaran 2024), we described Variable Priority (VarPro), a model-independent method designed to mitigate bias in variable importance estimation.

VarPro defines localized feature space regions using data-driven splitting rules and computes importance scores from these. By relying solely on observed data, VarPro avoids the biases introduced by permutations and artificial data and ensures a more robust approach to variable selection. Although VarPro addresses bias in population-level importance, it does not account for individualized effects. To bridge this gap, we introduce individual-VarPro (iVarPro), which shifts the focus to individual-level variable importance by explicitly estimating the gradient of the conditional mean function ψ and defining it as the individual variable importance value. This approach is particularly valuable in precision medicine, where subtle variations in patient characteristics can lead to markedly different prognoses. A large gradient indicates that even minor changes in a variable significantly impact an individual's prediction, whereas a near-zero gradient suggests minimal influence, regardless of its population-level significance.

The primary objective of this paper is to introduce and describe the iVarPro procedure and demonstrate its effectiveness, particularly in biomedical applications. Section 2 provides a detailed description of iVarPro, including its local linear regression approach for gradient estimation and methodological enhancements designed to improve stability and accuracy. Section 3 presents empirical evidence of iVarPro's capabilities. We begin with a regression simulation study evaluating its performance under complex gradient structures and correlated feature settings, comparing its effectiveness to existing procedures in both low- and high-dimensional settings. We further illustrate iVarPro's ability in a synthetic survival simulation. Section 4 applies iVarPro to a real-world survival dataset of patients who underwent treadmill exercise testing for suspected coronary artery disease. This serves as a real-world test case to distinguish variables with strong population-level effects from those with weaker but potentially meaningful individual contributions. Finally, Sect. 5 discusses our findings and their implications for precision medicine.

2 The iVarPro method

SHAP (Lundberg and Lee 2017) and LIME (Ribeiro et al. 2016) are widely used methods for understanding the effect of a feature on a model. However, in both approaches, a feature's importance value represents its contribution to the model's prediction for a specific instance relative to a baseline value, which may lack direct clinical relevance in precision medicine. Additionally, these methods assume local additivity in feature contributions, but their importance values do not imply direct effects. That is, a negative importance value for feature s in case i does not necessarily indicate that increasing feature s will decrease the outcome for case i.

To address these limitations, we propose iVarPro, which defines feature importance based on the local gradient of each feature, providing a more interpretable measure. The gradient quantifies how small perturbations in a variable influence an individual's predicted outcome, serving as a natural measure of sensitivity. Unlike SHAP, which fits a global prediction model, iVarPro constructs a local predictive model similar to LIME but estimates feature effects using local gradients rather than surrogate approximations.

A key challenge in local modeling is the small sample size, particularly in high-dimensional feature spaces. Increasing the sample size for fitting a model at a given instance i inevitably incorporates samples that are farther from i, reducing local relevance. To mitigate

this, we employ tree-based partitioning rules to define the local sample for instance *i*, selecting data points within the same terminal node of a decision tree. While this approach results in a small sample size, we introduce the notion of a *release region*, where the boundary constraint on feature *s* is relaxed when estimating its importance. This relaxation achieves two key objectives:

1. Increasing the effective sample size.
2. Enabling the estimation of the gradient of feature *s* using simple linear regression.

As will be explained shortly, an important aspect of iVarPro is that all other features remain constrained within the rule-defined local region, ensuring locality for instance *i* without requiring a multivariate regression model. Furthermore, this approach does not impose linearity assumptions on the remaining features. The key distinction between LIME and iVarPro is that iVarPro estimates feature importance using signficantly smaller local samples and computes feature gradients along one coordinate direction at a time.

In the following sections we describe the iVarPro method in detail. To set the stage, we first introduce some notation that will be helpful in laying out our proposed method.

Let $\psi(\mathrm{x}) = \mathbb{E}(\phi(Y)|\mathrm{x})$ be the unknown target function for the feature vector $\mathrm{x} = (x^{(1)}, \ldots, x^{(p)})^T$, where Y is the outcome and the choice of ϕ depends on the specific problem. Examples of ψ include the conditional mean in regression, class probabilities in classification, and survival probabilities in time-to-event analysis. Denote by $\mathrm{x}^{(S)} = \{x^{(s)}\}_{s\in S}$ the restriction of x to coordinates $s \in S$ for a set $S \subset \{1, \ldots, p\}$. The objective of population variable selection is to identify the smallest set of signal variables S_0 such that $\psi(\mathrm{x})$ depends only on $\mathrm{x}^{(S_0)}$ and is conditionally independent of the noise variables $\mathcal{N} = \{1, \ldots, p\} \setminus S_0$.

In contrast, when considering individual variable importance, the focus shifts to local behavior. Our approach is to examine the gradient of ψ, and our aim is to estimate:

$$\boldsymbol{\nabla}\psi(\mathrm{x}) = \left(\frac{\partial\psi(\mathrm{x})}{\partial x^{(1)}}, \ldots, \frac{\partial\psi(\mathrm{x})}{\partial x^{(p)}}\right)^T := \left(g^{(1)}(\mathrm{x}), \ldots, g^{(p)}(\mathrm{x})\right)^T.$$

The values $g^{(s)}(\mathrm{x})$ represent the individual variable importance for features $x^{(s)}$, $s = 1, \ldots, p$.

2.1 VarPro: using rules to obtain local gradients

Now we describe the iVarPro procedure in detail, beginning with a review of VarPro. A formal summary of the full iVarPro procedure is provided later in Algorithm 1.

VarPro defines feature space regions using data-driven rules, where local estimates of the target function are computed. The importance of a variable $s \in \{1, \ldots, p\}$ is assessed by comparing estimates within a rule's region R to those in its corresponding release region, defined as the region R expanded where constraints on $x^{(s)}$ are removed. These rules are constructed using random trees, with terminal nodes representing small rectangular regions R in the feature space (Lu and Ishwaran 2024). A key advantage of VarPro is its consistency property for selecting signal variables, ensuring that the constructed rules are based on features relevant to the target conditional mean ψ. This, in turn, enhances the ability to obtain

accurate gradient estimates, as the regions defined by these rules are more likely to reflect meaningful variations in ψ.

To harness this framework for iVarPro individual variable importance, we exploit the fact that each terminal node *R* provides a localized approximation of ψ. Since values within *R* are close with respect to a distance measure defined on the signal variables, we assume a common gradient value across all $\mathrm{x} \in R$. Specifically, let $\mathrm{x}_0 := \mathrm{x}_0(R)$ denote the centroid of *R*, and assign its gradient to all cases in *R*.

To define the gradient at $\mathrm{x}_0 = (x_0^{(1)}, \ldots, x_0^{(p)})^T$ along coordinate *s*, let $\psi_R^{(s)}(z)$ represent the evaluation of ψ at x_0, with the *s*-th coordinate perturbed by an increment *z*:

$$\psi_R^{(s)}(z) = \psi(x_0^{(1)}, \ldots, x_0^{(s-1)}, x_0^{(s)} + z, x_0^{(s+1)}, \ldots, x_0^{(p)}).$$

By construction, $\psi_R^{(s)}(0) = \psi(\mathrm{x}_0)$, and the gradient of ψ at x_0 along coordinate *s* is formally given by

$$g_R^{(s)} := g_R^{(s)}(\mathrm{x}_0) = \lim_{z \to 0} \frac{\psi_R^{(s)}(z) - \psi_R^{(s)}(0)}{z}.$$

Since ψ depends only on signal coordinates $s \in S_0$, the gradient is zero for noise coordinates $s \in \mathcal{N}$. This ensures that variable importance is derived solely from relevant features, preventing spurious contributions from irrelevant covariates.

2.2 Gradient estimation via local linear regression

To estimate the gradient, we employ a local linear regression model. In small regions around $\mathrm{x}_0(R)$ along the coordinate direction *s*, we approximate ψ using the linear model:

$$\psi_R^{(s)}(z) = \psi_R^{(s)}(0) + z g_R^{(s)}.$$

Given training data $(\mathrm{x}_1, y_1), \ldots, (\mathrm{x}_n, y_n)$, we regress $\{y_i\}$ on the *s*-coordinate features $\{x_i^{(s)}\}$ using an intercept-slope model based on data local to *R*. The gradient $g_R^{(s)}$ is estimated using the least squares estimate of the slope.

2.3 Removing spurious effects by using machine learning

To improve estimation accuracy, several key modifications are introduced. First, we replace $\{y_i\}$ with $\{\hat{\psi}(\mathrm{x}_i)\}$, where $\hat{\psi}$ represents any machine learning estimator of ψ. While iVarPro can operate directly on observed data, this substitution enhances gradient estimation by providing a more accurate response target, particularly in challenging settings.

2.4 Expanding the local sample: the release Region

For the local linear model assumption to hold, we must restrict the regression to a neighborhood of x_0. Using only observations within *R* is insufficient, as the sample sizes within these small regions are often too limited for stable estimation. To increase the sample size for local regression, we leverage the release region defined within the VarPro framework.

Releasing R along coordinate s expands the subset of individuals considered by removing constraints on $x^{(s)}$ while preserving all other constraints defining R. In this case, since the rule is rectangular, releasing on coordinate $x^{(s)}$ corresponds to removing the side of the rectangle in the direction of this coordinate (see left panel of Fig. 1). This targeted expansion introduces additional variation specifically along $x^{(s)}$, which is precisely what is needed to compute the directional derivative. Effectively, the release region, denoted $R^{(s)}$, consists of individuals who match those in R along all relevant coordinates $s' \neq s$ while expanding the available data along $x^{(s)}$, thereby stabilizing gradient estimation while avoiding artificial data generation.

2.5 Systematic expansion of the release region

To systematically control the extent of this expansion, we introduce an index $0 \leq \lambda \leq \lambda_0$ that determines the proportion of individuals selected from the release region $R^{(s)}$. Larger values of λ correspond to a greater number of included individuals. The expanded region is then defined as $R(s, \lambda) = R \cup R^{(s)}_{\lambda}$, where $R^{(s)}_{\lambda}$ is a restricted subset of $R^{(s)}$ determined by λ (see right panel of Fig. 1).

2.6 Standardized covariates for dimensionless gradient estimation

To further enhance the procedure, we standardize the variables within the expanded region $R(s, \lambda)$. Specifically, we center $x_i^{(s)}$ at the centroid value $x_0^{(s)}(R)$ and then scale it to have unit length, denoting the transformed values by $z_i^{(s)}$:

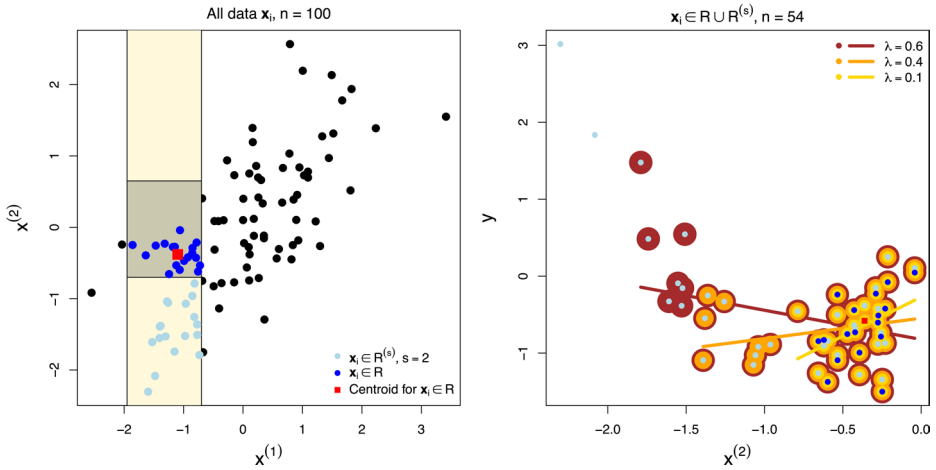

Fig. 1 Illustration of how iVarPro works. The left panel displays a region R (gray box) corresponding to a tree rule, where dark blue points represent observed data within R, and the centroid value is marked by a red square. Releasing variable $s = 2$ expands R to the release region $R^{(s)}$ (yellow area), where light blue points denote observed values in this expanded region. The right panel illustrates how the expansion region $R(s, \lambda) \subset R \cup R^{(s)}$, results in different data being incorporated as λ increases (dark blue points combining with light blue points), leading to distinct least-squares solutions (represented by lines). Cross-validation is employed to select the optimal λ

$$z_i^{(s)} = \frac{x_i^{(s)} - x_0^{(s)}(R)}{\sqrt{\sum_{\mathrm{x}_j \in R(s,\lambda)} (x_j^{(s)} - x_0^{(s)}(R))^2}}.$$

This transformation serves several purposes:

1. The value $z_i^{(s)} = 0$ corresponds to the centroid.
2. It ensures that the least-squares gradient estimator is standardized.
3. It enables a systematic expansion of $R(s, \lambda)$ from the centroid.

For instance, $\lambda = 1$ is calibrated so that it expands *R* by one standard deviation in $x^{(s)}$ from the centroid, which we adopt as the upper bound λ_0 in our analysis.

The local gradient is estimated via least squares regression using the data $\{z_i^{(s)}, \hat{\psi}(\mathrm{x}_i)\}$ within $R(s, \lambda)$:

$$\hat{\psi}(\mathrm{x}_i) = \alpha_R^{(s)} + z_i^{(s)} \beta_R^{(s)} + \varepsilon_i, \quad \mathrm{x}_i \in R(s, \lambda).$$

The least-squares slope is then used as the local gradient estimate: $\hat{g}_R^{(s)} := \hat{\beta}_R^{(s)}$.

2.7 Cross-validation for optimizing the fit

The expansion parameter λ controls the incorporation of the release region $R_\lambda^{(s)}$ into the local linear regression model. A smaller λ retains more localized data, preserving strong local assumptions, whereas a larger λ increases the sample size at the risk of including points that deviate from the local behavior of ψ. To select λ optimally, we use leave-one-out cross-validation based on the Predicted Residual Sum of Squares (PRESS) statistic.

For a given λ, the PRESS statistic is computed as follows. The leave-one-out residual for observation *i* is:

$$e_i = \hat{\psi}(\mathrm{x}_i) - \hat{f}^{(-i)}(z_i^{(s)}),$$

where $\hat{f}^{(-i)}(z_i^{(s)})$ is the least squares predicted value at $z_i^{(s)}$ when the *i*-th observation is excluded from the regression fit. The PRESS statistic is then given by:

$$\mathrm{PRESS}(s, \lambda) = \frac{1}{|R(s, \lambda)|} \sum_{i \in R(s,\lambda)} e_i^2.$$

Since local regression follows a linear model, the PRESS statistic simplifies in terms of leverage values h_i:

$$e_i = \frac{\hat{\psi}(\mathrm{x}_i) - \hat{f}(z_i^{(s)})}{1 - h_i},$$

where h_i is the leverage for $z_i^{(s)}$ and $\hat{f}$ is the least squares predictor. Thus, the PRESS statistic can be rewritten as:

$$\text{PRESS}(s,\lambda) = \frac{1}{|R(s,\lambda)|} \sum_{i \in R(s,\lambda)} \left(\frac{\hat{\psi}(\mathrm{x}_i) - \hat{f}(z_i^{(s)})}{1 - h_i} \right)^2 .$$

The optimal $\lambda^* = \text{argmin}_\lambda \text{PRESS}(s,\lambda)$ is selected, balancing the bias-variance trade-off in the local linear regression fit.

2.8 Averaging the gradient across rules

Recall that all cases within a rule are assigned the gradient for the centroid. To obtain a more stable estimate for a case x_i, we average its gradient values across all of its tree rules, denoted as $\mathcal{R}(\mathrm{x}_i)$, which is the set of all rules that contain x_i:

$$\hat{g}^{(s)}(\mathrm{x}_i) = \frac{1}{|\mathcal{R}(\mathrm{x}_i)|} \sum_{R \in \mathcal{R}(\mathrm{x}_i)} \hat{g}_R^{(s)}.$$

We call this value the individual importance value for x_i with respect to variable s.

2.9 Consistency of the gradient estimator

The iVarPro gradient estimator is based on local linear regression using observed values of the fitted prediction function $\hat{\psi}(\mathrm{x})$ over release regions defined by tree-based partitions. The consistency of local linear regression for estimating function derivatives is well-established under smoothness and design conditions (Fan J, Gijbels 1996; Ruppert and Wand 1994). To adapt these results to our setting, we consider the total estimation error of $\hat{g}_R^{(s)}(\mathrm{x})$ for $g_R^{(s)}(\mathrm{x})$ as arising from two sources: approximation error due to replacing the unknown regression function $\psi(\mathrm{x})$ with a fitted predictor $\hat{\psi}(\mathrm{x})$, and stochastic error from the local linear regression carried out within the expansion region $R(s,\lambda)$.

Suppose that $\psi(\mathrm{x})$ is continuously differentiable in a neighborhood of x, and that the predictor $\hat{\psi}(\mathrm{x})$ converges uniformly to $\psi(\mathrm{x})$ at rate $\sup_{\mathrm{x}} |\hat{\psi}(\mathrm{x}) - \psi(\mathrm{x})| = O_p(\eta_n)$. Then, by classical theory for local linear regression (Fan, Gijbels 1996), we obtain

$$\left| \hat{g}_R^{(s)}(\mathrm{x}) - g_R^{(s)}(\mathrm{x}) \right| = O_p\left(h^2 + \frac{1}{\sqrt{n_R h^3}} + \eta_n \right),$$

where h denotes the effective bandwidth of $R(s,\lambda)$ along coordinate s, and n_R is the number of samples in the expansion region. The term h^2 reflects the bias from local linear approximation, $1/\sqrt{n_R h^3}$ represents sampling variability under standard regularity conditions, and η_n quantifies the error from plugging in $\hat{\psi}$.

These results establish consistency of $\hat{g}_R^{(s)}(\mathrm{x})$ for $g_R^{(s)}(\mathrm{x})$, provided that $h^2 \to 0$, $n_R h^3 \to \infty$, and $\eta_n \to 0$ as $n \to \infty$. In terms of finite sample performance, we point out two things. First, that our PRESS-based tuning of the release region width λ is designed to balance the effect of the bias and the variance terms related to h, thus stabilizing estimation

by adapting to local curvature and noise. Secondly, the additive error term η_n appearing in the asymptotic expansion should not be interpreted as indicating that the use of an external predictor $\hat{\psi}$ necessarily worsens performance. In practice, replacing the raw response y with the smoothed function $\hat{\psi}(\mathrm{x})$ often reduces noise and yields improved stability and accuracy in gradient estimation, particularly in challenging problems. In Sect. 3, we will empirically evaluate the sensitivity and robustness of $\hat{g}_R^{(s)}$ to the choice of fitted model $\hat{\psi}$ and constrast it to the raw estimator using y.

Input: Training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ and machine learning predictor $\hat{\psi}$.

Step 1: Generate VarPro Rules

Obtain random tree terminal node regions $\{R\}$ using VarPro.

Identify signal variables S using population-level importance.

Step 2: Local Linear Regression for Gradient Estimation

for each feature $s \in S$ and rule R **do**

Extract data in R and define the release region $R^{(s)}$ by relaxing constraints on $X^{(s)}$.

Let $R_\lambda^{(s)} \subseteq R^{(s)}$ be the release subset within λ standard deviations of $x_0^{(s)}(R)$.

Expand R to $R(s,\lambda) = R \cup R_\lambda^{(s)}$, controlling expansion via λ.

Obtain $z_i^{(s)}$ by centering $x_i^{(s)}$ at $x_0^{(s)}(R)$ and standardizing to unit length.

Using the standardized features $z_i^{(s)}$, fit a local linear regression model:

$$\hat{\psi}(\mathbf{x}_i) = \alpha_R^{(s)} + z_i^{(s)}\beta_R^{(s)} + \varepsilon_i, \quad \mathbf{x}_i \in R(s,\lambda).$$

Perform leave-one-out cross-validation using the PRESS statistic:

$$\mathrm{PRESS}(s,\lambda) = \frac{1}{|R(s,\lambda)|} \sum_{i \in R(s,\lambda)} \left(\frac{\hat{\psi}(\mathbf{x}_i) - \hat{f}(z_i^{(s)})}{1 - h_i} \right)^2,$$

where $\hat{f}(z) = \hat{\alpha}_R^{(s)} + z\hat{\beta}_R^{(s)}$ is the least squares predictor.

Obtain $\hat{g}_R^{(s)} = \hat{\beta}_R^{(s)}$ for the optimized $\lambda^* = \mathrm{argmin}_\lambda \mathrm{PRESS}(s,\lambda)$

Assign all $\mathbf{x}_i \in R$ the value $\hat{g}_R^{(s)}$.

end for

Step 3: Compute Final Individual Importance Values

For a case $\mathbf{x}_i$, average its gradient estimates across all of its tree rules $\mathcal{R}(\mathbf{x}_i)$:

$$\hat{g}^{(s)}(\mathbf{x}_i) = \frac{1}{|\mathcal{R}(\mathbf{x}_i)|} \sum_{R \in \mathcal{R}(\mathbf{x}_i)} \hat{g}_R^{(s)}.$$

Output: Return $\hat{g}^{(s)}(\mathbf{x}_i)$ as the individual importance value for $\mathbf{x}_i$ with respect to variable s for $s = 1, \ldots, p$.

Algorithm 1 iVarPro: Individual Variable Priority

3 Synthetic experiments

In this section, we evaluate the empirical performance of iVarPro through a series of experiments designed to assess its effectiveness in different settings. We begin with a regression simulation study to examine iVarPro's ability to estimate local gradients under complex gradient structures and correlated feature settings. We then investigate its performance in a survival simulation, demonstrating its ability to identify variables with localized effects.

3.1 Benchmark regression study

To quantitatively assess the performance of iVarPro, we first apply a benchmark study where the true conditional mean and gradient are known. Two distinct simulation settings were used. In the first, all features were independently drawn from a Uniform(0, 1) distribution. In the second, dependence was introduced between features by applying a normal copula to impose a common correlation $\rho = 0.8$ among all predictors. For both settings, the response depends only on the first two predictors, while additional covariates act as noise variables. The model computed was:

$$y = \psi(x^{(1)}, x^{(2)}) + \epsilon, \quad \epsilon \sim N(0, 1).$$

We generated datasets with $n = \{250, 2000\}$ observations and a feature space of dimension $p = \{10, 200\}$. This experimental setup allows us to assess how well iVarPro recovers individual-level importance in both low- and high-dimensional settings, as well as under different levels of feature correlation and sample size configurations.

Three different ψ models were considered:

- Model 1 is generated according to a piecewise function:

$$\psi(x^{(1)}, x^{(2)}) = \begin{cases} 1, & x^{(2)} \leq 0.25, \\ 15x^{(2)}, & x^{(1)} \leq 0.5 \text{ and } x^{(2)} > 0.25, \\ 7x^{(1)} + 7x^{(2)}, & x^{(1)} > 0.5 \text{ and } x^{(2)} > 0.25. \end{cases}$$

- Model 2 is a quadratic function of the predictors, defining the response as a function of the squared Euclidean norm:

$$\psi(x^{(1)}, x^{(2)}) = 5r1_{\{r \leq 0.5\}}, \quad \text{where } r = (x^{(1)})^2 + (x^{(2)})^2.$$

- Model 3 follows a simple multiplicative interaction between the two predictors:

$$\psi(x^{(1)}, x^{(2)}) = 6x^{(1)}x^{(2)}.$$

The top row of Fig. 2a displays the 3 functions. The partial gradient with respect to $x^{(1)}$ and $x^{(2)}$ are displayed in rows (b) and (c), respectively. For example, for model 1, displayed on the extreme left of (a), there are three distinct regions in the feature space: a constant region where $x^{(2)} \leq 0.25$, a linear region with respect to $x^{(2)}$ when $x^{(1)} \leq 0.5$ and $x^{(2)} > 0.25$, and a bivariate linear region when $x^{(1)} > 0.5$ and $x^{(2)} > 0.25$. Correspondingly the gradi-

ent for $x^{(1)}$, displayed on the left of (b), is nonzero only in the bivariate linear region where $x^{(1)} > 0.5$ and $x^{(2)} > 0.25$. For $x^{(2)} > 0.25$, the gradient for $x^{(2)}$ shown on the left of (c) takes values of 15 when $x^{(1)} \leq 0.5$ and 7 when $x^{(1)} > 0.5$, reflecting a shift in the relative importance of $x^{(2)}$ depending on the value of $x^{(1)}$. Outside these regions, the function remains constant, resulting in a gradient of zero.

3.1.1 Implementing iVarPro

Algorithm 1 was used to implement iVarPro. For the external estimator $\hat{\psi}$ required by iVarPro, we used random forests to obtain an out-of-bag prediction for ψ. Rule generation and population-level variable selection were performed using the R-package `varPro`, available at https://github.com/kogalur/varPro. Random forests were implemented via the CRAN R-package `randomForestSRC` (Ishwaran and Kogalur 2025).

3.1.2 Comparison procedures

We compare iVarPro to three established methods for individualized feature importance:

- CSRF importance (Xu et al. 2016), implemented using the `ranger` R package (Wright and Ziegler 2017).
- SHAP, computed using the `treeshap` R package (Kozminski et al. 2024).
- LIME, computed using the `lime` R package (Pedersen and Benesty 2022).

Both SHAP and LIME were applied to a trained XGBoost model (Chen and Guestrin 2016), with hyperparameters selected via cross-validation.

We also evaluate three additional versions of iVarPro based on different choices of the external predictor $\hat{\psi}$, in order to gauge the robustness of the gradient estimator to the choice of fitted model:

- iVarPro-raw: No external model is used; local gradients are estimated directly using the raw response values y.
- iVarPro-gbm: The external predictor $\hat{\psi}$ is a gradient tree boosted model fit using the `gbm` R package (Greenwell et al. 2020).
- iVarPro-xgb: The external predictor $\hat{\psi}$ is an XGBoost model.

3.1.3 Performance evaluation and metrics

To evaluate the performance of the various procedures, we considered how well they recovered the gradient. The gradient of ψ serves as a natural gold standard because it quantifies the local rate of change of ψ with respect to each feature. An effective feature importance method should capture not only whether a variable is influential, but also the relative strength of its effect across instances. Since the gradient encodes this information, it provides a principled basis for comparison.

We employed ranking-based metrics, recognizing that the methods under study are not necessarily designed to estimate the gradient directly. Rather, it is sufficient that they correctly preserve the relative ordering of feature importance across dimensions. Using absolute

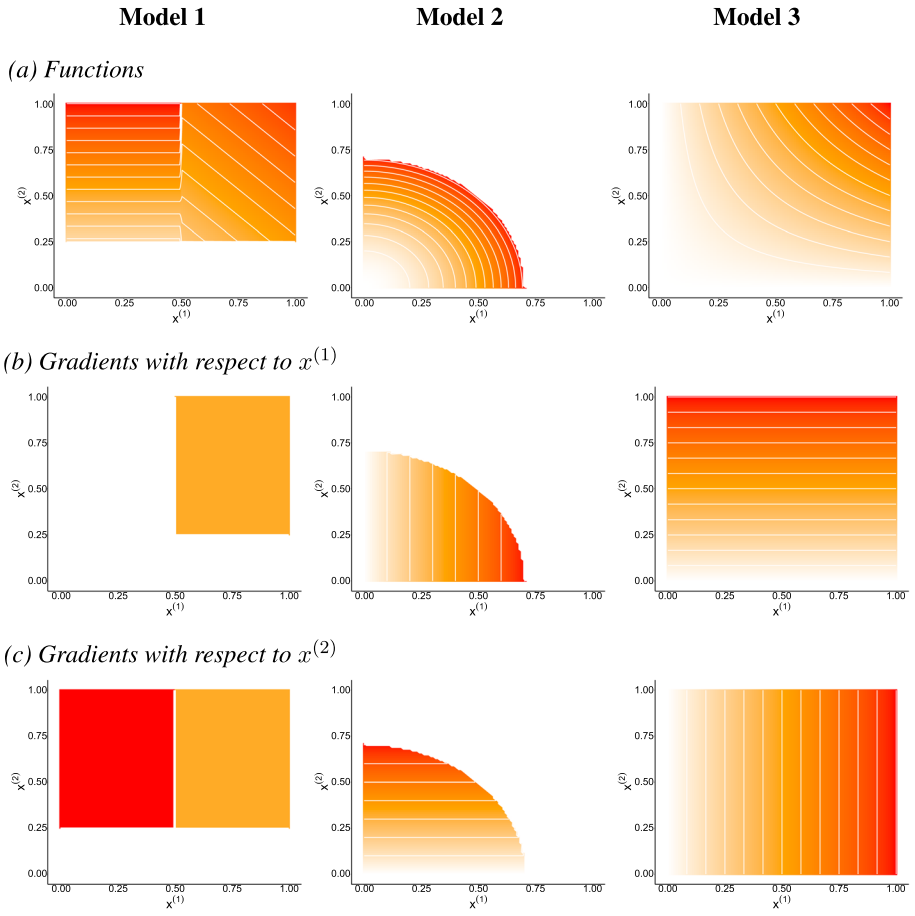

Fig. 2 Contour plots illustrating the simulation functions and their gradients. Darker regions indicate higher values, while white represents the lowest value (zero). **a** Top row: True function values for the three simulation models described in the text. **b** Middle row: Gradients of the functions with respect to $x^{(1)}$. **c** Bottom row: Gradients with respect to $x^{(2)}$

values for both the estimated and true gradients, we evaluated agreement using Kendall's τ_b, which measures the strength of monotonic association between two rank vectors while adjusting for ties. The metric was applied at the individual level by comparing each procedure's p-dimensional vector of predicted values to the true gradient vector for that instance. Final results were obtained by averaging the metric across all individuals.

In addition to Kendall's τ_b, we report two complementary ranking-based metrics: the concordance index (C-index) (Harrell et al. 1982) and the area under the precision-recall curve (PR-AUC) (Saito and Rehmsmeier 2015). The C-index generalizes Kendall's tau and measures the probability that, for a randomly selected pair of features, the estimated importance correctly ranks the true gradient magnitudes. Precision-recall analysis is particularly relevant in our setting due to the potential sparsity of gradient vectors. To apply it, we binarized the true gradient vector by labeling features as "positive" if their values exceeded the 25th percentile and "negative" otherwise.

3.1.4 Results

The simulations were repeated independently 100 times for each experimental condition. Results are summarized in Figs. 3, 4, and 5, which display performance across the three evaluation metrics: Kendall's τ_b, concordance index (C-index), and area under the precision-recall curve (PR-AUC), respectively. Each figure contains two panels: the left panel corresponds to the uncorrelated predictor setting, where features were sampled independently from a Uniform(0, 1) distribution; the right panel corresponds to the correlated setting, where dependence among features was introduced using a normal copula with pairwise correlation $\rho = 0.8$. Red error bars represent 95% confidence intervals for the mean and are slightly offset for clarity.

- iVarPro demonstrates the strongest overall performance, with its advantage becoming more pronounced as the sample size increases. This effect is especially clear for Kendall's τ_b and the C-index, which directly reflect the ranking of true gradient values. While the PR-AUC metric requires binarization and is therefore more sensitive to thresholding, iVarPro still shows a consistent upward trend.
- iVarPro maintains strong performance in high-dimensional settings, particularly as the sample size increases.
- These performance patterns remain consistent across both uncorrelated and correlated experimental conditions.

Because iVarPro explicitly targets the estimation of the gradient vector, we also evaluated its accuracy using mean squared error (MSE), which directly compares the estimated gradients to the true gradients. This metric is appropriate for iVarPro, as it produces scale-matched gradient estimates, unlike comparison methods that output relative importance scores. Results are shown in Fig. 6 and summarized below:

- The three procedures using externally fitted predictors (iVarPro, iVarPro-gbm, and iVarPro-xgb) performed similarly across all experimental settings, indicating that iVarPro is robust to the choice of standard machine learning predictor.
- In contrast, iVarPro-raw, which uses the raw outcome y without any smoothing, exhibited consistently higher MSE for Models 2 and 3 in the correlated settings, indicating a lack of robustness to noise and feature dependence. In Model 3, its error was so large that it exceeded the range of the plot and does not appear in the displayed results.

3.1.5 Visual comparison

To visually compare how the different procedures performed, we present results from a single run of the benchmark study using Model 2 for uncorrelated features with $d = 10$ and $n = 2000$, a setting where the comparison procedures performed reasonably well. Individual variable importance values for each procedure are shown in Fig. 7. The vertical and horizontal axes correspond to the observed values of $(x^{(1)}, x^{(2)})^T$, while point sizes are scaled according to each procedure's importance values. The left and right sides of the plots correspond to $x^{(1)}$ and $x^{(2)}$, respectively.

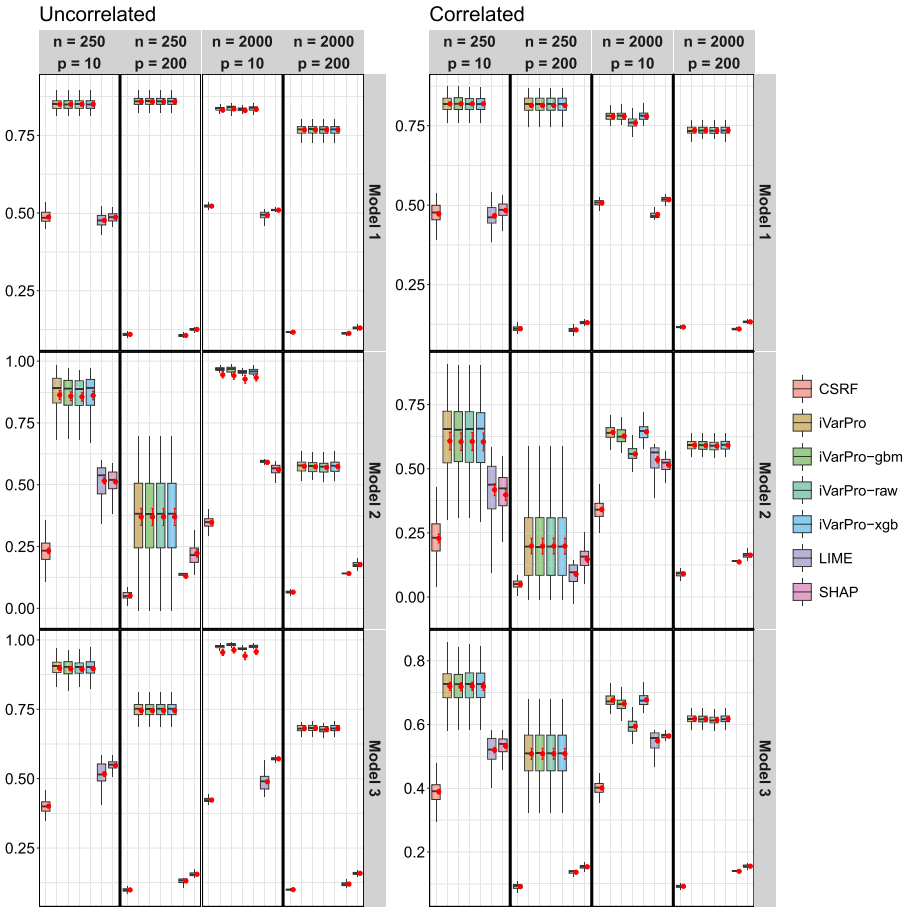

Fig. 3 Kendall's τ_b scores evaluating the performance of iVarPro and comparison procedures in a synthetic setting. Red error bars denote 95% confidence intervals for the mean, slightly offset to the right for visual clarity. The left panel corresponds to the uncorrelated setting, where features are independently drawn from a Uniform(0, 1) distribution. The right panel shows results for the correlated setting, where dependence among predictors is introduced via a normal copula with common correlation $\rho = 0.8$

Comparing these results to the true gradient (middle column of the second and third rows (b) and (c) of Fig. 2), we observe that iVarPro more accurately captures the true gradient structure, in agreement with our benchmark findings.

To gain further insight into iVarPro, we refer to Fig. 8. The figure displays data points from the simulation (black points) with superimposed tree rules R used by iVarPro to construct its estimator. These rules correspond to small rectangular regions, representing the terminal nodes of random trees, and are color-coded to indicate the estimated gradient. Notice how the rules blanket the feature space, enabling iVarPro to form an accurate estimator. The top panels show the true gradient, while the bottom panels display the gradient values estimated by iVarPro. A strong agreement between the two highlights iVarPro's effectiveness in capturing the underlying gradient structure.

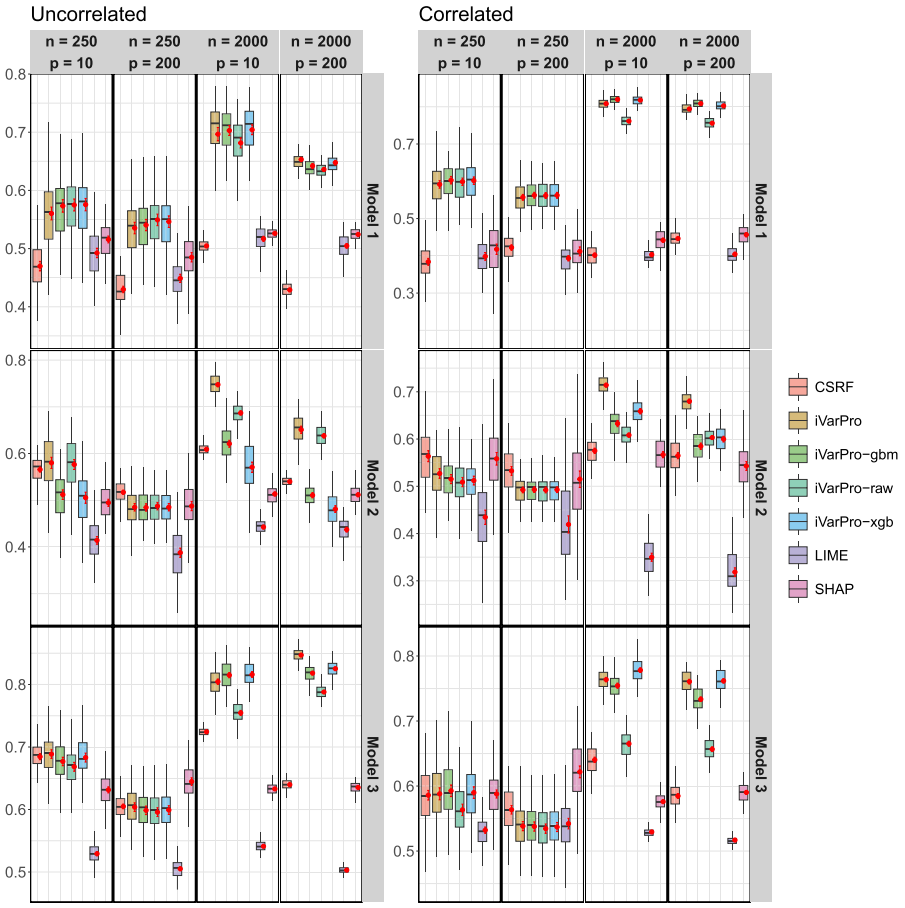

Fig. 4 C-index scores evaluating the performance of iVarPro and comparison procedures in a synthetic setting

Figure 9 displays the same type of figures for Model 2, but under the correlated setting ($\rho = 0.8$). Despite this challenging scenario, iVarPro effectively captures the underlying gradient structure, demonstrating its robustness in the presence of feature dependence.

3.2 Synthetic survival example

In the next example, we simulated a survival setting where the true (potentially unobserved) survival time was given by

$$T^o = \log\left[1 + V + \exp\sum_{s=1}^{4} \beta_s X^{(s)} + \beta_5 X^{(5)} 1_{\{X^{(2)} > 0.5\}}\right).$$

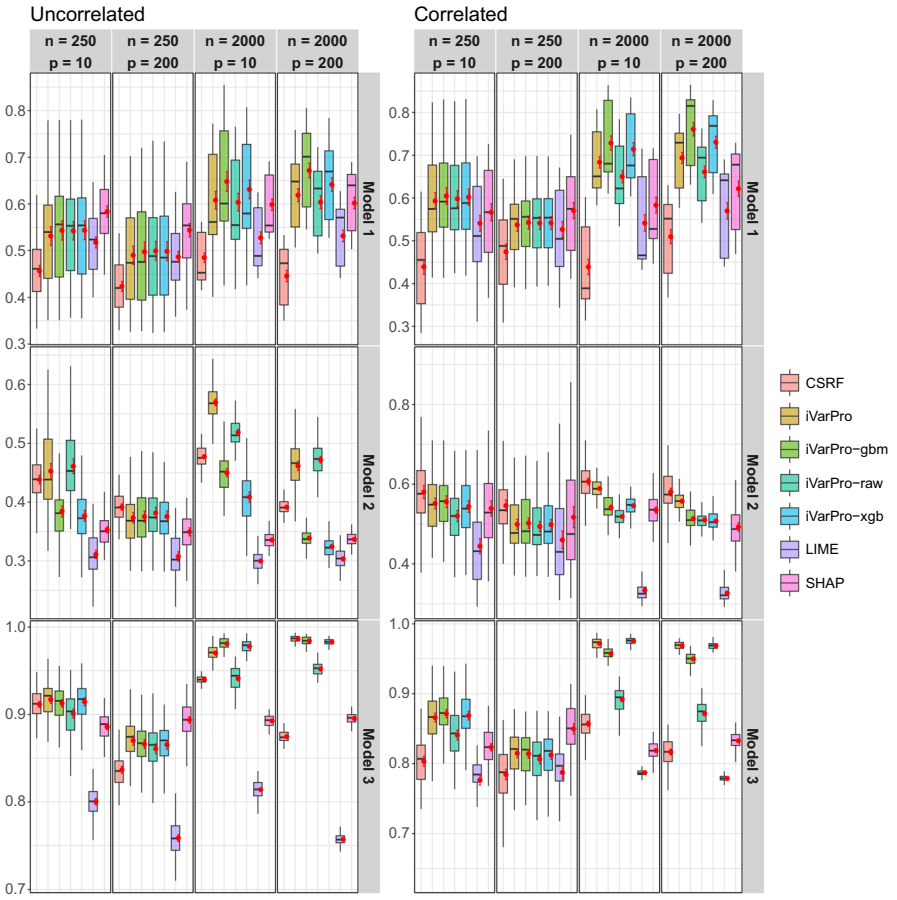

Fig. 5 PR-AUC scores evaluating the performance of iVarPro and comparison procedures in a synthetic setting

The variable V follows a standard exponential distribution and is sampled independently of the features $X^{(s)}$, which are drawn independently from a Uniform(0, 1) distribution.

We take $\psi(\mathrm{x})$ to be the restricted mean survival time (RMST) (Irwin 1949; Andersen et al. 2004; Royston and Parmar 2011; Kim et al. 2017). The RMST provides a meaningful summary of survival and is defined as the integral of the survival function up to a specified time horizon $\tau > 0$:

$$\psi(\mathrm{X}) = \int_0^\tau S(t|\mathrm{X})\, dt,$$

where the survival function is given by

$$S(t|\mathrm{x}) = \mathbb{P}\{T^o > t|\mathrm{X} = \mathrm{x}\}.$$

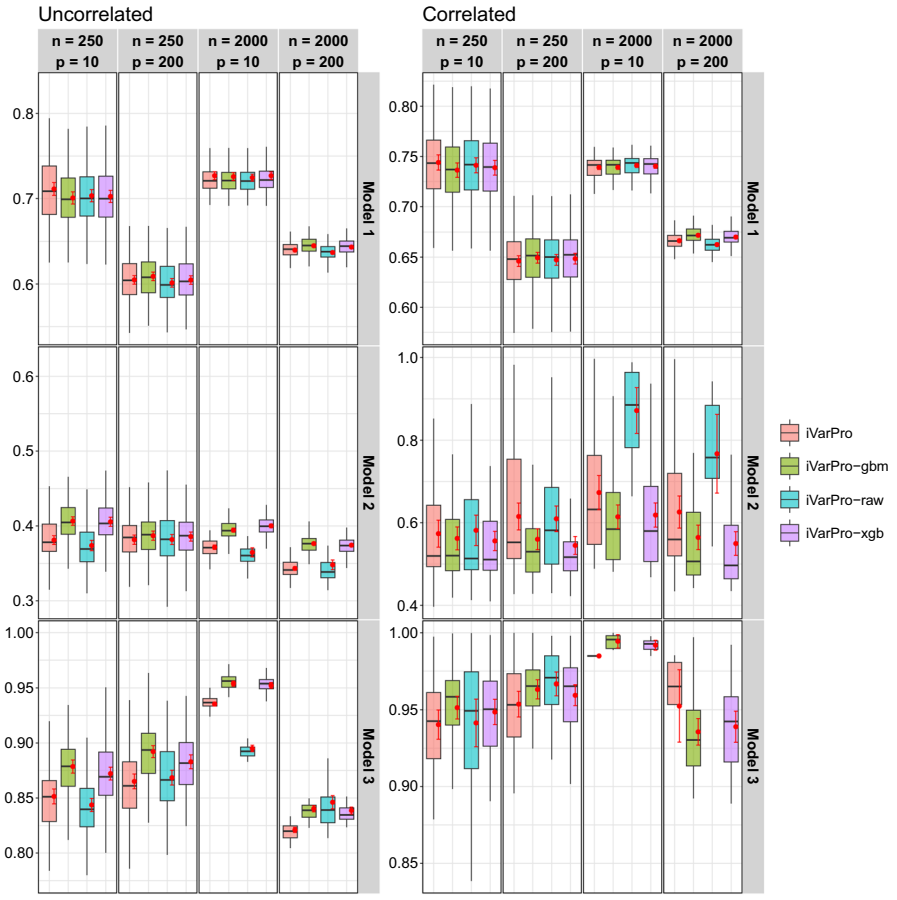

Fig. 6 MSE scores comparing the performance of the different iVarPro procedures. For Model 3 under the correlated setting, the MSE for iVarPro-raw was significantly large, exceeding the plotted range and therefore not visible in the displayed results

In this example, we can derive a closed-form expression for the gradient of $\psi(\mathrm{x})$, allowing us to exactly assess the performance of iVarPro. We have

$$S(t|\mathrm{x}) = \mathbb{P}\left\{\log[1 + V + A(\mathrm{x})] > t\right\},$$

where

$$A(\mathrm{x}) = \exp\left(\sum_{s=1}^{4} \beta_s x^{(s)} + \beta_5 x^{(5)} 1_{\{x^{(2)} > 0.5\}}\right).$$

Since V follows a standard exponential distribution, its survival function is given by

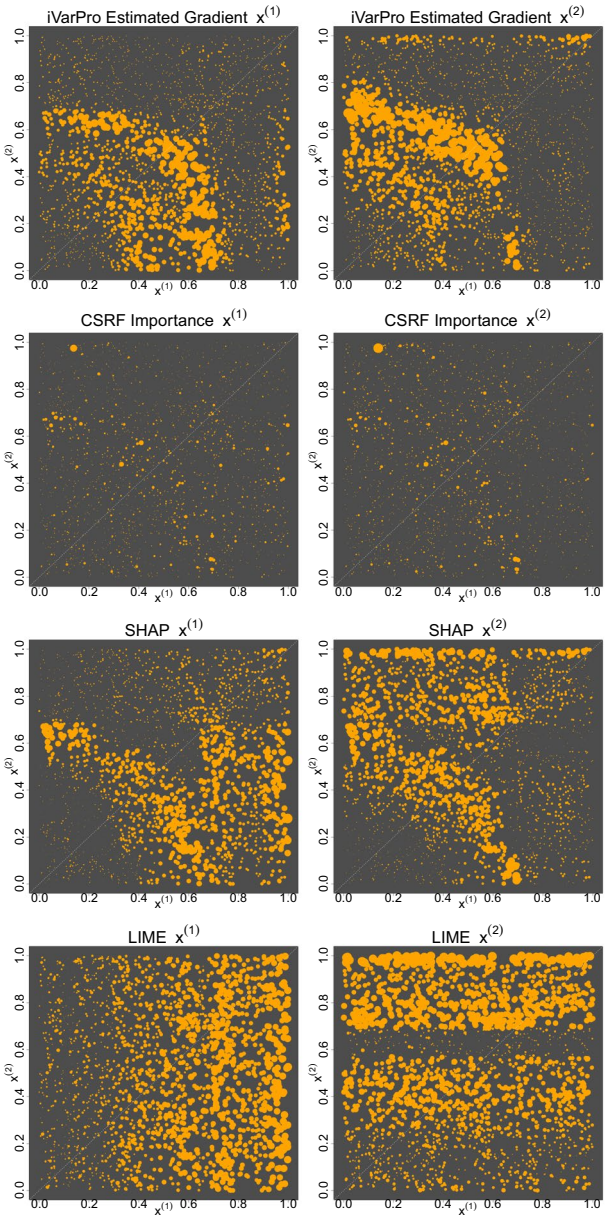

Fig. 7 Individual variable importance values for each procedure in Model 2 with uncorrelated features ($d = 10$, $n = 2000$). The vertical and horizontal axes correspond to $(x^{(1)}, x^{(2)})$, with point sizes scaled to the procedure's importance values. The left and right sides of the plots correspond to $x^{(1)}$ and $x^{(2)}$, respectively. Compared to the true gradient (middle column of (b) and (c) of Fig. 2), iVarPro more accurately captures the underlying gradient structure

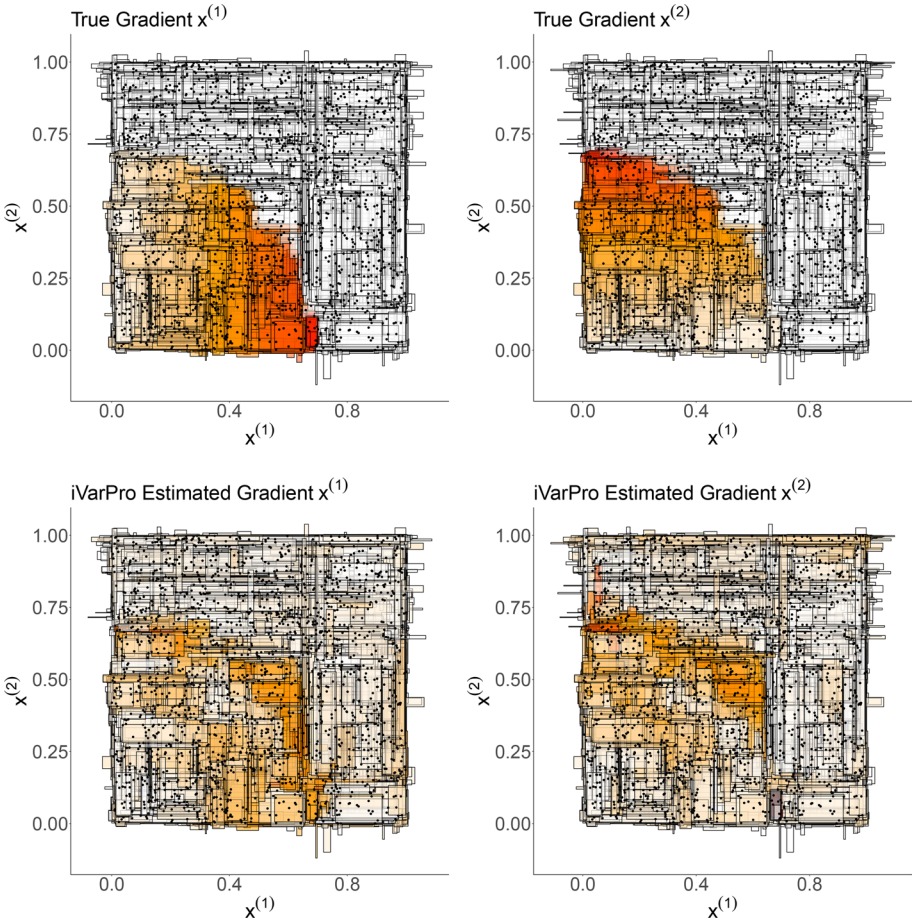

Fig. 8 Tree rules and gradient estimates from iVarPro for Model 2. The figure displays simulation data points (black) along with tree rules R (colored rectangles), which represent terminal nodes from random trees. Rule colors indicate the estimated gradient. The top panels show the true gradient, while the bottom panels display iVarPro's gradient estimates. The close agreement between the top and bottom panels highlights iVarPro's ability to accurately capture the gradient structure

$$\mathbb{P}\{V > v\} = e^{-v}.$$

Thus, the survival function simplifies to

$$S(t|\mathrm{x}) = \mathbb{P}\left\{V > e^t - 1 - A(\mathrm{x})\right\} = e^{A(\mathrm{x})}\lambda(t), \qquad \text{where } \lambda(t) = \exp(-e^t + 1)$$

and the RMST becomes

$$\psi(\mathrm{x}) = e^{A(\mathrm{x})}\Lambda(\tau), \qquad \text{where } \Lambda(\tau) = \int_0^\tau \lambda(t)dt.$$

Differentiating, we obtain the partial derivatives:

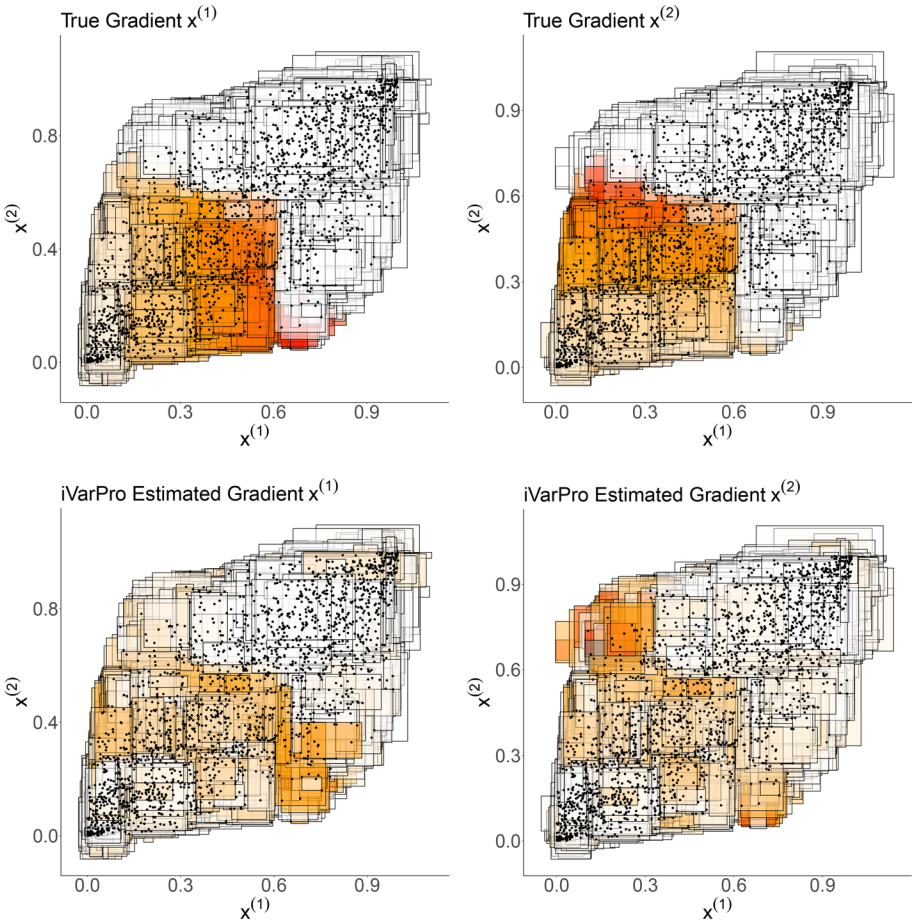

Fig. 9 iVarPro gradient estimates under correlated features. The figure displays results for Model 2 with features correlated at $\rho = 0.8$. The top panels show the true gradient, while the bottom panels display iVarPro's estimated gradient. Despite the feature dependence, iVarPro closely matches the true gradient, highlighting its robustness in capturing the underlying structure even in correlated settings

$$\begin{aligned} \frac{\partial \psi(\mathrm{x})}{\partial x^{(s)}} &= \beta_s A(\mathrm{x}) e^{A(\mathrm{x})} \Lambda(\tau), \quad s = 1, 2, 3, 4 \\ \frac{\partial \psi(\mathrm{x})}{\partial x^{(5)}} &= \beta_5 A(\mathrm{x}) e^{A(\mathrm{x})} \Lambda(\tau) 1_{\{x^{(2)} > 0.5\}}. \end{aligned}$$

Notice that the gradient for variables $x^{(1)}, \ldots, x^{(4)}$ depends on all signal variables, including $x^{(5)}$. These variables serve as population variables, meaning they have a strong population-level effect that would be immediately identified in a large proportion of the target population. On the other hand, $x^{(5)}$ has a smaller effect that is modulated by $x^{(2)}$ and influences survival only within a specific subpopulation—namely, individuals for whom $x^{(2)} > 0.5$. In particular, the gradient for $x^{(5)}$ is nonzero only in this subregion, highlighting its localized effect.

3.2.1 Results

Parameters were $\beta_s = 2$ for $s = 1, \ldots, 5$, and additionally, 15 noise variables were included, resulting in a total of $p = 20$ features. Random censoring at a 25% rate was applied, and a total of $n = 2000$ data points were sampled.

Since the effect of τ acts only as a global scaling parameter to the gradient, its value is not critical. However, for concreteness, we set τ to the largest follow-up time for our time evaluation point. The survival function for determining the RMST was estimated using random survival forests (RSF) (Ishwaran et al. 2008).

Figure 10 displays the individual importance values for $x^{(s)}$, $s = 1, \ldots, 4$, compared to $x^{(5)}$, with point sizes in the figure scaled according to the importance values of $x^{(5)}$. For the plot of $x^{(2)}$ versus $x^{(5)}$, we observe that individual importance is only positive in the subregion $x^{(2)} > 0.5$, agreeing with the true gradient calculations previously derived.

4 Precision survival analysis using ECG and clinical data

As our next illustration, we apply iVarPro to a dataset previously analyzed in Gorodeski et al. (2009). This dataset originates from a large cohort of 18,964 patients who underwent treadmill exercise testing for the evaluation of suspected coronary artery disease. Notably, all patients had a clinically normal resting electrocardiogram (ECG) and no known history of cardiovascular disease at the time of testing. The primary outcome of interest was all-cause mortality. Over a median follow-up of 10.7 years (range for survivors: 5-17 years), 1,585 patients (8%) died.

A key aspect of this dataset is the diverse nature of its covariates, which include over 150 features spanning both clinical and electrocardiographic (ECG) measures. Clinical variables encompass demographic and medical history factors such as age, gender, diabetes status, hypertension, smoking history, and exercise-related parameters including exercise capacity and heart rate recovery. These variables are well-established predictors of cardiovascular outcomes. In contrast, the ECG-derived features provide finer physiological details, including quantitative measures related to heart rate, conduction, repolarization, and left ventricular mass (Gorodeski et al. 2009). While ECG variables have been identified as potential prognostic markers in various populations, their relative importance compared to clinical variables remains uncertain, particularly in patients with clinically normal resting ECGs.

This setting presents an ideal test case for iVarPro, as it reflects a common challenge in biomedical research: distinguishing variables with strong population-level effects from those with weaker, but potentially meaningful, patient-specific contributions. Clinical variables are expected to exhibit higher overall importance across the cohort, whereas ECG measures may carry individualized relevance that standard population-based methods may not fully capture. By applying iVarPro, we can better quantify the extent to which ECG-derived variables contribute to risk at the individual level.

4.1 Results

For ψ, we used the integrated cumulative hazard function (CHF), with RSF employed to estimate ψ. The integrated CHF can be interpreted as a mortality value, representing, for a

given patient, the expected number of individuals in the dataset who would be expected to die if they shared similar characteristics.

As expected, the top population-level variables identified were clinical variables, including Peak METs and the Duke Treadmill Score (DTS). These findings are not unexpected: Peak METs (Metabolic Equivalents) quantifies exercise capacity by measuring the maximum level of oxygen consumption during exertion. It is a strong prognostic indicator of cardiovascular fitness, with higher values generally associated with lower mortality risk (Vivekananthan et al. 2003). The Duke Treadmill Score (DTS) (Mark et al. 1991), though often classified as a clinical measure, also incorporates ECG-based components. DTS is a well-established prognostic index derived from exercise stress testing and integrates multiple factors, including exercise duration, ST-segment depression, and the presence of angina during exertion. Because it combines both clinical and ECG-derived features, DTS serves as a comprehensive cardiovascular risk measure. Higher DTS values are generally associated with a lower risk of cardiovascular events and mortality.

In addition to these predictors, several ECG-specific variables also emerged as informative, including the primary lead ST-segment value. To assess the individual importance of this variable, we contrast it with Peak METs. Figure 11 displays the observed values of these two variables, stratified by patient mortality. In the top panel, point sizes are scaled to reflect the magnitude of the estimated gradient for Peak METs, while in the lower panel, point sizes correspond to the gradient for the ST-segment value.

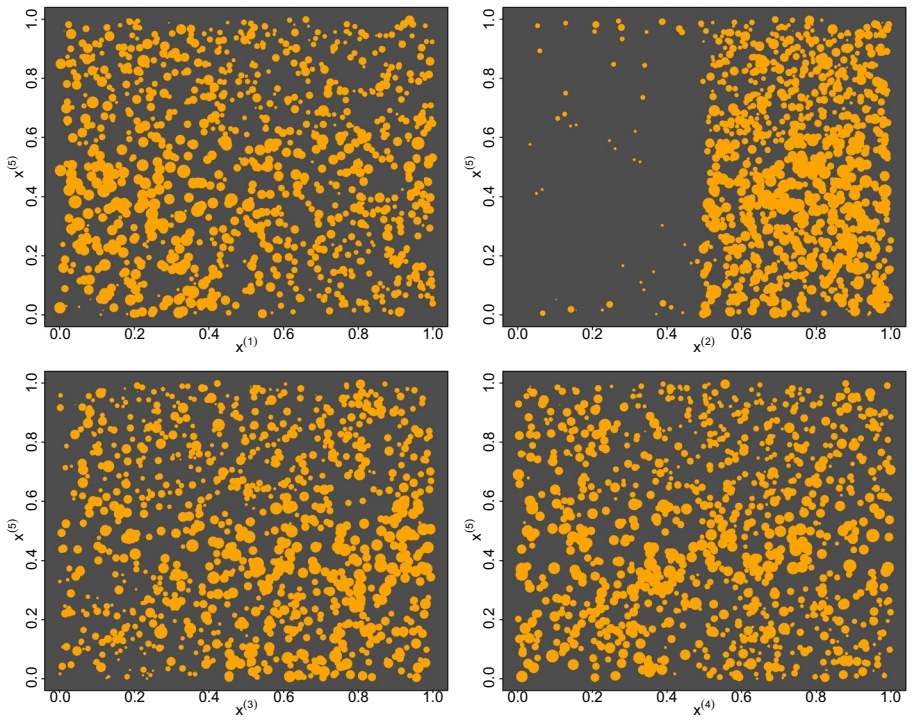

Fig. 10 Individual variable importance values from the synthetic survival example where $x^{(1)}, \ldots, x^{(4)}$ have a strong population-level effect, whereas $x^{(5)}$ has a smaller effect that is modulated by $x^{(2)}$ and influences survival only within a specific subpopulation $x^{(2)} > 0.5$

In these figures, the gradient becomes more significant with increasing mortality (expected number of deaths) from left to right, with Peak METs generally exhibiting the largest gradients. For the highest mortality category (rightmost figures), the gradient for Peak METs (top right) is particularly large for individuals with low recorded Peak METs values, indicating that its effect is especially pronounced for high-risk individuals regardless of ST-segment. On the other hand, large gradient values for the ST-segment are observed for moderate Peak METs, suggesting that this ECG variable may have a moderating effect on Peak METs.

Figure 12 presents a similar set of figures as above, but with the vertical axis representing the Duke Treadmill Score (DTS). Recall that DTS incorporates ECG information, therefore it is interesting to study whether it also exhibits a localized effect. Circle sizes in the top and bottom panels are scaled to the gradients for Peak METs and DTS, respectively. For both variables, gradient values increase with cardiovascular risk. Low Peak METs are associated with a large gradient, regardless of DTS (top right), whereas the DTS gradient is large for moderately low Peak METs. This suggests that patients with very low Peak METs may have the potential to reduce their risk by improving their DTS.

Some confirmation of these findings can be discerned from Fig. 13. The figure displays Kaplan-Meier survival curves for low ST-segment (black) versus high ST-segment (red) across different levels of Peak METs and risk groups, as defined by mortality. Specifically, the analysis stratifies patients into moderate- and high-risk groups and further considers cases where Peak METs is low or moderate. We observe that the ST-segment has a pronounced effect in moderate-risk patients, even when Peak METs is low. In high-risk patients, the effect of the ST-segment is also pronounced, but only when the individual has a moderate Peak METs. This agrees with our finding that in high-risk patients with low Peak METs, the ST-segment does not significantly modulate Peak METs' gradient.

Figure 14 presents the corresponding results for DTS, where low DTS values are shown in black and high values in red. Similar to the previous figure for ST-segment, we observe a pattern in which DTS influences survival outcomes across different Peak METs levels and risk groups. However, the effect of DTS appears less pronounced compared to the ST-segment, suggesting that while DTS remains an important prognostic factor, its relative impact may be weaker in certain subgroups.

5 Discussion

5.1 Challenges with estimating the gradient and contrasting methods

Estimating local gradients in high-dimensional settings poses several challenges. Accurate estimation of ψ is needed to ensure reliable importance measures, while defining appropriate neighborhoods is critical to balance bias and variance. Small neighborhoods risk instability, whereas overly large ones may dilute local structure.

iVarPro addresses these issues by combining local linear regression with rule-based partitions derived from VarPro. The release region for each rule identifies a neighborhood of cases by relaxing the constraint on a target variable, introducing sufficient local variation for estimating coordinate-wise derivatives. These directional gradients are then assembled into an interpretable, individualized estimate of feature importance.

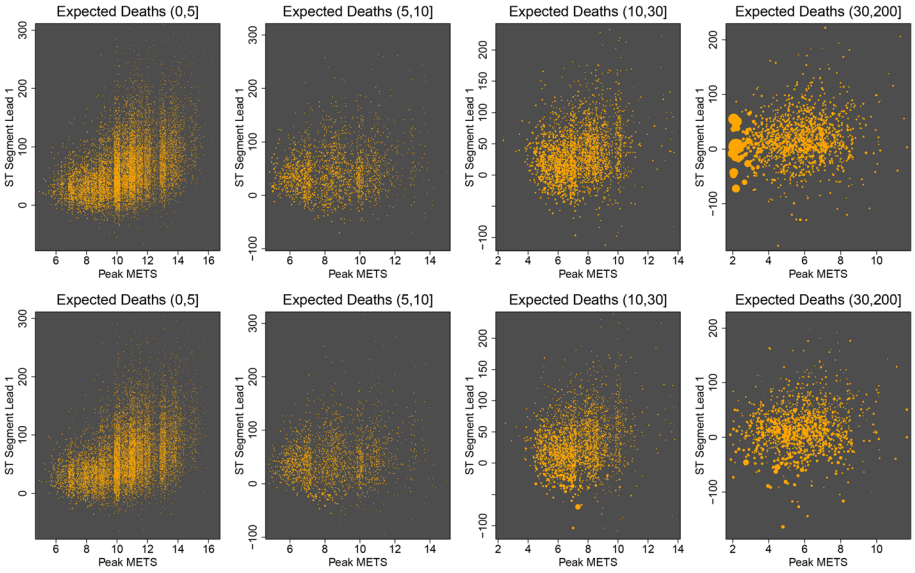

Fig. 11 Clinical variable (Peak METs: Metabolic Equivalents at Peak Exercise) versus ECG variable (ST-segment above the baseline). In the top panel, circle size is scaled to the Peak METs gradient, while in the bottom panel, it is scaled to the ECG gradient, with values jittered for clarity. Data is stratified by patient mortality, measured as the expected number of deaths

A known limitation of tree-based rules is the presence of sharp boundaries, which can introduce discontinuities in estimation. iVarPro mitigates this in two ways: first, by averaging over many randomized tree partitions, which is an approach known to reduce boundary artifacts (Scornet et al. 2015), and second, by using expansion regions around terminal nodes, allowing nearby observations to be included and reducing edge sensitivity.

Several existing model-agnostic methods also aim to assess local feature influence, including LIME (Ribeiro et al. 2016), SHAP (Lundberg and Lee 2017), and ICE/ALE plots (Goldstein et al. 2015; Apley and Zhu 2020). These approaches rely on perturbing or marginalizing over features to approximate local effects. LIME fits surrogate models on perturbed inputs; SHAP computes reweighted model outputs based on conditional expectations; ICE varies one feature at a time to visualize its effect; and ALE improves upon ICE by averaging local differences only within observed regions of the feature space, thereby avoiding extrapolation and offering greater robustness to feature correlations. However, ALE produces population-level summaries of local effects, rather than individualized importance scores. In contrast, iVarPro directly estimates directional gradients at the level of the individual using only observed data, offering a stable, localized, and interpretable measure of feature sensitivity.

5.2 Empirical results and implications for precision medicine

The empirical results from applying iVarPro demonstrate strong and robust performance across a range of conditions. Simulation studies show that iVarPro reliably recovers true gradient structures, even in challenging high-dimensional and correlated settings, outper-

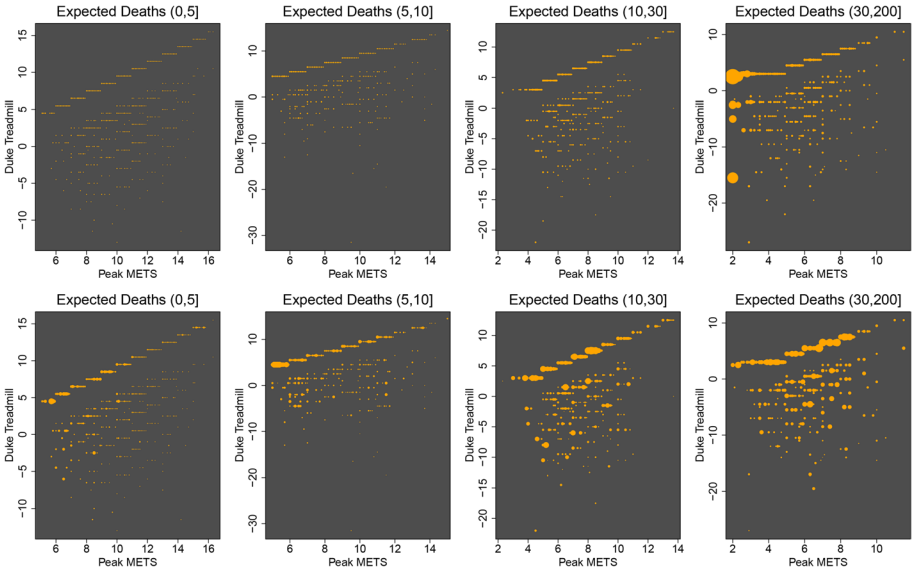

Fig. 12 Same as Fig. 11, but with the vertical axis representing the Duke Treadmill Score (DTS). Circle sizes in the top and bottom panels are scaled to the gradients for Peak METs and DTS, respectively

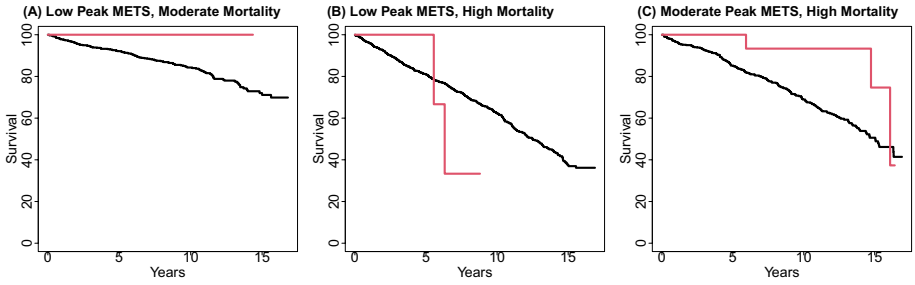

Fig. 13 Kaplan-Meier survival curves for low ST-segment (black) versus high ST-segment (red) under different conditions for Peak METS (low and moderate values) and mortality (moderate risk, high risk)

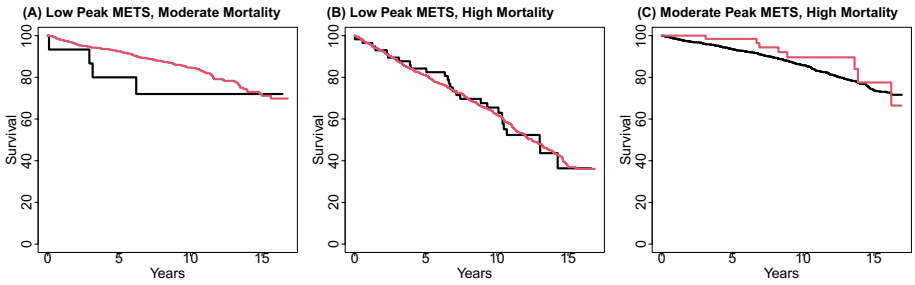

Fig. 14 Kaplan-Meier survival curves for low DTS (black) versus high DTS (red) under different conditions for Peak METS and mortality

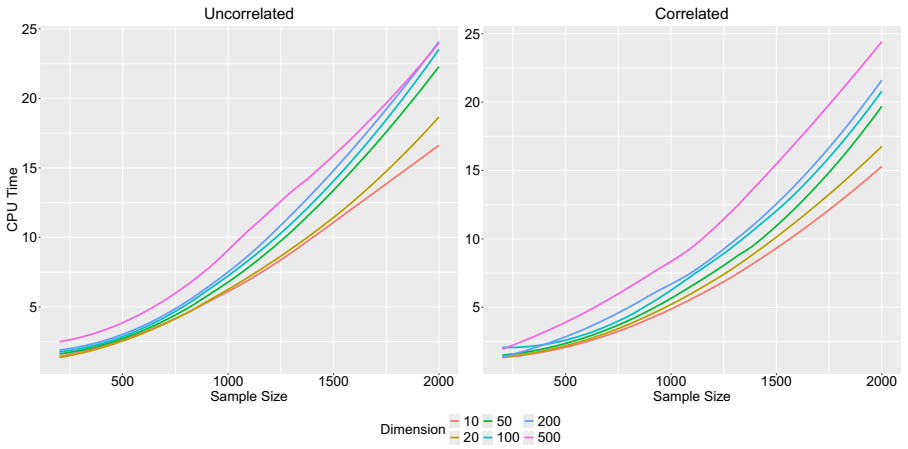

Fig. 15 Computation times (in seconds) for iVarPro across varying sample sizes (n), dimensions (p), and correlation settings. Each line reflects the average CPU time over replicated simulations from Experiment Model 2

forming existing model-agnostic methods. Moreover, comparisons among iVarPro variants highlight the importance of using a well-calibrated external predictor for ψ: procedures based on standard machine learning models (e.g., XGBoost, GBM or random forests) yielded similar and stable performance, indicating robustness to the choice of estimator. In contrast, iVarPro-raw, which uses unsmoothed outcomes, showed substantially degraded accuracy under correlated features. These results emphasize that smoothing the prediction surface is critical for reliable gradient estimation. When applied to real-world clinical and ECG data, iVarPro was able to distinguish between strong population-level predictors, such as Peak METs, and more subtle, patient-specific features like ST-segment deviations and DTS–highlighting its utility for individualized interpretation in precision medicine.

The implications for precision medicine are significant. Traditional variable selection methods prioritize predictors with high average importance, potentially obscuring individual heterogeneity in risk factors. iVarPro provides a personalized framework that enhances our understanding of how specific variables contribute to risk at the individual level. This distinction is especially critical in clinical decision-making, where two patients with similar overall risk scores may have different underlying drivers of that risk. By identifying patient-specific predictors, iVarPro can help tailor interventions to the variables that matter most for each individual.

5.3 Computational burden

To make such personalized modeling feasible in real-world clinical settings, computational efficiency must scale with both sample size and feature dimension. We assessed iVarPro's computational speed using Experiment Model 2, varying sample sizes (n) and feature dimensions (p) under both uncorrelated and correlated settings. Results are shown in Fig. 15. As expected, computation time scaled approximately linearly with n, reflecting the cost of local gradient estimation and tree-based partitioning. Notably, CPU times remained within a few seconds across increasing p, even up to $p = 500$. This stability can be attributed to VarPro's

preliminary feature screening step, which effectively reduces the working dimension before local modeling. By focusing estimation only on signal variables, iVarPro maintains computational feasibility even in high-dimensional settings.

5.4 Future work

Future research could explore extending iVarPro to additional medical domains, incorporating time-dependent effects in survival analysis, and refining its integration with causal inference techniques. Ultimately, this approach offers a promising avenue for enhancing patient-centered decision-making and advancing the goals of precision medicine.

Acknowledgements Research for the authors was supported by the National Institute Of General Medical Sciences of the National Institutes of Health, Award Number R35 GM139659 and the National Heart, Lung, and Blood Institute of the National Institutes of Health, Award Number R01 HL164405.

Author contributions M.L. and H.I. have contributed equally to this work.

Data availability Our code is publicly available as an R-package varPro, which can be accessed at the repository https://github.com/kogalur/varPro.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

Aas K, Jullum M, Løland A (2021) Explaining individual predictions when features are dependent: More accurate approximations to shapley values. Artif Intell 298:103502

Andersen PK, Hansen MG, Klein JP (2004) Regression analysis of restricted mean survival time based on pseudo-observations. Lifetime Data Anal 10(4):335–350

Apley DW, Zhu J (2020) Visualizing the effects of predictor variables in black box supervised learning models. J Royal Stat Soc Series B (Stat Methodol) 82(4):1059–1086

Breiman L (2001) Random forests. Mach Learn 45:5–32

Chen T, Guestrin C (2016) 'Xgboost: A scalable tree boosting system', Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16) pp. 785–794. https://doi.org/10.1145/2939672.2939785

Dai G, Shao L, Chen J (2024) Moving beyond population variable importance: concept, theory and applications of individual variable importance. J R Stat Soc Ser B Stat Methodol 87(3):816–832

Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Monographs Stat Appl Prob 66:360

Goldstein A, Kapelner A, Bleich J, Pitkin E (2015) Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J Comput Graph Stat 24(1):44–65

Gorodeski EZ, Ishwaran H, Blackstone EH, Lauer MS (2009) Quantitative electrocardiographic measures and long-term mortality in exercise test patients with clinically normal resting electrocardiograms. Am Heart J 158(1):61–70

Greenwell B, Boehmke, B, Cunningham, J, Developers G (2020) gbm: Generalized Boosted Regression Models. R package version 2.1.8. https://CRAN.R-project.org/package=gbm

Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA (1982) Evaluating the yield of medical tests. JAMA 247(18):2543–2546

Irwin J (1949) The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. Epidemiol Infection 47(2):188–189

Ishwaran H, Kogalur UB (2025), Random Forests for Survival, Regression, and Classification (RF-SRC). R package version 3.3.3. https://CRAN.R-project.org/package=randomForestSRC

Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS (2008) Random survival forests. Ann Appl Stat 2(3):841–860

Kim DH, Uno H, Wei L-J (2017) Restricted mean survival time as a measure to interpret clinical trial results. JAMA Cardiol 2(11):1179–1180

Komisarczyk K, Kozminski P, Maksymiuk S, Kapsner LA, Spytek M, Krzyzinski, Biecek P (2024) treeshap: Compute SHAP Values for Your Tree-Based Models Using the 'TreeSHAP' Algorithm. R package version 0.3.1. https://cran.r-project.org/package=treeshap

Lu M, Ishwaran H (2024) 'Model-independent variable selection via the rule-based variable priority', arXiv **2409.09003**. arxiv:abs/2409.09003

Lundberg SM Lee S-I (2017) A unified approach to interpreting model predictions, in 'Advances in Neural Information Processing Systems', Vol. 30

Mark DB, Shaw L, Harrell FE Jr, Hlatky MA, Lee KL, Bengtson JR, McCants CB, Califf RM, Pryor DB (1991) Prognostic value of a treadmill exercise score in outpatients with suspected coronary artery disease. N Engl J Med 325(12):849–853

Pedersen TL, Benesty M (2022) lime: Local Interpretable Model-Agnostic Explanations. R package version 0.5.3. https://cran.r-project.org/package=lime

Ribeiro MT, Singh S, Guestrin C (2016) Why should I trust you? explaining the predictions of any classifier, In: 'Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', ACM, pp. 1135–1144

Royston P, Parmar MK (2011) The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. Stat Med 30(19):2409–2421

Ruppert D, Wand M (1994) Multivariate locally weighted least squares regression. Ann Statist 22(1):1346–1370

Saito T, Rehmsmeier M (2015) The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE 10(3):e0118432

Scornet E, Biau G, Vert J-P (2015) Consistency of random forests. Ann Stat 43(4):1716–1741

Strobl C, Boulesteix A-L, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 8(1):1–21

Vivekananthan DP, Blackstone EH, Pothier CE, Lauer MS (2003) Heart rate recovery after exercise is a predictor of mortality, independent of the angiographic severity of coronary disease. J Am Coll Cardiol 42(5):831–838

Winn-Nuñez ET, Griffin M, Crawford L (2024) A simple approach for local and global variable importance in nonlinear regression models. Comput Stat Data Anal 194:107914

Wright MN, Ziegler A (2017) Ranger: a fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw 77(1):1–17

Xu R, Nettleton D, Nordman DJ (2016) Case-specific random forests. J Comput Graph Stat 25(1):49–65