# Model Descriptions

# Model Performance

Loss = the average loss over the testing dataset (the sum of all the losses divided by the number of items in the testing dataset).

**Note**: Binary cross-entropy is specified in training

Base Model test loss: 0.04847599193453789  
Base Model test accuracy: 0.988304078578949

Small Model test loss: 0.05162016674876213  
Small Model test accuracy: 0.988304078578949

BigNet5 Model test loss: 0.11611512303352356  
BigNet5 test accuracy: 0.9649122953414917

BigNet7 Model test loss: 0.14744716882705688  
BigNet7 test accuracy: 0.9707602262496948

BigNet10 Model test loss: 0.25518128275871277  
BigNet10 test accuracy: 0.9590643048286438

BigNet15 Model test loss: 0.14141316711902618  
BigNet15 test accuracy: 0.9824561476707458

**Final Remarks:** Top 3 best models are:

1. Base Model ~ minimizes loss best
2. Small Model ~ same accuracy as base
3. BigNet15 Model ~ 98.2% accuracy

# DALEX Interpretations

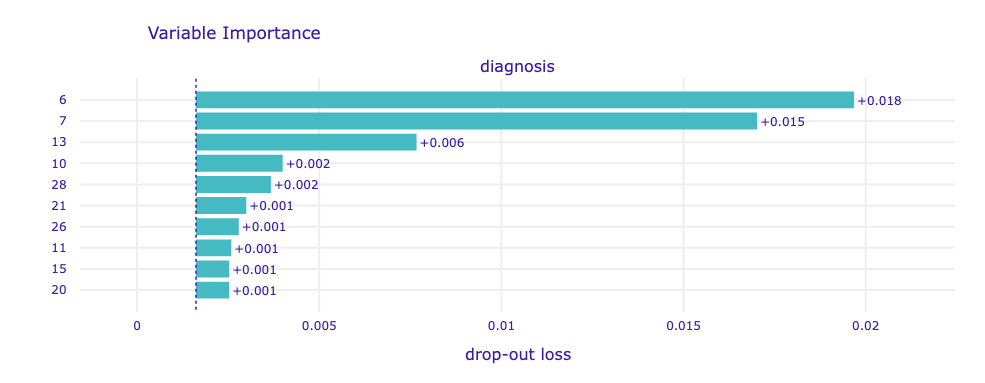
**Important Note**: DALEX interpretations evaluate trained model when fitted to test data

## Model-level Explanations

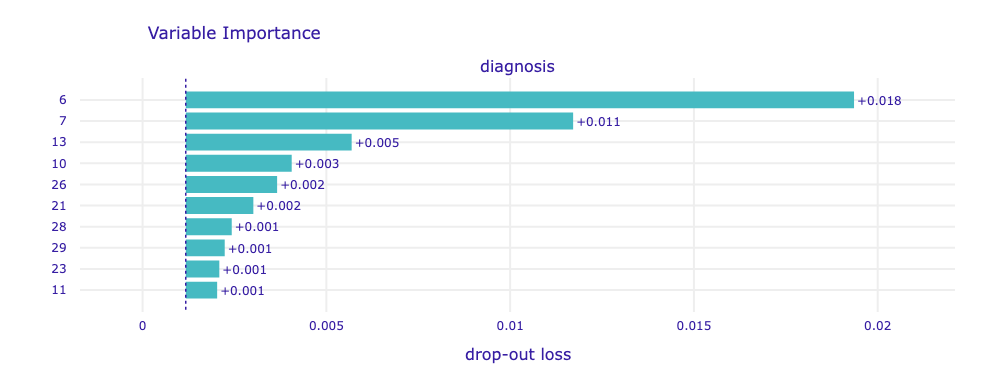
**Permutation-based Variable Importance Plots:**

*Params*: loss function = ‘auc’ (for binary classification)

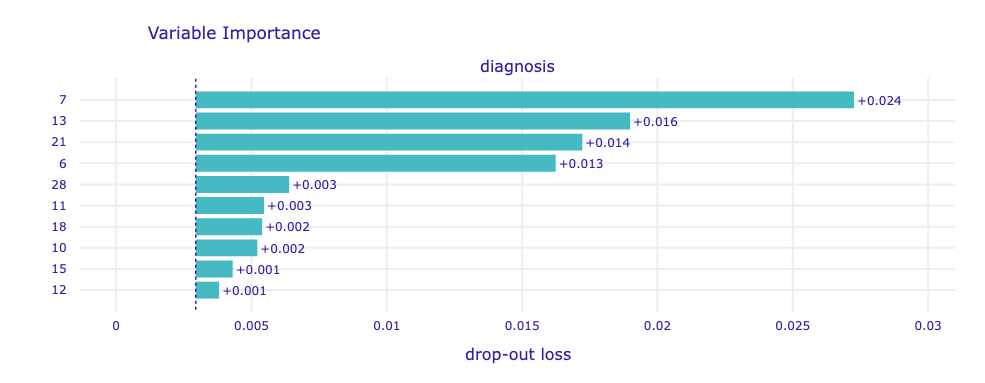
Base Model



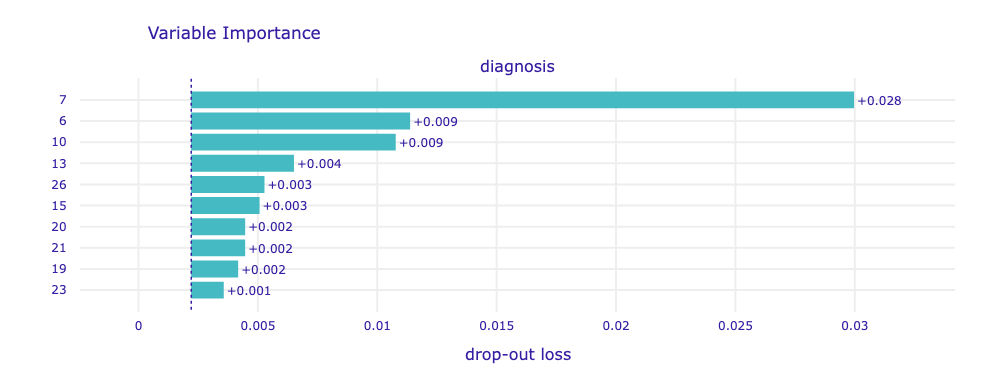
Small Model



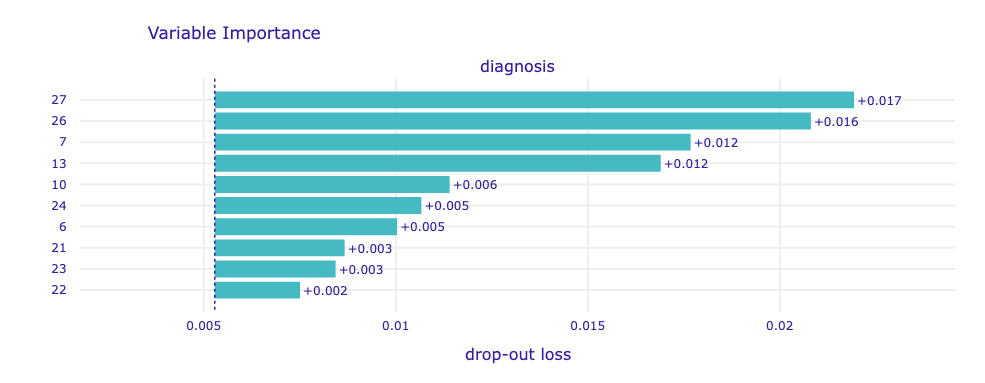
BigNet5 Model



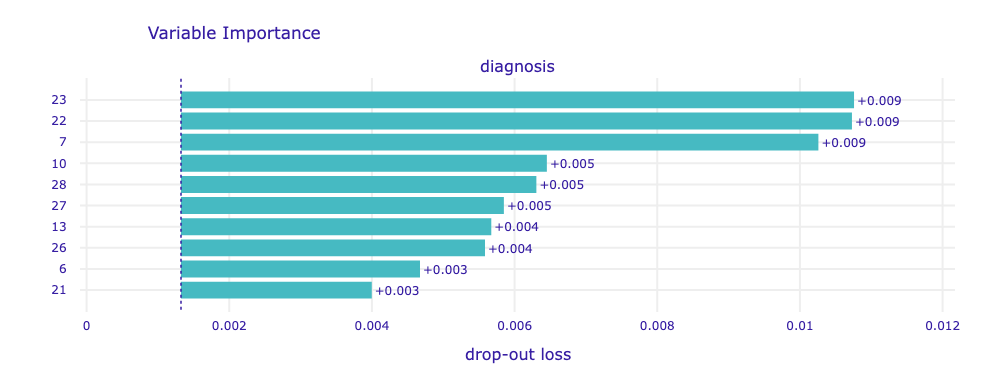
BigNet7 Model



BigNet10 Model



BigNet15 Model



**Final Remarks:**

* All models identify **concave.points\_mean** (i.e., 7), **concavity\_mean** (i.e., 6), **area\_se** (i.e., 13), **radius\_se** (i.e., 10), and **texture\_worst** (i.e., 21) as highly important features
* **concave.points\_mean** is identified in the top 3 most important features across all models
* The bars start in drop-out loss value for the model on the original data (x-axis). The length of the bar corresponds to the drop-out loss *after* permutations

**Diagnostic Residual Plots:**

*Params*: loss function = ‘auc’ (for binary classification)

Note: The diagnostic function used to generate the below plots performs model diagnostics of residuals. Residuals are calculated and plotted against predictions, true y values or selected variables.

## Prediction-level Explanations

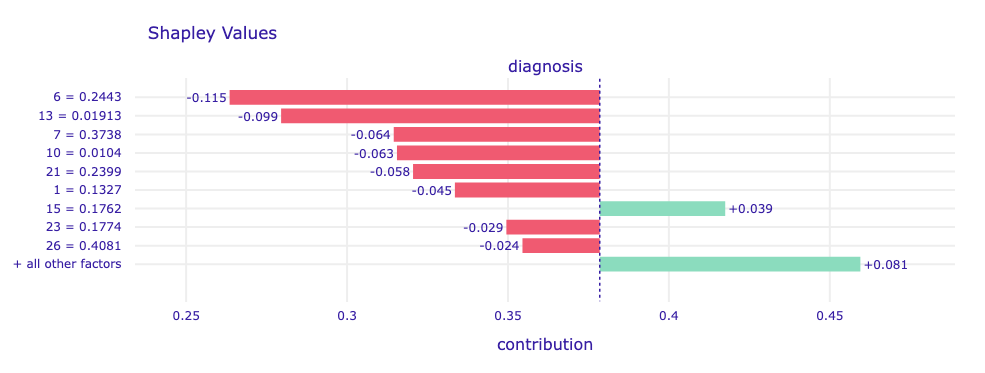
**SHAP Plots:**

*Params*: type = ‘shap’

**Important note**: **Negative** values are indicative of negative cancer diagnoses; **positive** values are indicative of positive cancer diagnoses

**Important note**: On the x-axis we have prediction value; on the y-axis, we have variables and their value for the observation

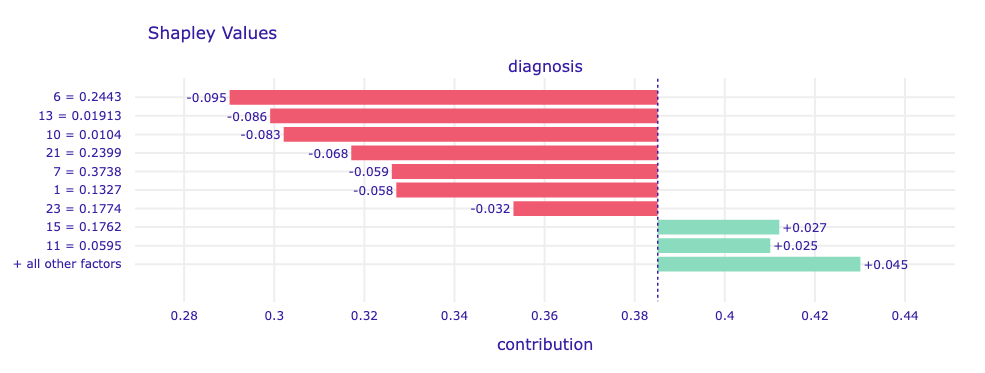
Base Model



Notes:

* The biggest influence negative and overall on diagnosis in this model is **concavity\_mean**
* The biggest positive influence on diagnosis in this model is **compactness\_se**
* **concavity\_mean, area\_se, concave.points\_mean, radius\_se, texture\_worst, texture\_mean area\_worst** and **concavity\_worst** contribute to negative diagnoses
* **compactness\_se** and all other features contribute to positive diagnoses

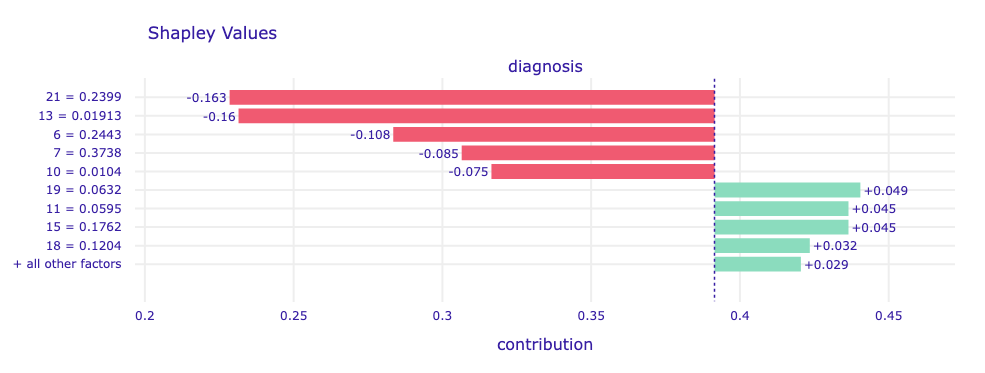
Small Model



Notes:

* The biggest influence negative and overall on diagnosis in this model is **concavity\_mean**
* The biggest positive influence on diagnosis in this model is **compactness\_se**
* **concavity\_mean, area\_se, radius\_se, texture\_worst, concave.points\_mean, texture\_mean,** and **area\_worst** contribute to negative diagnoses
* **compactness\_se, texture\_se** and all other features contribute to positive diagnoses

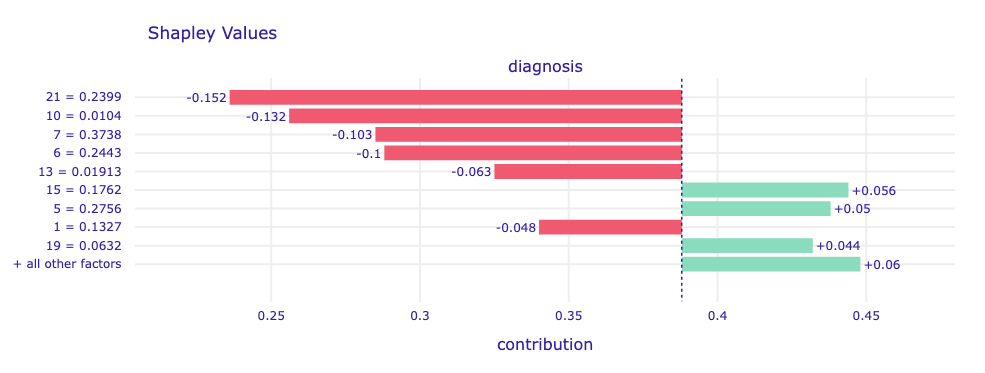
BigNet5 Model



Notes:

* The biggest influence negative and overall on diagnosis in this model is **texture\_worst**
* The biggest positive influence on diagnosis in this model is **fractal\_dimension\_se**
* **texture\_worst, area\_se, concavity\_mean, concave.points\_mean,** and **radius\_se** contribute to negative diagnoses
* **fractal\_dimension\_se, texture\_se, compactness\_se, symmetry\_se** and all other features contribute to positive diagnoses

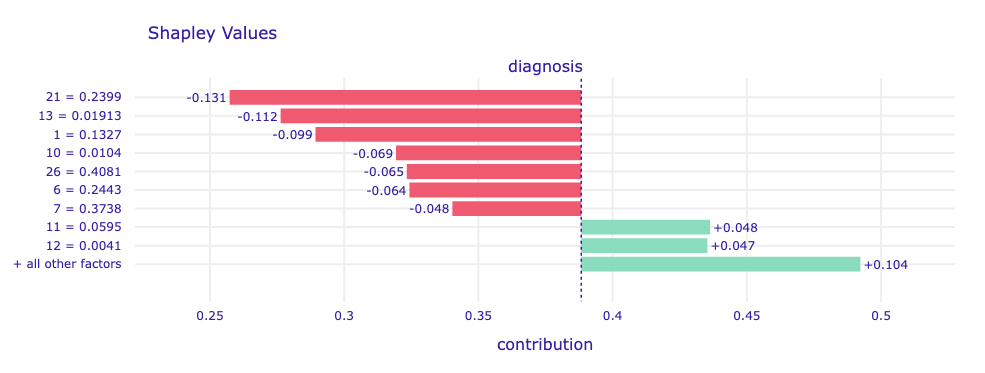
BigNet7 Model



Notes:

* The biggest influence negative and overall on diagnosis in this model is **texture\_worst**
* The biggest positive influence on diagnosis in this model is **compactness\_se**
* **texture\_worst, radius\_se, concave.points\_mean, concavity\_mean, area\_se,** and **texture\_mean** contribute to negative diagnoses
* **compactness\_se, compactness\_mean, fractal\_dimension\_se** and all other features contribute to positive diagnoses

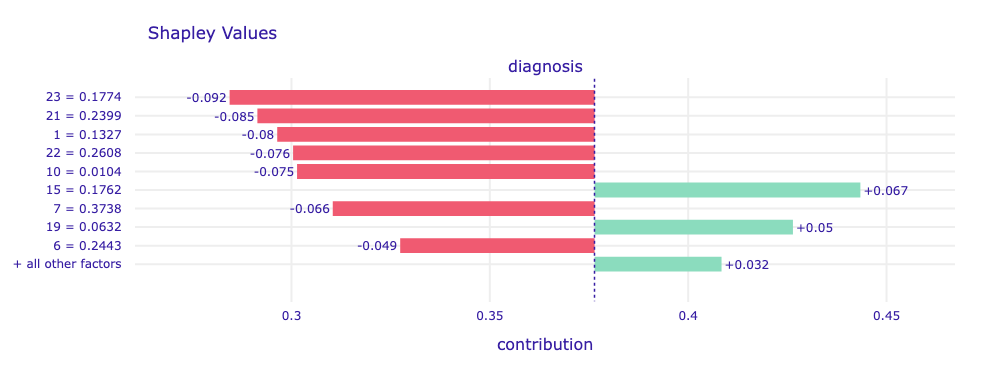
BigNet10 Model



Notes:

* The biggest influence negative and overall on diagnosis in this model is **texture\_worst**
* The biggest positive influence on diagnosis in this model is **texture\_se**
* **texture\_worst, area\_se, texture\_ mean, radius\_se, concavity\_worst, concavity\_mean,** and **concave.points\_mean** contribute to negative diagnoses
* **texture\_se, perimeter\_se,** and all other features contribute to positive diagnoses

BigNet15 Model



Notes:

* The biggest influence negative and overall on diagnosis in this model is **area\_worst**
* The biggest positive influence on diagnosis in this model is **compactness\_se**
* **area\_worst, texture\_worst, texture\_mean, perimeter\_worst, radius\_se, compactness\_se, concave.points\_mean,** and **concavity\_mean** contribute to negative diagnoses
* **compactness\_se, fractal\_dimension\_se** and all other features contribute to positive diagnoses

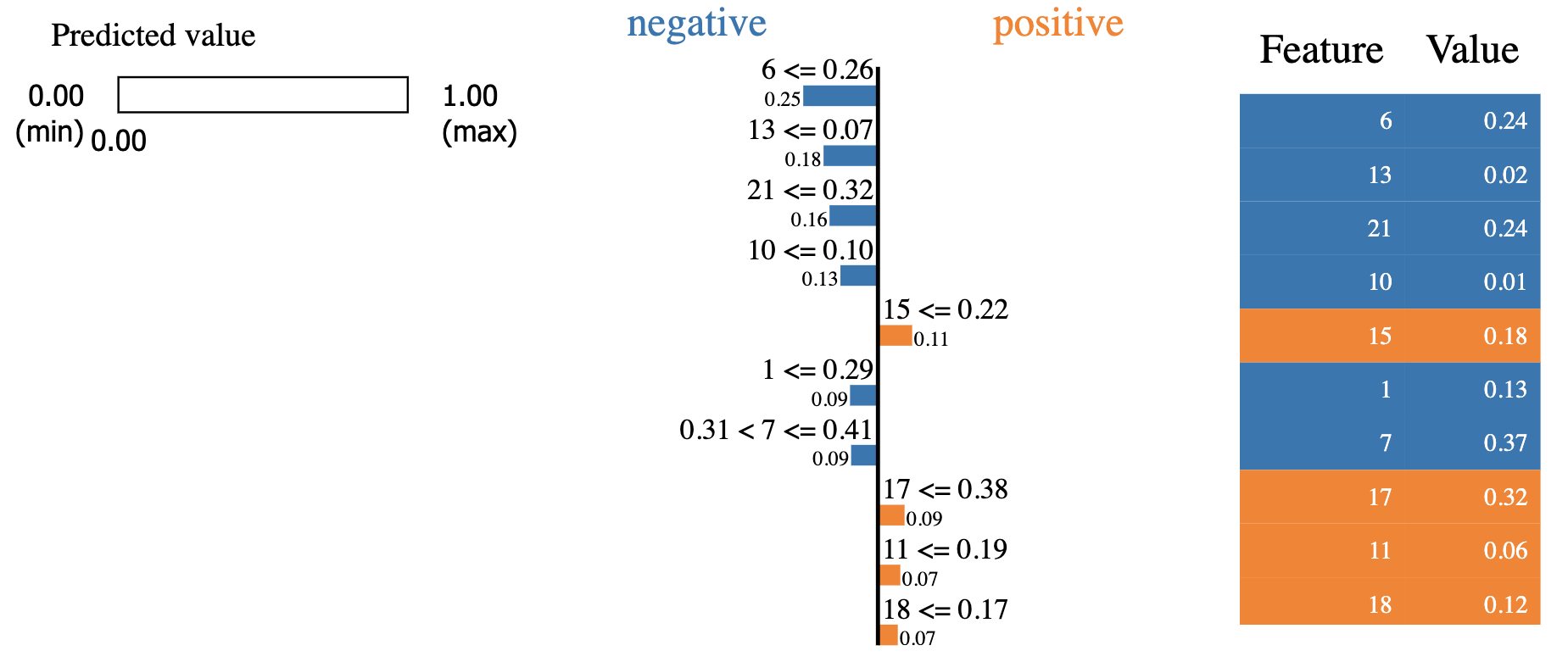
**Important Note**: Caution should be exercised when deriving insights from SHAP analyses. It is important not to become over-invested in conclusions that certain features *cause* certain outcomes unless the experiment has been conducted within a causal framework (this is rarely the case). SHAP only tells you what the model is doing within the context of the data on which it has been trained: it doesn't necessarily reveal the true relationship between variables and outcomes in the real world. Decision-makers are often tempted to view features in SHAP analyses as dials that can be manipulated to engineer specific outcomes, so this distinction must be communicated.

**Lime Plots:**

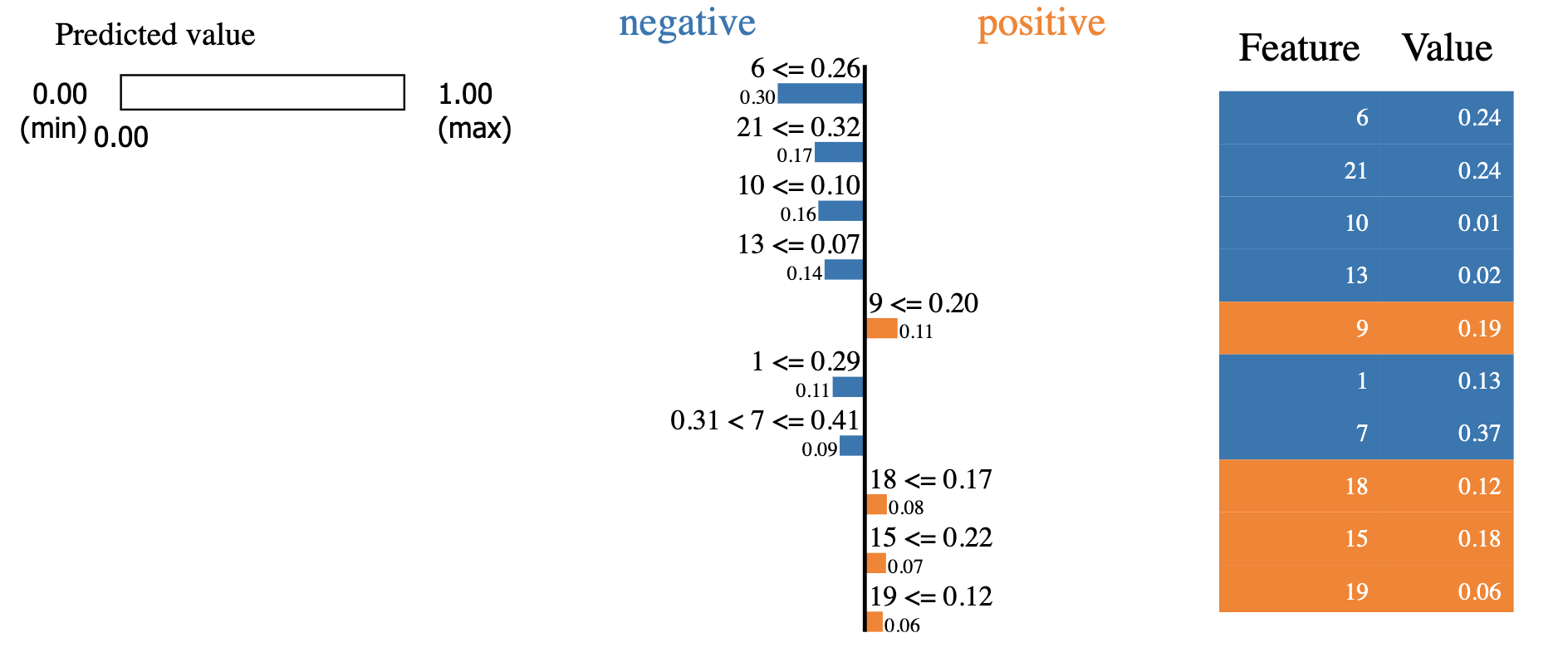
*Params*:

**Important note**:

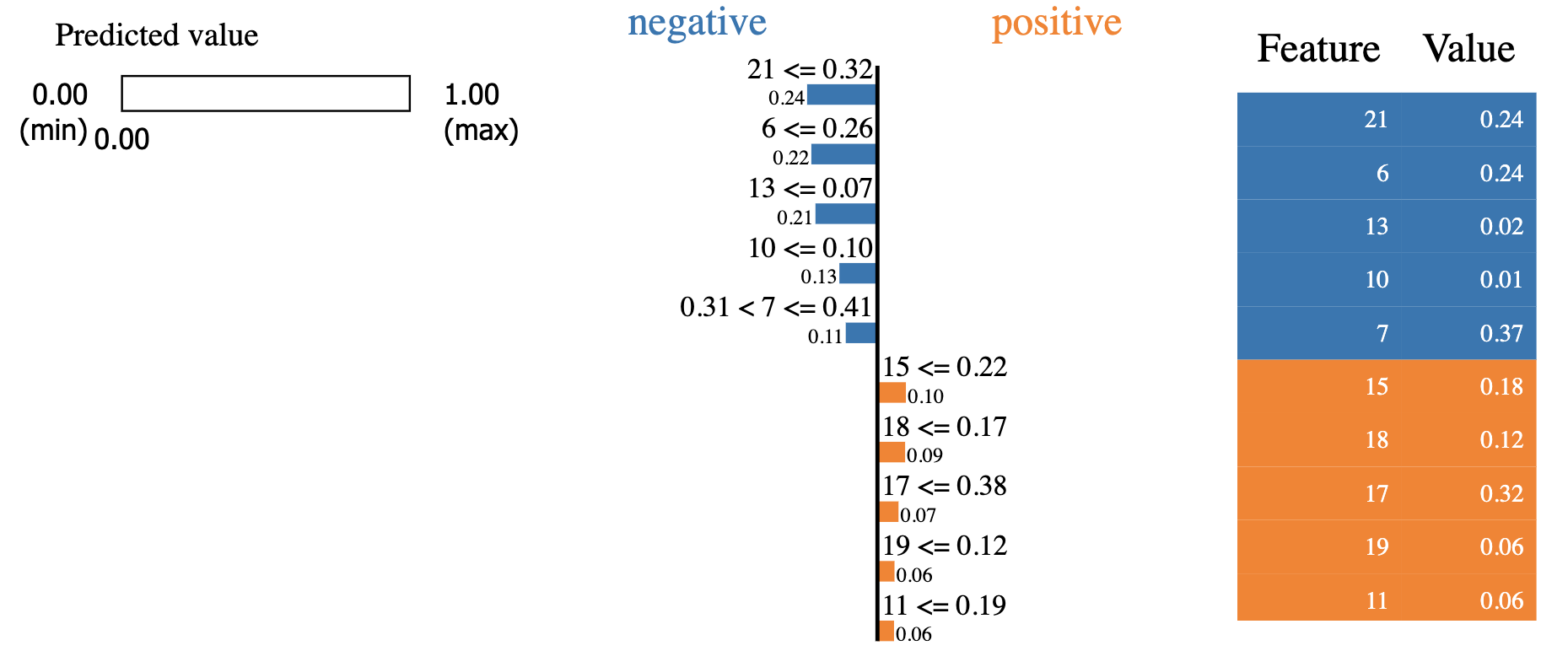
Base Model



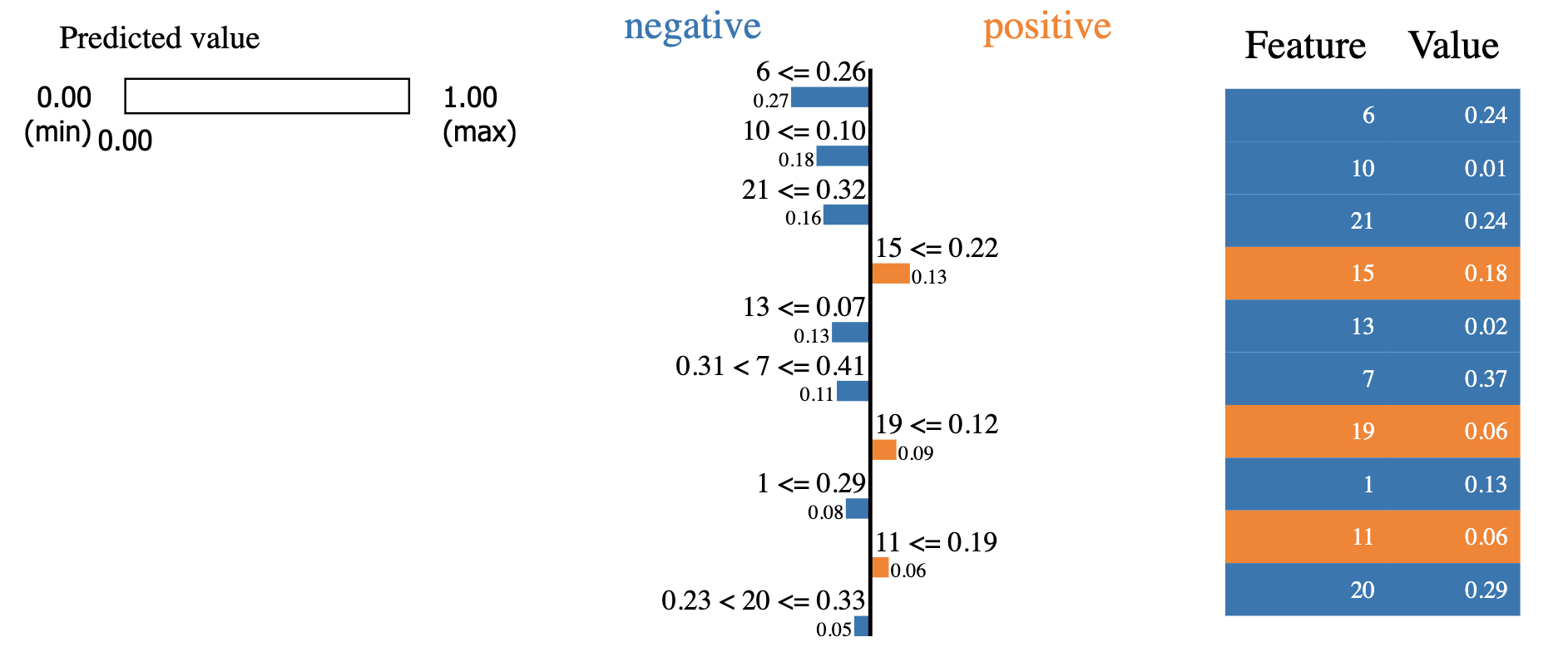
Small Model



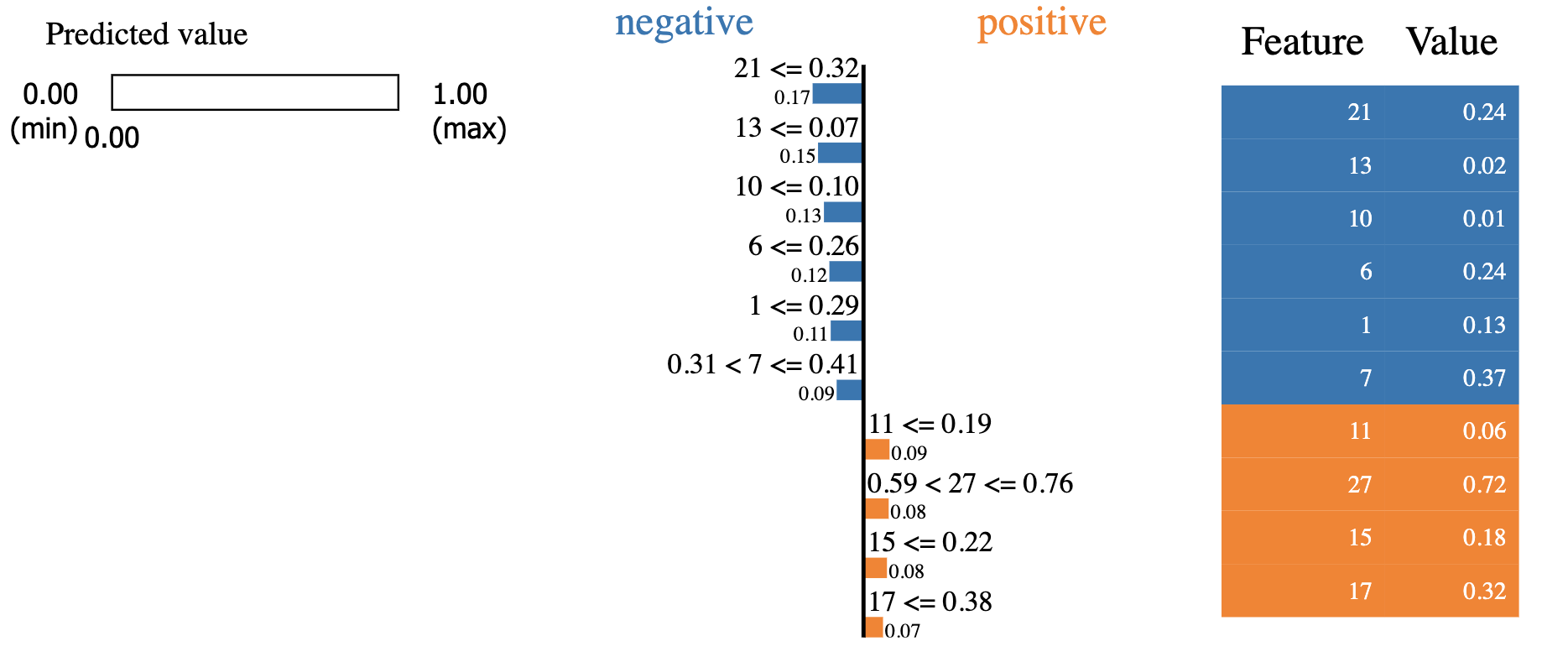
BigNet5 Model



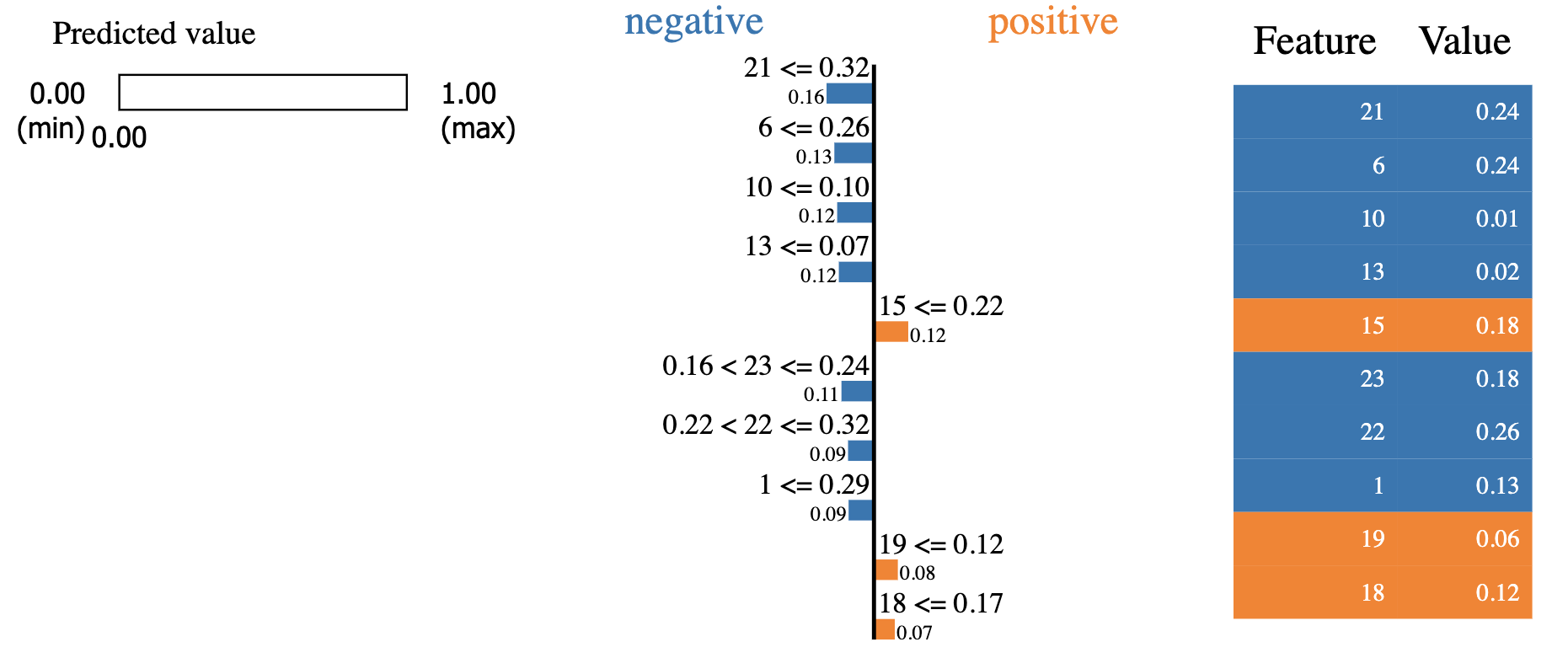
BigNet7 Model



BigNet10 Model



BigNet15 Model



# Alibi Interpretations

**Integrated Gradients Plots:**

*Params*:

**Important note**:

Base Model

Small Model

BigNet5 Model

BigNet7 Model

BigNet10 Model

BigNet15 Model

**ALE Plots:**

*Params*:

**Important note**:

Base Model

Small Model

BigNet5 Model

BigNet7 Model

BigNet10 Model

BigNet15 Model