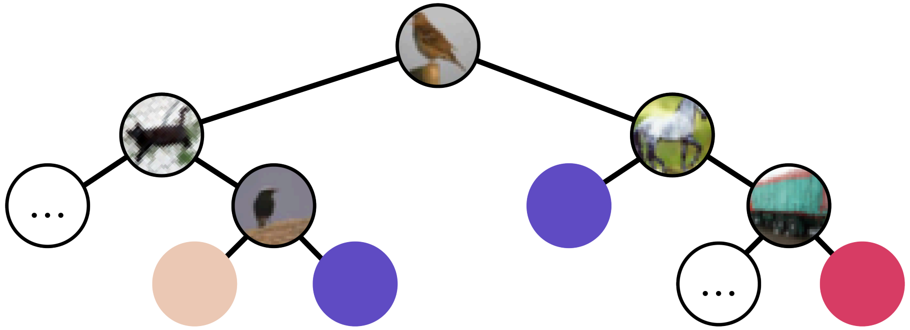
# Data-Agnostic Pivotal Instances Selection for Decision-Making Models

*Case-based Trees for relational and non-relational data*




Alessio Cascione *University of Pisa*
*Mattia Setzu* *University of Pisa*
Riccardo Guidotti *University of Pisa*

EXPLANATION OF AI DECISION MAKING.
MI MO SA
FAIR Future Artificial Intelligence Research
PIANO
Funded by the European Union NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

# Explainability and its many forms

- Feature importance: **impacting** the task
- Decision rules: **describing** the task
- Counterfactuals: **defining** a behavior


Age >= 50,
Smokes: Yes,
Respiratory history: Yes
**=> Ill**

Age >= 50,
*Smokes: Yes to Smokes: No*
=> Ill to Not ill

# Explainability and its many questions

- Feature importance: What is important?
- Decision rules: What decides?
- Counterfactuals: What to change?


Age >= 50,
Smokes: Yes,
Respiratory
history: Yes
**=> Ill**

Age >= 50,
*Smokes: Yes to*
*Smokes: No*
=> Ill to Not ill

# Explainability and its many scopes

- Feature importance: Local
- Decision rules: Local/Semi-global
- Counterfactuals: Local


Age >= 50,
Smokes: Yes,
Respiratory history: Yes
**=> Ill**

Age >= 50,
*Smokes: Yes to Smokes: No*
=> Ill to Not ill

# Non-relational data: breaking the mold

Non-relational data has an inherent difficulty in tackling the above tasks, often due to

- Complex relations among features
- Violated assumptions
- Sparsity

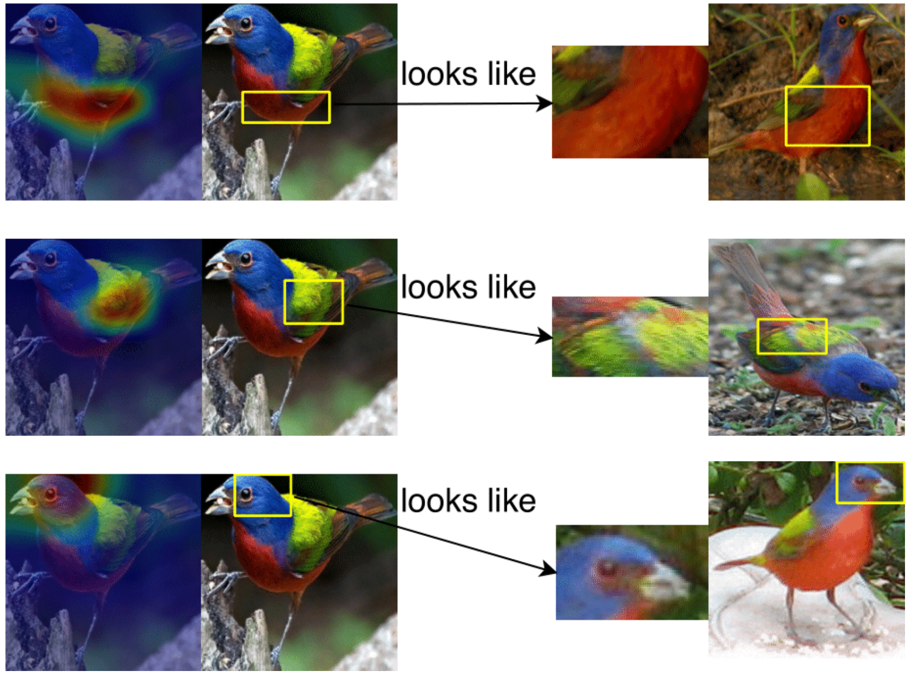
A saliency map requires a perturbation algorithm how to *jointly* perturb pixels, often under the assumption of their independence, thus creating a large amount of possibly out-of-distribution synthetic instances.

# Case-based explanations

A.k.a. prototypes, pivots. Define a set of instances, which define...

- Importances
- Counterfactuals

By design they tackle sparsity and inter-instance relationships!




The bird in the picture is classified as *Clay-colored sparrow* on the basis of its similarity with some prototypes.

# The case-based landscape

- By design
- Data-agnostic
- Factual explanations
- Local *or* global explanations

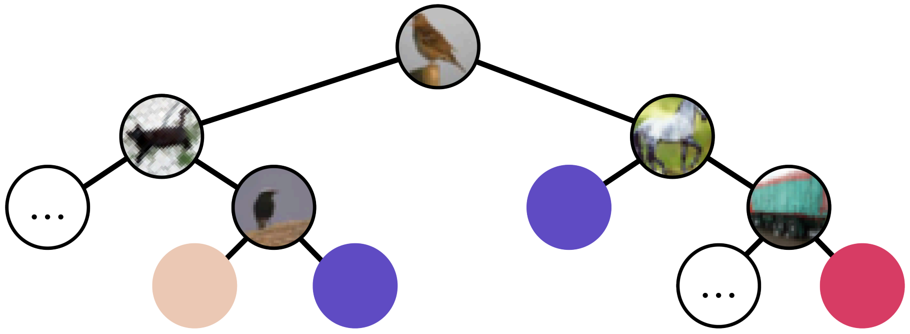
- No counterfactuals
- Local *and* global explanation
- No inter-case relationship

Prototype selection for interpretable classification. J. Bien et al.
Deep learning for interpretable image recognition. C. Chen et al.
Interpreting CNNs via Decision Trees. Q. Zhang et al.
Proximity Forest: an effective and scalable distance-based classifier for time series. Lucas et al.
Prototype selection for interpretable classification. Bien & Tibshirani.

# Pivot Tree

Our proposal, **Pivot Tree** is a case-based model for explainable classification

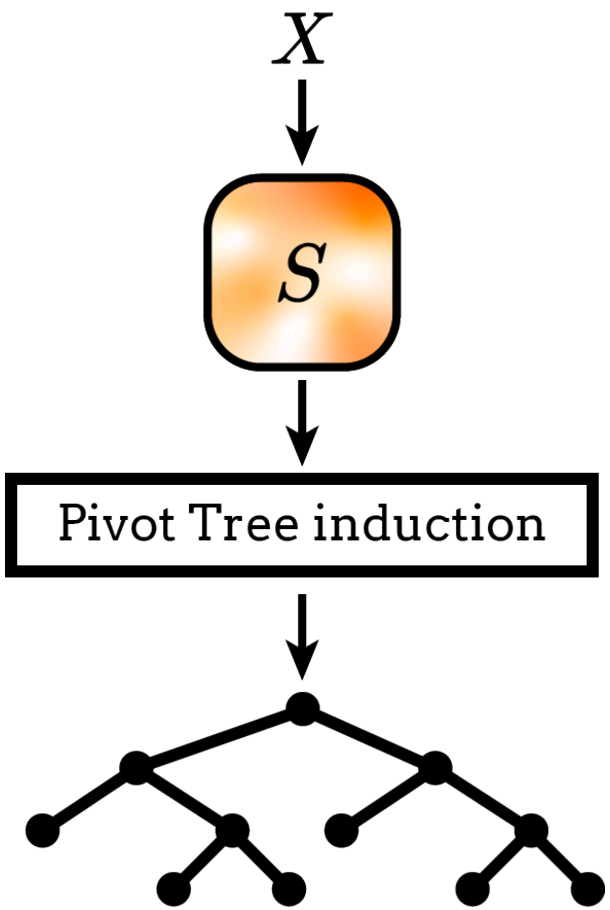
- Explainable by design
- Data-agnostic
- Factual explanations
- Counterfactuals
- Local and global explanations
- Hierarchical case structure



A Pivot Tree: instances laid on a tree structure route predictions. Similarity to instances determines routing, and leaves are also associated with a classification label.

# Learning a Pivot Tree (PTC)

**Pivot Tree Classifier (PTC)**



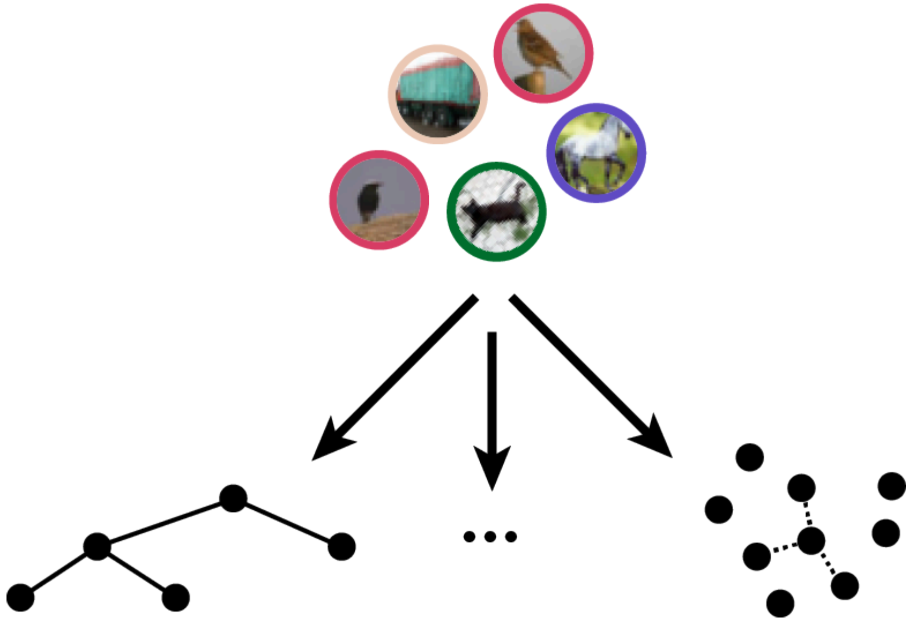

1. Compute a similarity matrix $S$, possibly embedding data
2. Induce a Decision Tree on $S$, filtering similarities to
   - **discriminative** pivots: maximizing entropy
   - **descriptive pivots**: maximizing class centrality

# Learning a Pivot Selector (PTS)

Pivots in the nodes constitue a dataset themselves!

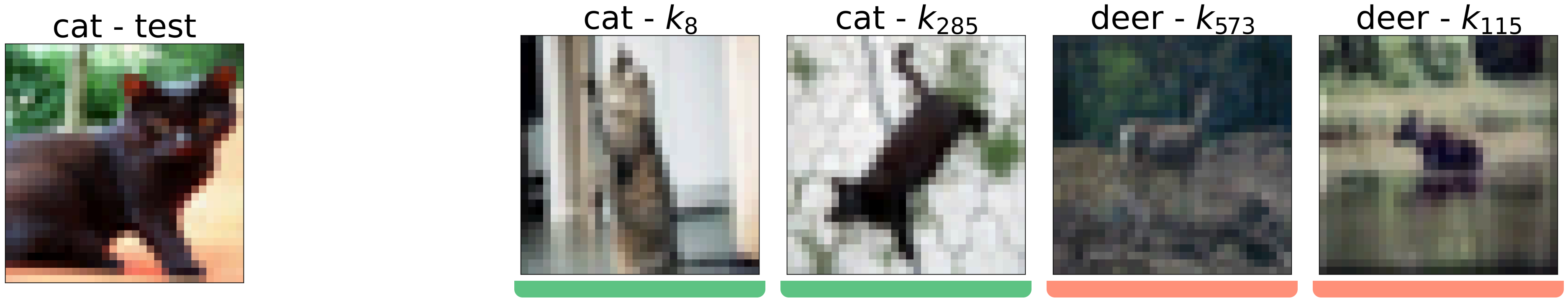
1. Extract pivots from nodes
2. Train an interpretable model **on top** of Pivot Tree

**Pivot Tree Selector (PTS)**



# A qualitative result

## Cifar 10




An instance (left), and pivots extracted by Pivot Tree. The cat has high similarity to the first two, and low to the latter two.

# Experiments

**Explainability**

- Fidelity
- Complexity
- Stability

**Competitors**

- $\varepsilon$-ball
- $k$-NN

**Datasets**

- 11 tabular
- 5 time series
- 3 image + 1 real world use case of oral cancer detection
- 5 text

# How faithful is Pivot Tree?

| | $mean \uparrow$ | $std \downarrow$ | $rank \downarrow$ |
|---|---|---|---|
| **PTS + DT** | .62 | .22 | 3.1 |
| K-medoids | .61 | .22 | 3.6 |
| $\varepsilon$-ball | .61 | .23 | 3.7 |
| **Pivot Tree** | .58 | .27 | 3.9 |
| K-means | .62 | .22 | 4.0 |
| Random per class | .60 | .22 | 4.6 |
| Random | .60 | .22 | 5.0 |

On $f1$ score...

- Pivot Tree works **better** as a pivot **selector than** a standalone **classifier**
- Models built directly on data lag behind: good data matters

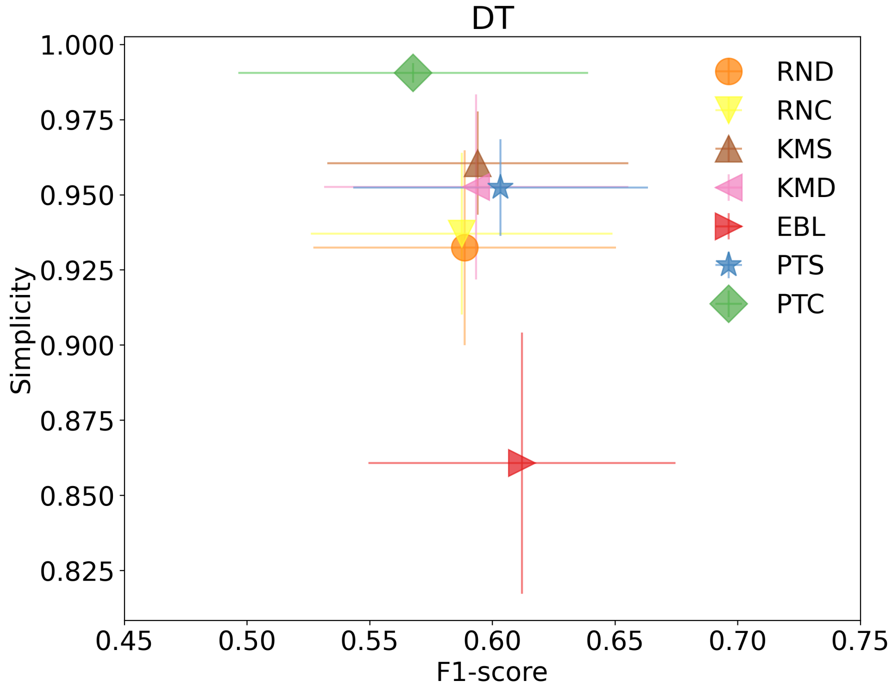
# A real-world study: oral cancer detection

A real-world case study on an oral cancer detection dataset.

- 535 images
- 3 classes
- Publicly available

| | **Pivot Tree** | **Decision Tree** | $k$**-NN** | **CNN** |
|---|---|---|---|---|
| $f1$ | **.834** | .833 | .811 | .854 |
| $complexity$ | 9 | 47 | **5** | |

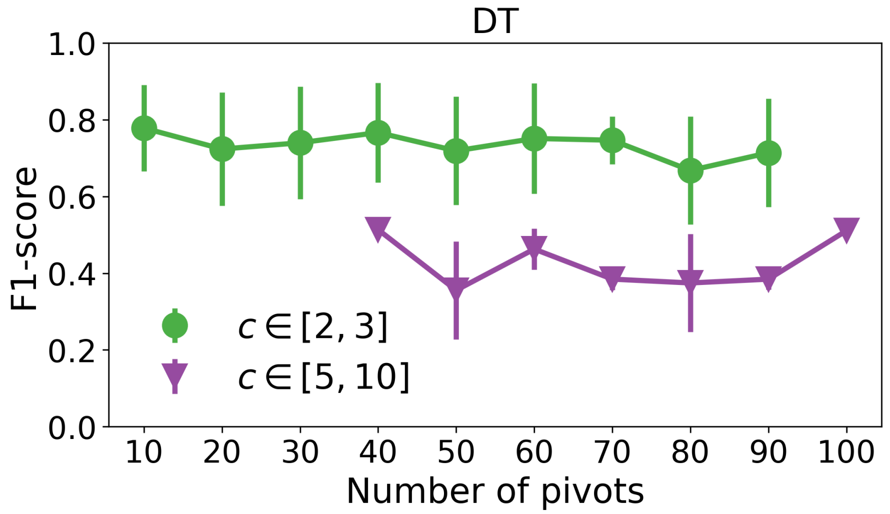
# How complex is Pivot Tree?




Scatter plot of $f1$ score and simplicity of Pivot Tree and its competitors, regularized to use a maximum of $20$ pivots.

- When not regularized, Pivot Tree slightly worse than competitors
- More complex models (bottom) have higher variance

# How stable is Pivot Tree?




Scatter plot of $f1$ score of Pivot Tree as the maximum number of pivots increases. $C$ indicates the number of classes.

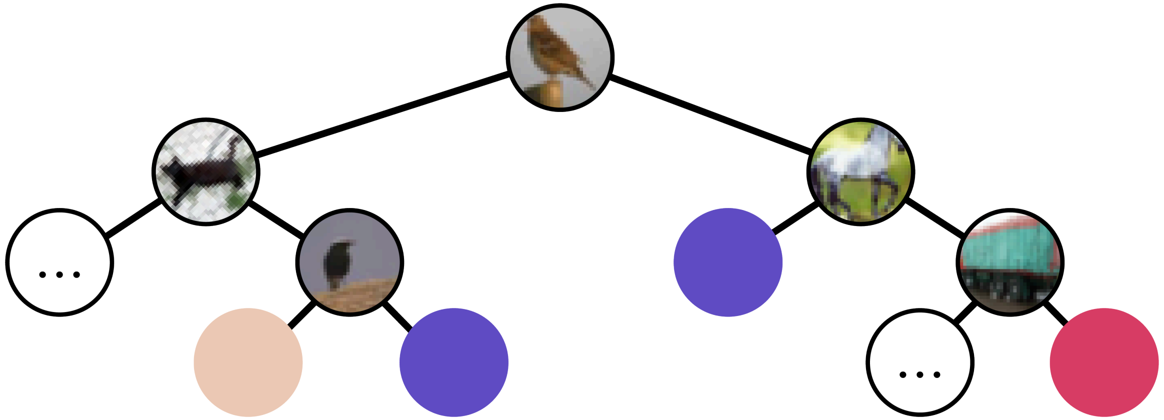
Low variance but decreasing trend as regularization decreases: Occam's razor!

# Pivot Tree

An explainable by-design case-based algorithm.

Future work:

- Proximity splits: to which pivots are instances more similar?
- More powerful downstream models




github.com/msetzu/pivottree

**See you at the poster session!**