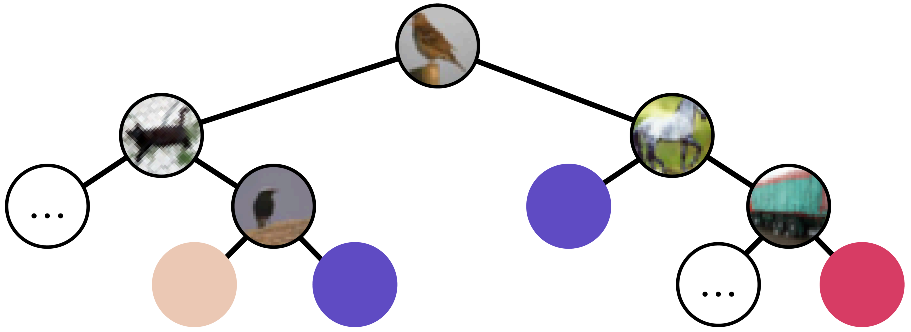
Data-Agnostic Pivotal Instances Selection for Decision-Making Models

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.
MIMOSA
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: Define a set of features **impacting** the task
- Decision rules: Define a set of constraint **defining** a behavior
- Counterfactuals: Define a set of constraints **defining** a behavior

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

Age >= 50,
<u>Smokes: Yes to Smokes: No</u>
=> Ill to Not ill

Explanations types: rules, counterfactual rules, and feature importance.

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

Explanations types: rules, counterfactual rules, and feature importance.

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

Explanations types: rules, counterfactual rules, and feature importance.

An XAI roadblock: non-relational data

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

- Complex relations among features
- Violated assumptions
- Sparsity

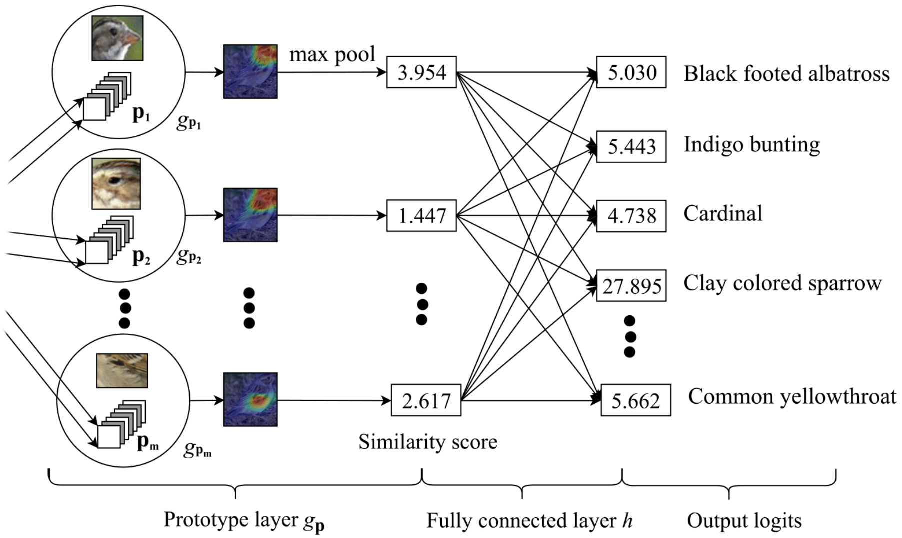
A saliency map requires a perturbation algorithm how to *jointly* perturb pixels, often under the assumption of their independence, thus creating a large amount of possibly out-of-distribution synthetic instances.

Case-based explanations

Define a set of instances, which can...

- Indicate which instances are important
- Define neighborhoods
- Provide counterfactuals

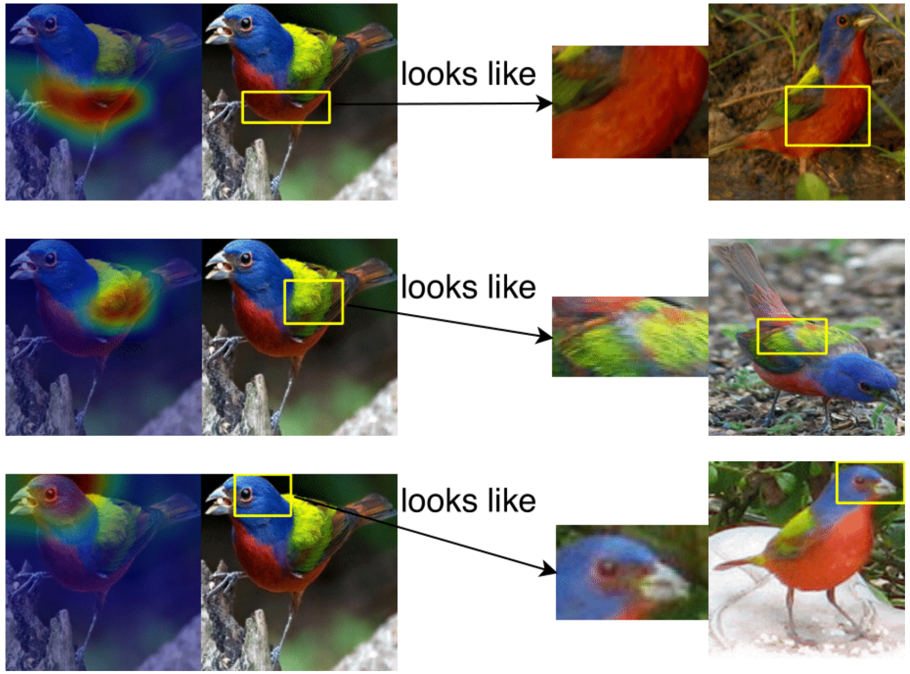
By design they tackle sparsity, inter-instance relationships, and **push interpretation to the user!**



A case-based model: the bird in the picture is classified as *Clay-colored sparrow* on the basis of its relationship with other prototypes: the model predicts by leveraging instances.

Case-based explanations

- Often universal, and applicable to any data type
- Inherently understandable, since we often rely on case-based reasoning ourselves
- Applicable at different scopes, both local and global



Cases lend themselves to explanation too: here, feature importance of the retrieved cases.

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.

The case-based landscape

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

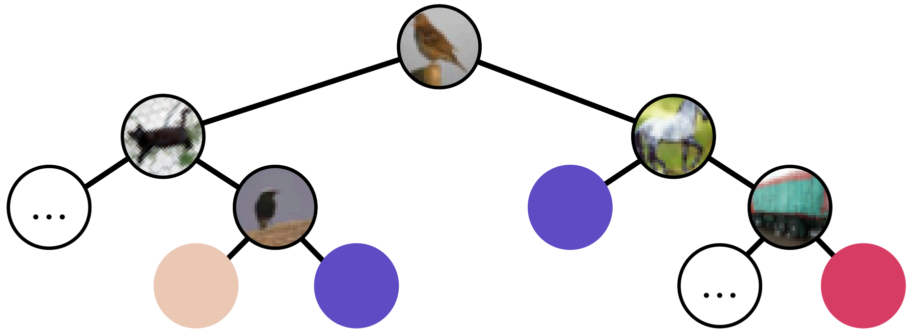
- No counterfactuals
- Local *and* global explanation
- No inter-case relationship

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

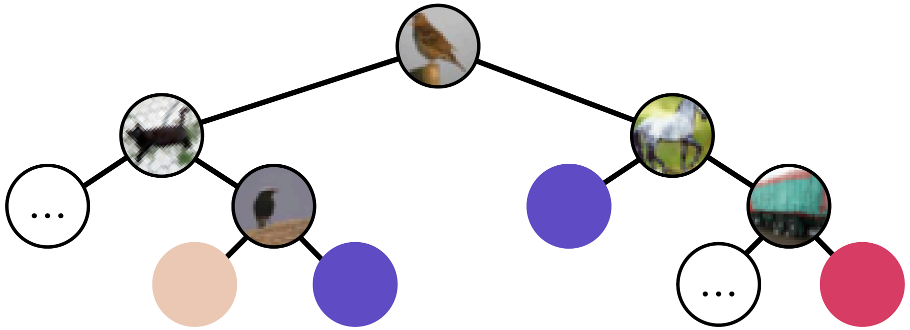


A Pivot Tree: instances laid on a tree structure route predictions. Similarity to instances determines routing, and leaves are also associated with a classification label.

Pivot Tree

Learning a Pivot Tree

- Select a set of pivots P with $\pi(X)$
- Learn a tree $f(\pi(X))$ on $P = \pi(X)$

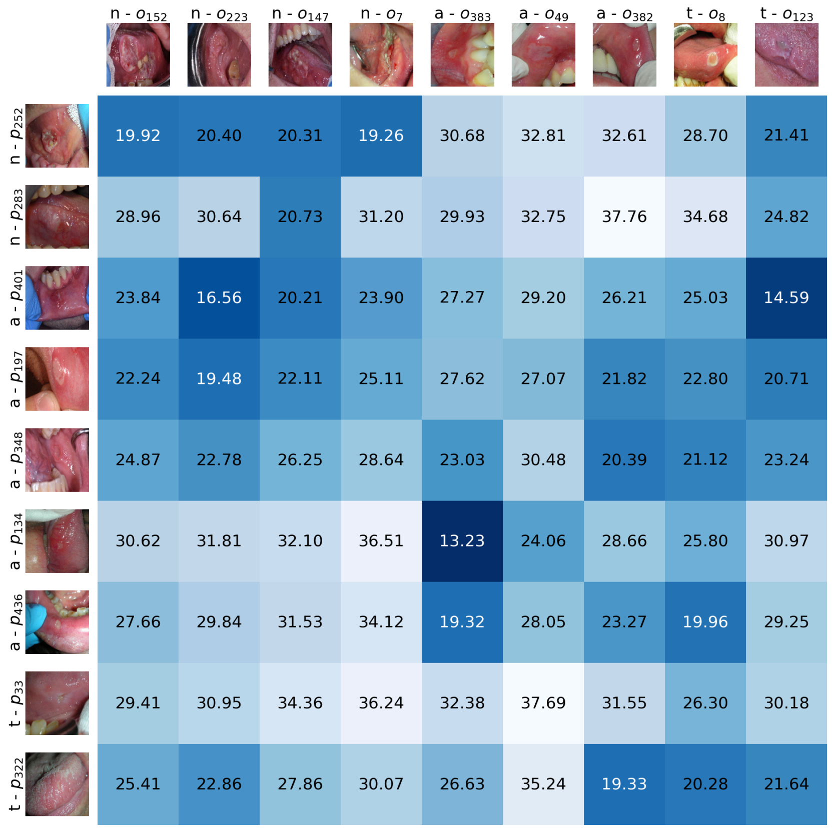


A Pivot Tree: instances laid on a tree structure route predictions. Similarity to instances determines routing, and leaves are also associated with a classification label.

Pivot Tree: selecting pivots

How do we choose *pivots*? We don't. We compute a similarity matrix S: **learning pivots is data-driven!**

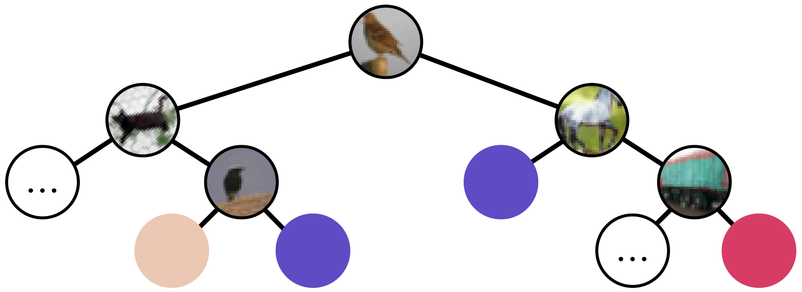
Provided a similarity metric $s : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is given, constructing S is data-agnostic!



A similarity matrix induced on Pivot Tree's training data (oral cancer detection).

Pivot Tree: inducing a tree

With a (new) feature matrix S, inducing a Pivot Tree can be reduced to any of the existing state-of-the-art tree-induction algorithms, e.g., CART, C.5, etc.



An induced Pivot Tree.

```
def induce_split(similarities, labels):
  splits_values = candidate_splits_per_feature(similarities)
  splits = partition(labels, splits)
  impurities = entropy(partitions)

  return partitions[argmin(impurities)]
```

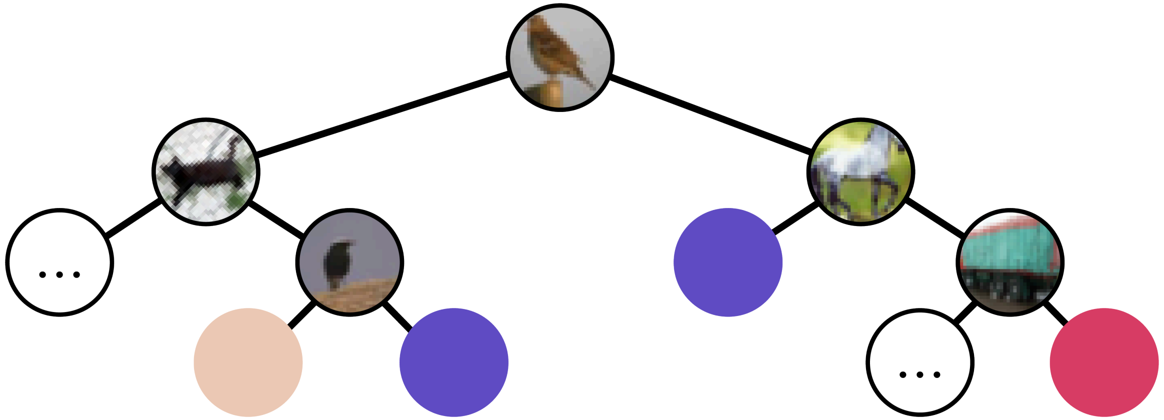
Pivot... anything

Pivots selected by Pivot Tree constitute a set P of data... that we can later re-use for learning another downstream model!

- k-nearest neighbor... defined on P
- decision tree... induced on P
- ...

Pivot Tree

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



An induced Pivot Tree.

Experiments

Explainability

- How faithful is Pivot Tree?
- How complex is Pivot Tree?
- How stable is Pivot Tree?

Competitors

- ε-ball
- k-NN

Datasets

- 11 tabular
- 5 time series
- 3 image
- 5 text

How faithful is Pivot Tree?

Downstream model: CART decision tree.

F1-score

	Random	Random per class	K-means Pivots	K-means Pivots	ε-ball	Pivot Tree S	Pivot Tree C
$mean \uparrow$	.60	.60	**.62**	*.61*	*.61*	**.62**	.58
$std \downarrow$	**.22**	**.22**	**.22**	**.22**	*.23*	**.22**	.27
$rank \downarrow$	5.0	4.6	4.0	*3.6*	3.7	**3.1**	3.9

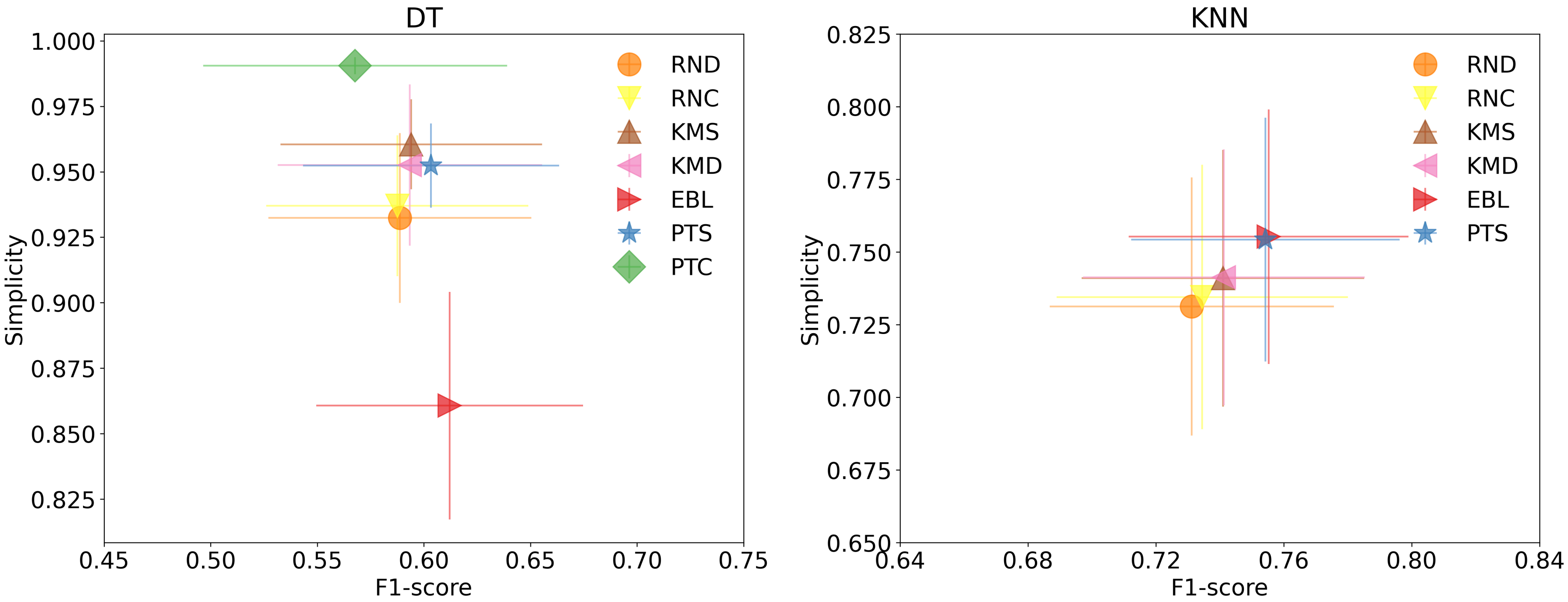
How faithful is Pivot Tree?

Downstream model: k-NN.

F1-score

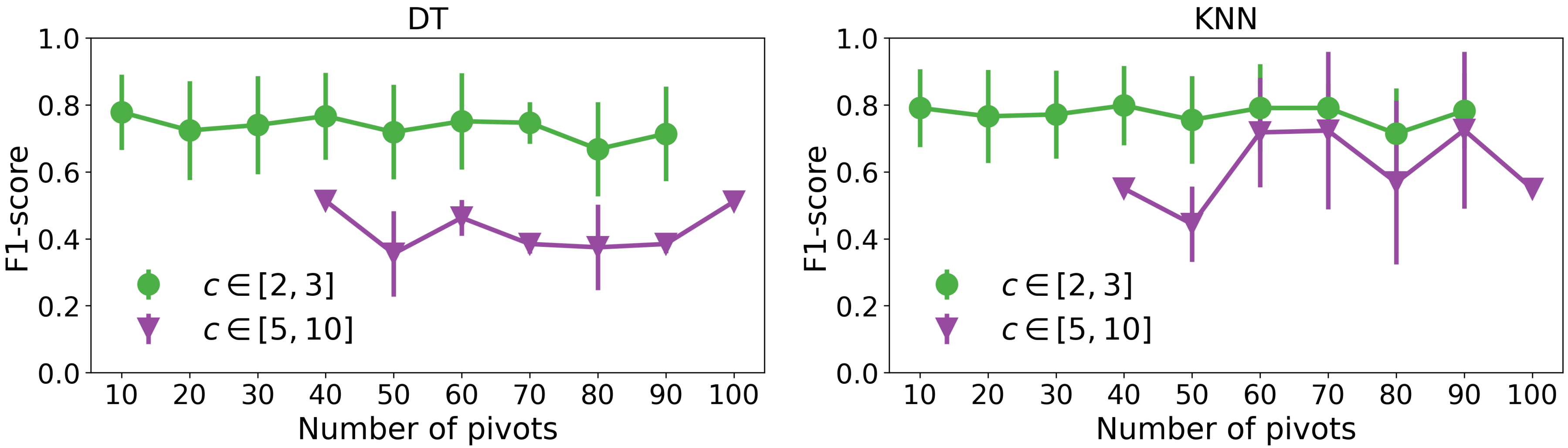
	Random	**Random per class**	**K-means Pivots**	**K-means Pivots**	**ε-ball**	**Pivot Tree S**	**Pivot Tree C**
$mean \uparrow$	.72	.72	*.73*	*.73*	**.74**	*.73*	*.73*
$std \downarrow$	*.19*	*.19*	**.18**	*.19*	*.19*	*.19*	*.19*
$rank \downarrow$	4.8	4.1	*3.0*	3.1	**2.2**	3.5	3.5

How complex is Pivot Tree?



Scatter plot of $f1$ score and simplicity of Pivot Tree and its competitors, on a decision tree (left) and k-NN (right) model.

How stable is Pivot Tree?



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

A real-world study: oral cancer detection

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

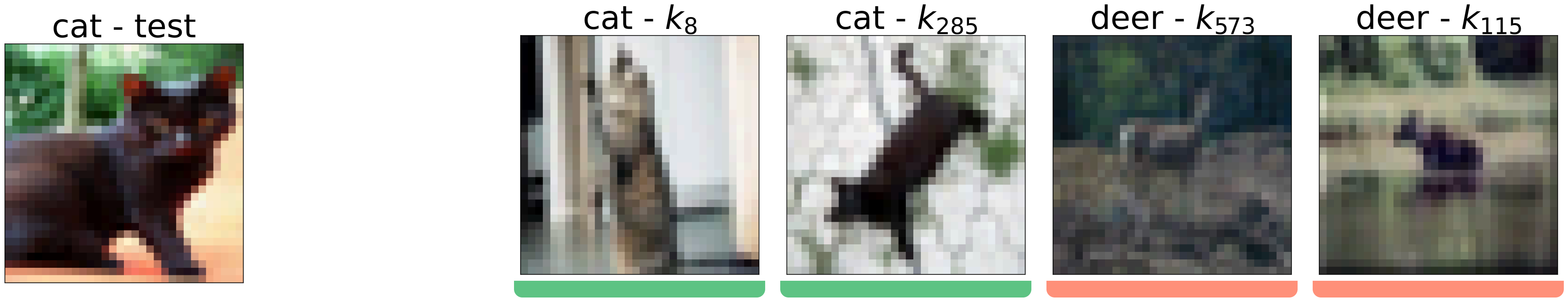
- 535 images
- 3 classes
- Publicly available

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

Interpretable Machine Learning for Oral Lesion Diagnosis through Prototypical Instances Identification. Cascione et al.

Qualitative results

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.

Qualitative results

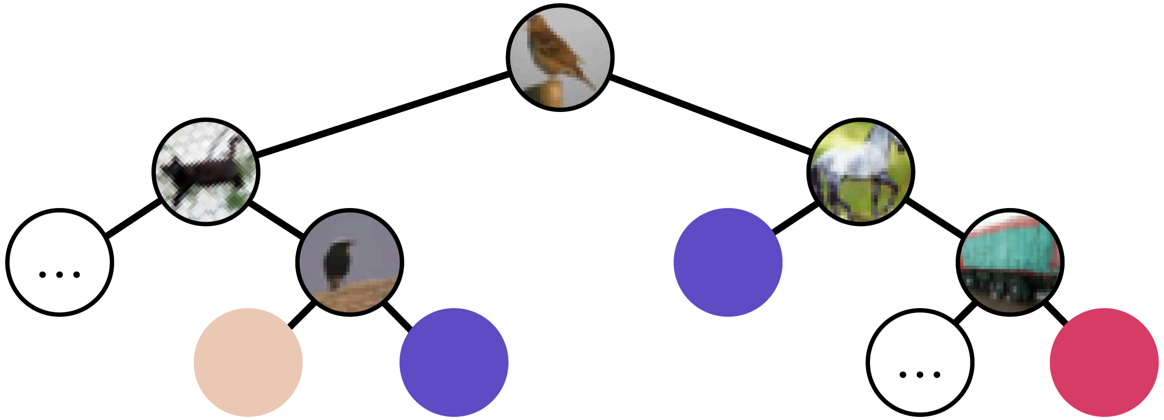
Cifar 10

An instance (left), and pivots extracted by Pivot Tree. The bird has high similarity with all the pivots.

Pivot Tree

An explainable by-design case-based algorithm.

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



github.com/msetzu/pivottree

See you at the poster session!