

Evaluation of XAI on ALS 6-Months Mortality Prediction

Evaluation for the iDPP Lab on Intelligent Disease Progression Prediction at CLEF 2022

T.M. Buonocore, G. Nicora, A. Dagliati, E. Parimbelli



Evaluation of XAI on ALS 6-Months Mortality Prediction

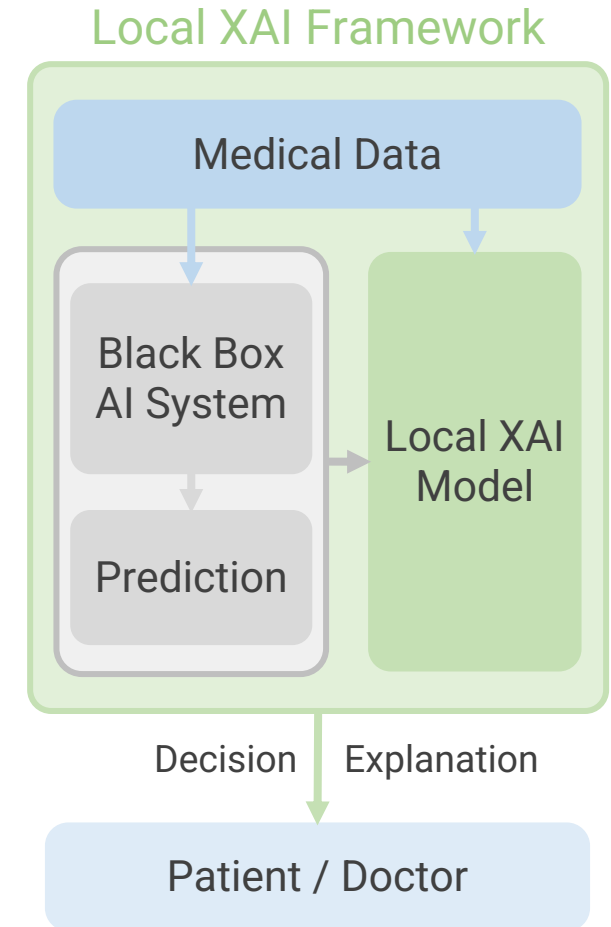
Evaluation for the iDPP Lab on Intelligent Disease Progression Prediction at CLEF 2022

T.M. Buonocore, G. Nicora, A. Dagliati, E. Parimbelli

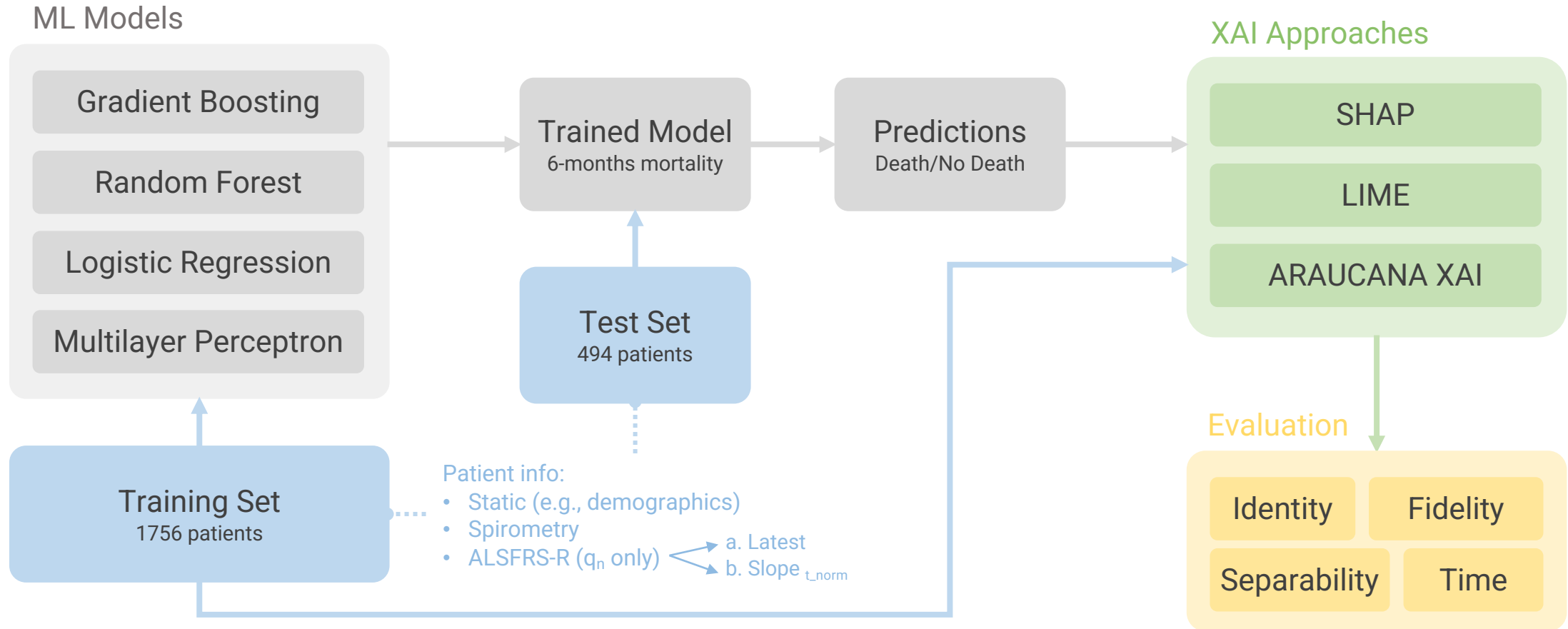


eXplainable AI (XAI)

- Focus on the implementation of explanatory models for **opaque AI systems** (e.g., DNNs) to provide:
 - A general understanding of the classifier behaviour (**global explainability**)
 - An insight about the internals of the prediction for a single instance (**local explainability**)
- The “right for an explanation” is also demanded by regulations (EU’s GDPR and Artificial Intelligence Act)
- Local XAI is crucial in **high stakes applications** such as **healthcare**, where patients/stakeholders often wish to understand the **reasons** behind a particular prediction

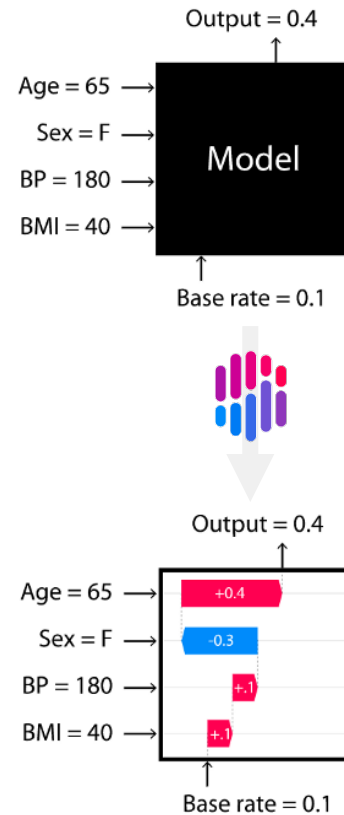


ALS XAI Experimental Workflow



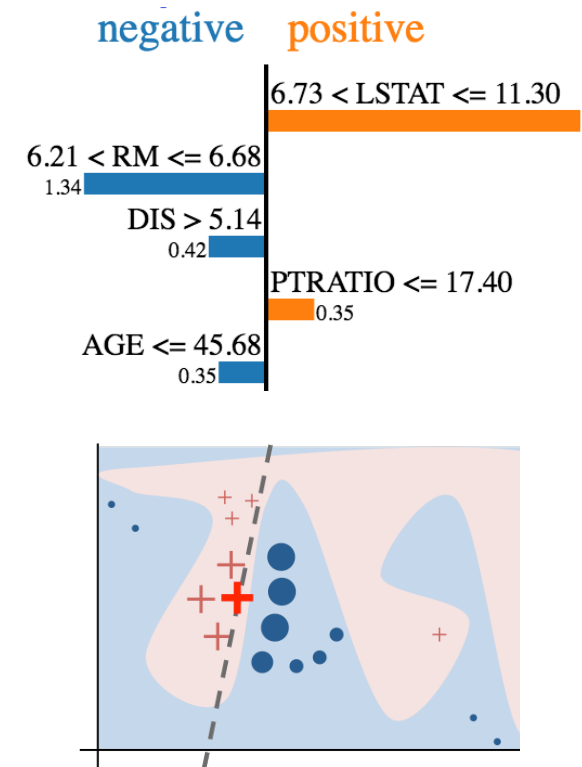
SHAP

- Game theoretic approach
- Local XAI conveyed through **feature importance**: Shapley values are used to decompose the final predicted probability assigning contribution to each feature



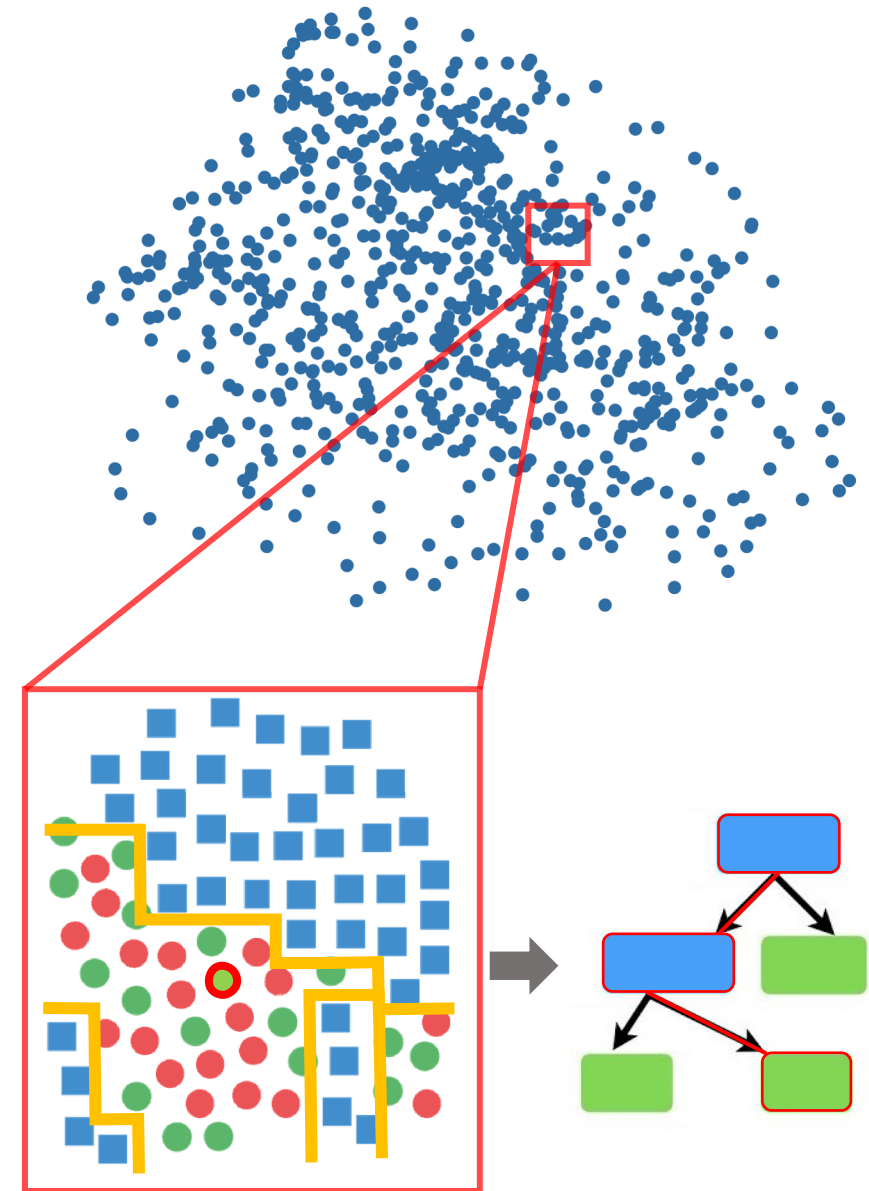
LIME

- Linear glass-box surrogate approach
- Local XAI conveyed through **feature importance**: the black-box model behaviour is approximated locally by using an interpretable-by-design model, e.g., logistic regression.



Araucana XAI

- Post-hoc local XAI approach based on **CART**
 - Ability to deal with **non-linear decision boundaries**
 - Improved fidelity to the original model
 - Native support to both classification and regression prob.
 - Explanation = **navigable tree structure**
- Given a single instance x :
 1. Compute $D = \text{dist}(x, z)$ for each training element z
 2. Define subset T_n as the closest N elements to x
 3. Augment T_n with **SMOTE oversampling** (optional)
 4. **Re-label** the samples of T_n (or $T_n \cup S$) with the class predicted by the **predictive function** f of the black-box classifier. Define the explainer set E as $T_n \cup S \cup x$
 5. Train e as a decision tree on E . Optionally **prune** it
 6. **Navigate** e according to the feature values of x



Quantitative Evaluation of Explanations

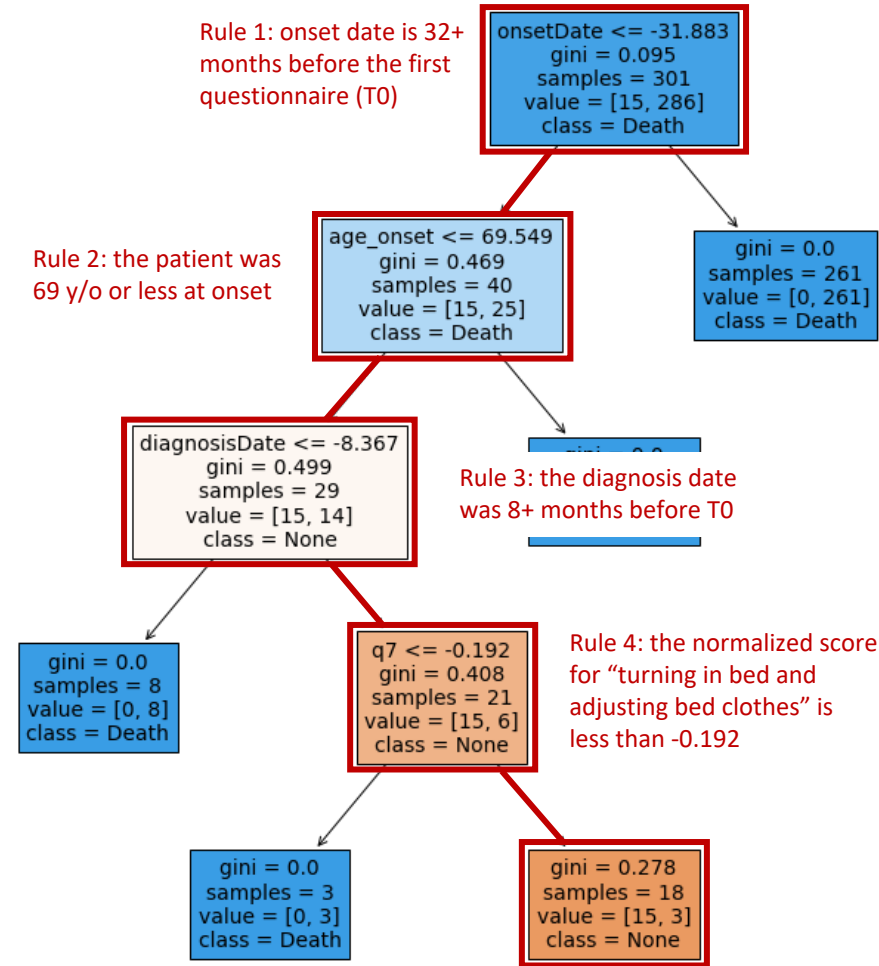
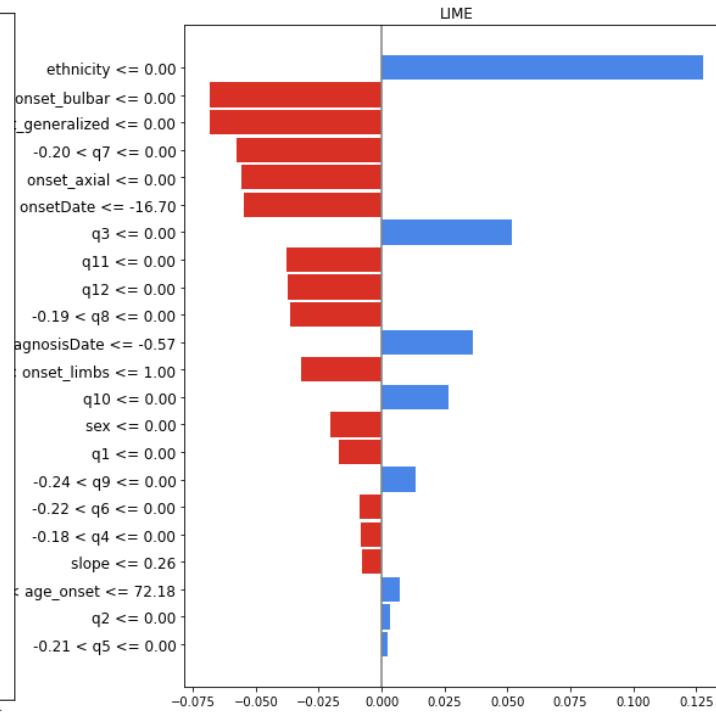
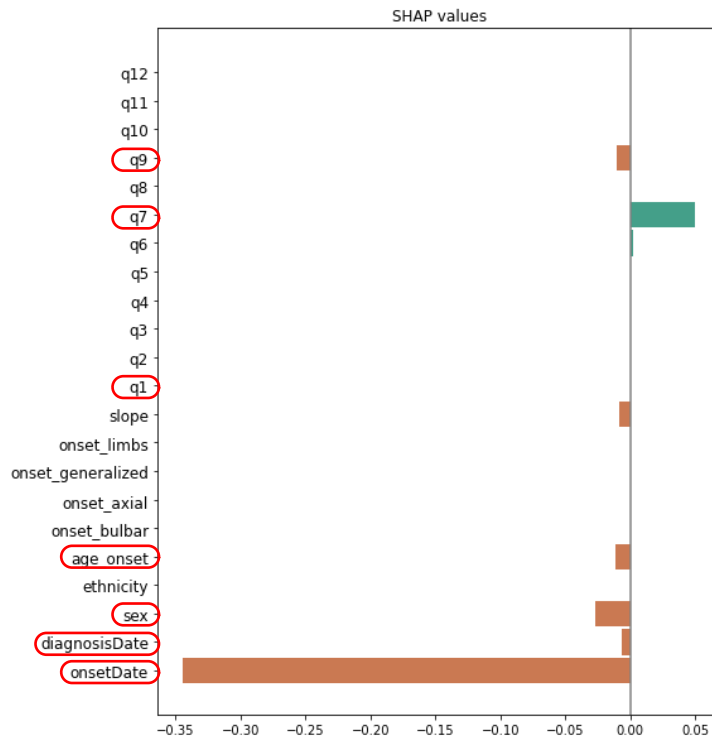
- **Identity:** if there are two identical instances, they must have the same explanations
- **Fidelity:** concordance between the predictions of the XAI surrogate model and the original ML model
- **Separability:** if there are 2 dissimilar instances, they must have dissimilar explanations
- **Time:** average time required by the XAI method to output explanations across the entire test set

	Identity*	Fidelity (higher is better)			Separability (lower is better)			Time (s) (lower is better)		
		SHAP	LIME	ARAU	SHAP	LIME	ARAU	SHAP	LIME	ARAU
GB	N.A.	1	0.99	1	0.0005	0	0.03	0.01	484	13.5
RF		1	0.99	1	0	0	8.2E-6	7.7	539	48.4
LR		1	0.99	1	0	0	0.08	4.6	480	9.6
MLP		1	0.99	1	0	0	0.001	4.8	484	12.3

* no 2 identical instances are present in the provided test set

Qualitative Evaluation

onsetDate	Age_onset	Onset_bulbar	diagnosisDate	slope	q1	q7	y	\hat{y}
-66.8	68.40	0	-0.866	0.029	0	-0,17	0	0



Conclusion: the patient has been classified as *Alive* because his onset date is 32+ months before T0 **and** his age at that time was less than 69 **and** the diagnosis date was less than 8 months before T0 **and** the normalized score for "turning in bed and adjusting bed clothes" is greater than -0,192

Conclusion

- **Major takeaways**

- **Quantitative:** the three different XAI approaches are all valid for ALS 6-months mortality prediction, no definitive superior performance over the others
- **Qualitative:** ARAU offers the opportunity of delivering explanations through a simple and understandable navigable tree interface, while LIME and SHAP don't. In terms of feature importance, ARAU and SHAP often agree while LIME usually provides different explanations

- **Limitations**

- No clinical experts were involved in the pilot
- Unoptimized ML models
- Unrefined approach for time-dependent data



Thanks for your attention