Understanding Machine Learning Explainability Models in the context of Pancreatic Cancer Treatment

José Bobes-Bascarán, Ángel Fernández Leal, Eduardo Mosqueira-Rey, David Alonso-Ríos, Elena Hernández-Pereira, Vicente Moret-Bonillo

Centro de Investigación CITIC, University of Coruña, Spain
Correspondence: jose.bobes@udc.es

DOI: *https://doi.org/10.17979/spudc.000024.28*

Abstract: The increasing adoption of artificial intelligent systems at sensitive domains where humans are particularly, such as medicine, has provided the context to deeply explore ways of making machine learning models (ML) understandable for their final users. The success of such systems require the trust of their users, and thus there is a need to design and provide methods to understand the decisions made by such systems. We start from a public Pancreatic Cancer dataset and experiment with different ML models on a diagnosis scenario with the goal to decide whether a patient should be prescribed with a chemotherapy treatment. To validate the diagnosis results we explore different explainability approaches: Decision Tree, Random Forest, and model agnostic ad-hoc models, and compare them against a standard Pancreatic Cancer treatment set of rules. The increasing adoption of artificial intelligent systems at sensitive domains where humans are particularly, such as medicine, has provided the context to deeply explore ways of making machine learning models (ML) understandable for their final users. The success of such systems require the trust of their users, and thus there is a need to design and provide methods to understand the decisions made by such systems. We start from a public Pancreatic Cancer dataset and experiment with different ML models. To validate the diagnostic results we explore different explainability approaches: Decision Tree based approach, Random Forest based approach, and different model agnostic ad-hoc approaches, and we compare them against a standard Pancreatic Cancer treatment set of rules.

1 Introduction

When creating Machine Learning (ML) models there is normally a trade of between interpretability and accuracy. While the former aims to create models that can be understand by their end-users, the search for accuracy often require complex models that are not easy to interpret.

In recent years ML models are being deployed covering a wide range of scenarios were it is of crucial importance the ability to understand how those models behave and reach a certain conclusion. It is equally important to reach a certain level of accuracy so that the models could be trust in real scenarios.

In this research, we continued from our previous work on AL Bobes-Bascarán et al. (2021) Bobes-Bascarán et al. (2023), where we first experimented with generated synthetic data and an Active Learning approach, and then followed with a real dataset introducing medical doctors in the loop of a therapy selection model for pancreatic cancer. On our previous work the goal was to overcame the scarcity of data available on the Pancreatic Cancer context by incorporating humans into the ML loop Mosqueira-Rey et al. (2022b). The focus is now on applying several explainability techniques to get useful insight about the underlining models.

We found two different concepts in the literature that refer to the quality of a system to be understood by its end-users: Explainability and Interpretability.

Explainability is the ability to describe how a model could reach a certain prediction or classification result so that it can be understand by its end-users.

Interpretability is intrinsic to the model itself and refers to the fact that end-users could interpret the relationship between the model inputs and its outputs.

On the one hand, we found that better interpretability and understandability leads to better trust and eases the adoption of AI systems. On the other hand, better accuracy require more complex models. This trade-of is a key factor when dealing with models deployed on domains such as healthcare.

To overcome the understandability issue eXplainable AI (XAI) Gunning (2017): "XAI will create a suite of machine learning techniques that enables human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners".

Some authors implicitly assumes that the ease of understanding and clarity targeted by XAI techniques for the model at hand, results in different application purposes, such as a better trustworthiness of the model's output by the audience. They do synthesize and enumerate definitions for these XAI goals, in example: Trustworthiness, Causality, Informativeness and Confidence Arrieta et al. (2020).

Nowadays, there can be found several methods that help to enhance ML models with easier to understand ad hoc exlainability models. Among others features, they provide several charts for both local and global explanations which is a very interesting characteristic for non-experts users. A set of tools helping in the different processes has been described by Mosqueira-Rey et al. (2022a).

2 Dataset

The dataset used in this work was obtained from The Cancer Genome Atlas Program Tomczak et al. (2015) published by the USA National Cancer Institute (NCI) and the National Human Genome Research Institute that contains a complete database of pancreatic cancer cases. It is composed of several research projects, among them, the TCGA-PAAD, currently with 185 diagnosed cases with all the necessary details.

There are 185 cancer positive patients (83 female and 102 male) and for each one we were interested in determining whether chemotherapy treatment was indicated or not based on the diagnostic information available. From the available cases 117 (60%)of them correspond to a chemotherapy treatment and 64 correspond to "Other".

It includes patient demographic information, family history, diagnosis, treatments, and genomic, epigenomic, transcriptomic and proteomic data. It issues information such as: the "stage event" which describes the pathological state of the tumor, the "clinical data" which describes the characteristics of the tumor and the occurrence of "new tumor events" which describes the patient follow-up.

3 Data Preparation

The aim of the experiment we have designed is to classify whether to prescribe a Chemotherapy treatment or not, based on some of the most relevant patient features available in the dataset.

A data curation process has been performed including (1) handling Missing Values/Null Values, (2) removing redundancy, (3) simplifying the target values with only *Chemotherapy* or *Other*, (4) converting some of the categorical variables into numerical, (5) using One-hot encoding for the rest of categorical features, (6) and dropping unneeded features.

The selected features are: pathologic_stage, pathologic_t, pathologic_m, pathologic_n, primary_diagnosis, tissue_or_organ_of_origin, age_at_index, and gender. We refer the reader to the

official NIH GDC documentation were a detailed description is provided.

The target variable is the therapy type to be prescribed to the patient. It can take several values, but we are interested only in whether or not it is chemotherapy. The possible values are 'Chemotherapy', or 'Other', as we have recoded all the alternative values (Hormone Therapy, Vaccine, and Ancillary) as 'Other'.

After preparing the database we performed a split between the training and the test data of 70 and 30% respectively.

4 Generating the models

With the goal of illustrating the balance between the accuracy and the explainability capabilities of the models, we have created several models and discuss their advantages and disadvantages.

4.1 Decision tree

A decision tree is a decision support hierarchical model that uses a tree-like model of decisions and their possible consequences.

We do create a decision tree with only 6 levels, and a minimum of 5 samples per leaf, using the training data.

	precision	recall	f1-score	support
Chemotherapy	0.70	0.88	0.78	34
Other	0.69	0.41	0.51	22
accuracy			0.70	56
macro avg	0.69	0.65	0.65	56
weighted avg	0.70	0.70	0.68	56

Text representation

```
|--- primary_diagnosis_Neuroendocrine carcinoma, NOS <= 0.50
|   |--- age_at_index <= 79.50
|   |   |--- age_at_index <= 62.50
|   |   |   |--- age_at_index <= 59.50
|   |   |   |   |--- age_at_index <= 56.50
|   |   |   |   |   |--- gender_female <= 0.50
|   |   |   |   |   |   |--- class: Chemotherapy
|   |   |   |   |   |--- gender_female >  0.50
|   |   |   |   |   |   |--- class: Other
...
```

Decision Tree Graphical representation (partial chart)

For the explainability aspect, we can count on the graph representation itself that clearly identifies which is the decision path of an instance. For each of the decision nodes, the condition is evaluated and we descent through the appropriate branch, until we get to a leaf level. For each of the nodes the *Gini index*, the number of samples and the specific class is represented.

Furthermore, we enhance the explanation using **Permutation Importance** that is an algorithm that computes importance scores for each of the feature variables of a dataset. These scores are determined by computing the sensitivity of a model to random permutations of feature values. After permuting a certain feature, t-he increase in the prediction error of the model determines the importance score of the feature (see 2).

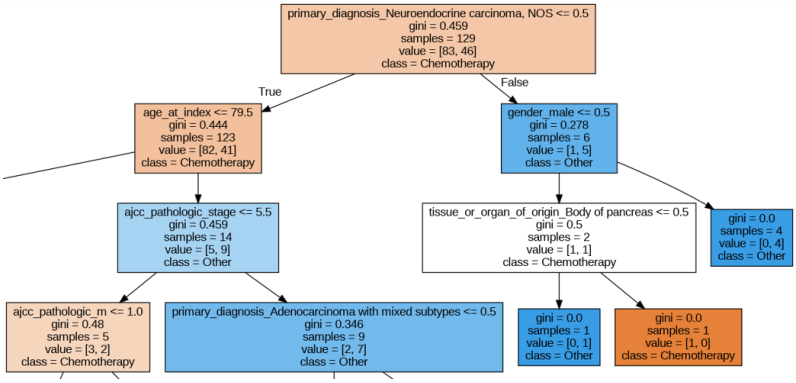

Figure 1: Decision Tree: graphical representation (partial chart).

Weight	Feature
0.4251	age_at_index
0.1077	primary_diagnosis_Neuroendocrine carcinoma, NOS
0.0990	ajcc_pathologic_stage
0.0922	tissue_or_organ_of_origin_Head of pancreas
0.0653	gender_female
0.0565	tissue_or_organ_of_origin_Pancreas, NOS
0.0512	primary_diagnosis_Adenocarcinoma with mixed subtypes
0.0402	ajcc_pathologic_m
0.0376	tissue_or_organ_of_origin_Body of pancreas
0.0251	gender_male
0	ajcc_pathologic_n
0	ajcc_pathologic_t
0	tissue_or_organ_of_origin_Tail of pancreas
0	primary_diagnosis_Carcinoma, undifferentiated, NOS
0	primary_diagnosis_Infiltrating duct carcinoma, NOS
0	primary_diagnosis_Mucinous adenocarcinoma
0	tissue_or_organ_of_origin_Overlapping lesion of pancreas
0	primary_diagnosis_Adenocarcinoma, NOS

Figure 2: Decision Tree: permutation importance.

4.2 Random forest

A Random Forest is an ensemble method which consist of a set of decision trees that combined, normally provide better accuracy than a single tree. Even if Random Forests are more accurate than Decision Trees, they are way more complex in terms of understandability.

	precision	recall	f1-score	support
Chemotherapy	0.64	1.00	0.78	34
Other	1.00	0.14	0.24	22
accuracy			0.66	56
macro avg	0.82	0.57	0.51	56
weighted avg	0.78	0.66	0.57	56

5 Understanding the models

Firstly, we would like to emphasize that the decision tree is already a white-box model and offers a good explainable capabilities by nature. Both the textual and the graphical representations suit many scenarios as it is easy to understand and interpret on single instances.

Nevertheless, using Permutation importance we can synthesize the rules produced by the tree data structure in an aggregated manner. Observing figure 2 we can see that *age at diagnosis*, the *primary diagnosis* being *Neuroendocrine carcinoma*, the *pathologic stage* of the cancer, and the fact of having the *tumor localized at the head of the pancreas* are the more relevant features of the model.

Secondly, for the Random Forest (RF) created using bootstrap aggregation by combining several decision trees, even if the first results deliver a lower accuracy than the DT, the ensemble models in general provide better generalization capabilities. As the RF is more complex than a simple DT we enhanced the RF by means of creating several ad-hoc explainability models on top or it. We have chosen widespread methods such SHAP and LIME.

SHAP is a method based on the cooperative game theory proposed by Shapley Shapley (1953).

It handles explainability by attributing a numerical value to each of the features of the model. This number represents the contribution of the feature to the model prediction or classification result.

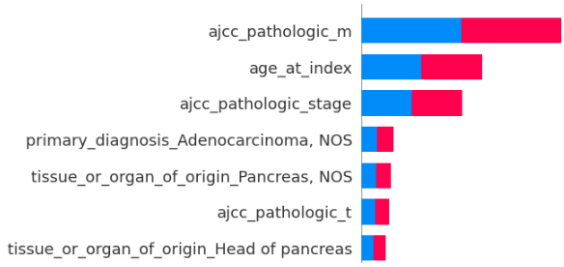

Figure 3: SHAP summary plot for variable importance.

The summary plot 3 give us the variable importance. Features with high predicting power are shown at the top and the ones with low predicting powers are shown at the bottom.

We could also see the contribution of each feature into the prediction probability. The redder the color, the higher the value and vice versa. Also, when the value is on the positive side, it contributes to the class 'Chemotherapy' prediction result probability and vice versa.

LIME, proposed by Ribeiro et. al Ribeiro et al. (2016), is the acronym for Locally Interpretable

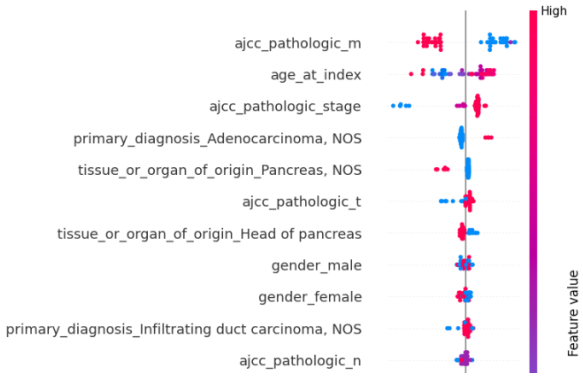

Figure 4: SHAP summary plot for impact on output.

Model-agnostic Explanations. It tries to understand the features that influence a prediction on a single instance focusing on a local level where a linear model is enough to explain the behavior.

We present local explanations for two of the patients available on the dataset.

The first patient (index 3) is a woman who was diagnosed an infiltrating duct carcinoma at the tail of the pancreas when she was 58 years old. The TNM is T3, N0, M1 and the pathologic stage is IV. The model selects a Chemotherapy treatment 6.

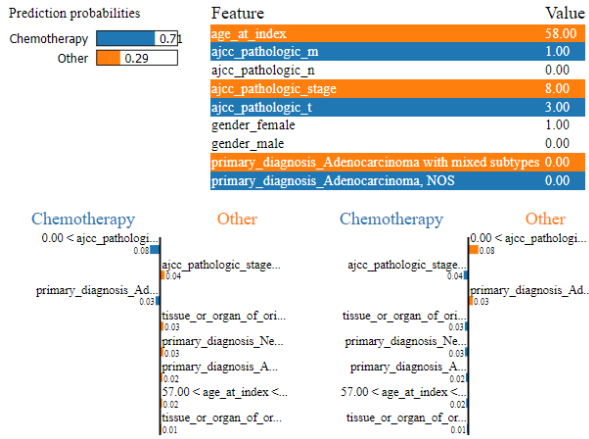

Figure 5: LIME patient 3 resume.

The second patient (index 4) is a woman who was diagnosed and infiltrating duct carcinoma in a none specific area of the pancreas when she was 66 years old. The TNM is T3, N0, M0 and the pathologic stage is IIA. The model selects a Chemotherapy treatment. The model selects a Chemotherapy treatment 5.

6 Conclusions

We have described the importance of interpreting and understanding a ML models, especially in a sensible context, as it is the case of the pancreatic cancer domain.

Through the construction of two different ML models we have illustrated how a simple model could be easily understood with a visual representation, and how difficult it will be to compre-

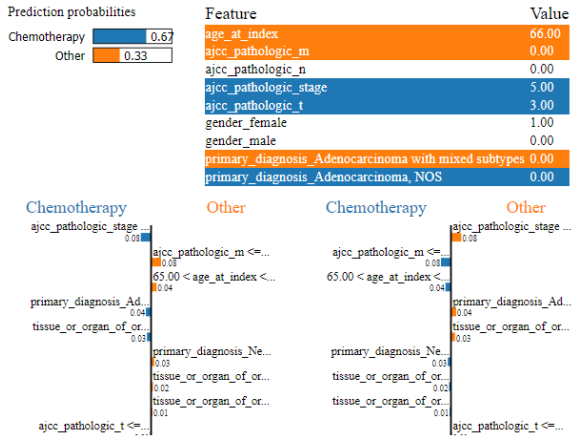

Figure 6: LIME patient 4 resume.

hend a more sophisticated one, as the ensemble random forest.

Several ad-hoc models have been created using the SHAP and LIME methods to provide further explainability features over the original models. Those models provide an easy to understand variable importance for both global and local explanations, that can be combined to fully interpret and rely on a ML model.

We plan to compare the results obtained, to a set of guidelines widely adopted among pancreatic cancer professionals. The idea will be to determine if the explainable model is able to reach the same expert criteria.

In our study, the pathologic M, and the age at the first diagnosis are the most relevant characteristics, but perhaps the medical literature do not agree on those attributes.

7 Acknowledgments

This work has been supported by the State Research Agency of the Spanish Government (grant PID2019-107194GB-I00/AEI/10.13039/501100011033) and by the Xunta de Galicia (grant ED431C2022/44), supported in turn by the EU European Regional Development Fund. We wish to acknowledge support received from the Centro de Investigación de Galicia CITIC, funded by the Xunta de Galicia and the European Regional Development Fund (Galicia 2014-2020 Program; grant ED431G 2019/01).

Bibliography

A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.

J. Bobes-Bascarán, E. Mosqueira-Rey, and D. Alonso-Ríos. Improving medical data annotation including humans in the machine learning loop. *Engineering Proceedings*, 7(1):39, 2021.

J. Bobes-Bascarán, A. Pérez-Sánchez, E. Mosqueira-Rey, D. Alonso-Rıos, and E. Hernández-Pereira. Using active learning to improve the treatment selection on pancreatic cancer patients. In *Proceedings of V XoveTIC Conference. XoveTIC*, volume 14, pages 70–72, 2023.

D. Gunning. Explainable artificial intelligence (xai). *Defense advanced research projects agency (DARPA), nd Web*, 2(2):1, 2017.

E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, and J. Bobes-Bascarán. A classification and review of tools for developing and interacting with machine learning systems. In *Proceedings of the 37th Annual ACM Symposium on Applied Computing*, pages 1083–1092. Association for Computing Machinery, New York, NY, USA, 2022a. ISBN 9781450387132.

E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and A. Fernández-Leal. Human-in-the-loop machine learning: A state of the art. *Artificial Intelligence Review*, 2022b.

M. T. Ribeiro, S. Singh, and C. Guestrin. "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*, KDD '16, page 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322.

L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games II*, pages 307–317. Princeton University Press, Princeton, 1953.

K. Tomczak, P. Czerwińska, and M. Wiznerowicz. The cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology*, 19(1A):68–77, 2015.