



HAL
open science

Data Exploration Based on Local Attribution Explanation: A Medical Use Case

Elodie Escriva, Emmanuel Doumard, Jean-Baptiste Excoffier, Julien Aligon,
Paul Monsarrat, Chantal Soulé-Dupuy

► **To cite this version:**

Elodie Escriva, Emmanuel Doumard, Jean-Baptiste Excoffier, Julien Aligon, Paul Monsarrat, et al..
Data Exploration Based on Local Attribution Explanation: A Medical Use Case. 27th European
Conference Advances in Databases and Information Systems (ADBIS 2023), Sep 2023, Barcelone,
Spain. pp.315-323, 10.1007/978-3-031-42941-5_27 . hal-04239173

HAL Id: hal-04239173

<https://hal.science/hal-04239173>

Submitted on 12 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Data exploration based on local attribution explanation: a medical use case

Elodie Escriva^{1,2}[0000-0003-3618-967X], Emmanuel
Doumard^{3,4}[0000-0003-1185-9630], Jean-Baptiste Excoffier¹[0000-0002-9313-2429],
Julien Aligon²[0000-0002-1954-8733], Paul Monsarrat^{3,5,6}[0000-0002-5473-6035],
and Chantal Soulé-Dupuy²[0000-0002-2637-724X]

¹ Kaduceo, Toulouse (FR)

`elodie.escriva@kaduceo.com`

² Université de Toulouse-Capitole, IRIT, (CNRS/UMR 5505), Toulouse (FR)

³ RESTORE Research Center, Toulouse (FR)

⁴ Université de Toulouse-Paul Sabatier, IRIT, (CNRS/UMR 5505), Toulouse (FR)

⁵ Artificial and Natural Intelligence Toulouse Institute ANITI, Toulouse (FR)

⁶ Oral Medicine Department, Toulouse (FR)

Abstract. Exploratory data analysis allows to discover knowledge and patterns and to test hypotheses. Modelling predictive tools associated with explainability made it possible to explore more and more complex relationships between attributes. This study presents a method to use local explanations as a new data space to retrieve precise and pertinent information. We aim to apply this method to a medical dataset and underline the benefit of using explanations to gain knowledge. In particular, we show that clusters based on local explanations, combined with decision rules, allow to better characterise patient subgroups.

Keywords: Medical Data Exploration · Explainable IA · Machine Learning.

1 Introduction & Related Work

As data availability increased in the last decades, exploratory data analysis techniques have arisen to investigate data and discover patterns, make and test hypotheses with the help of statistics, graphical representation, clustering or predictive tools. In particular, Bottom-Up approaches consists in finding patterns and gaining insight by analysing data without making *a-priori* hypotheses [15, 18]. Among the tools for exploratory data analysis, predictive approaches, primarily through machine learning, have made it possible to capture more complex statistical phenomena in the data that classical statistical techniques cannot understand. However, due to the lack of explanation of the predictions, the Machine Learning (ML) black box effect is a limitation for sensitive areas, such as those involving human lives. In the medical field, patients may legally ask for the reasons behind a decision, which may be problematic when ML modelling is used in the decision-making process [9].

A way to better understand machine learning modelling and the prediction they produce lies in the Explainability domain (XAI). In particular, local explanations allow investigation of the reasons behind the model prediction for each instance. Local attribution methods like LIME [16], SHAP [13] or Coalitional-based methods [7] explain the prediction by computing the impact of each attribute for each instance. All these methods produce explanations called "influences", each with different strengths and weaknesses as detailed in [4].

Research has focused on the applicability, evaluation and uses of explanations, especially in the medical field. Influences can be used for multiples purposes: select attributes [11,17], find attributes relationships [2,10], determine subgroups and recommend instances based on influences [5,6], extract knowledge in data from influences [14]. Each paper shows that using influences is of great interest in the modelling pipeline and uses influences as a new data space to explore. However, not all papers strictly compare the benefit of using explanations to gain knowledge, compared to a classical analysis of raw data. And as explanations provide information on the modelling and complex interactions of the dataset, the contribution of influences must be assessed against raw data.

Then, our objective is to apply a bottom-up exploratory data analysis approach on a medical dataset, on both explanations and raw data, to highlight and compare the knowledge retrieved in both data spaces. We show that explanations can allow a deeper dataset investigation. This study can also show the usefulness of seeing explanations not only as an outcome but also as a tool.

The paper is structured as follows: we introduce our method and the dataset used in Section 2, demonstrate the usefulness of explanation-based analysis in Section 3 and discuss results and perspectives in Section 4.

2 Methods

2.1 Dataset

To enable reproducible results, we use an open-source dataset: Acute Inflammation dataset⁷. The Acute Inflammation dataset was created to develop an expert system for urinary disease. It consists of 120 patients, described by six attributes: Temperature (35°C - 42°C), Occurrence of nausea (*yes-no*), Lumbar pain (*yes-no*), Urine pushing (continuous need for urination, *yes-no*), Micturition pain (*yes-no*) and Burning of urethra, itch, swelling of urethra outlet (abbreviated as Urethra burning, *yes-no*). Each patient can have two different diseases of the urinary system: acute inflammation of urinary bladder (AIUB) and acute nephritis of renal pelvis origin. Patients may suffer from both diseases simultaneously, so this dataset is a multi-output problem. We only focus on the AIUB disease to have a binary classification problem. Medical staff defined AIUB as "*a sudden occurrence of pains in the abdomen region and the urination in form of constant urine pushing, micturition pains and sometimes lack of urine keeping. Temperature of the body is rising, most often not above 38C. The excreted urine is turbid and sometimes bloody*" [3].

⁷ Dataset: <https://archive.ics.uci.edu/ml/datasets/Acute+Inflammations>

2.2 Modelling

The proposed method aims to analyse and explore datasets through modelling and influences. Based on a dataset of interest consisting of patients' medical records and their disease diagnosis, this method allows an understanding of interactions between patients' characteristics and the disease. It is divided into three parts, inspired by [5]:

(1) The first one consists of *ML predictive modelling*, to evaluate the risk of AIUB disease for each patient based on the understanding of the complex statistical relationship of the dataset. An XGBoost model, a boosted tree ensemble technique [1], is used for its efficiency. We use a nested cross-validation (CV) procedure to provide unbiased modelling (hyperparameters optimization with an inner 5-fold CV) and to evaluate performances and compute local explanations (through an outer 5-fold CV).

(2) Second step is the *explanation of the modelling* to provide individual explanations of the prediction for each patient, corresponding to individual risk and protective factors. TreeSHAP [12], a local attribution XAI method for tree-based predictive models, is used to compute influence explanations.

(3) Last step consists of *identifying subgroups of similar patients* to discover local patterns in the data and explain the subgroups characteristics. K-Medoids algorithm is used for the clustering task to ensure robustness against outliers, while the optimal number of groups was chosen with the Silhouette score. K-medoids algorithm is used on the influence explanations from step (2), with the advantages of taking into account the non-linear interactions discovered by the model while having all features at the same unit. Decisions rules for all clusters are computed with Skope-Rules algorithm [8]. Rules are computed to ensure perfect precision and recall of all rules: all instances of the cluster respect the rule, and all instances respecting the rule belong to the cluster.

3 Results

3.1 Raw data Analysis

Populations and statistical tests. Table 1 shows the main characteristics of the dataset using raw data only, with results from statistical tests performed on AIUB and Non-AIUB patients: Student tests for quantitative attributes and Chi-squared test for qualitative attributes. Three attributes are defined as statistically significant to detect AIUB: Lumbar pain, Urine pushing and Micturition pain. Patients with lumbar pain seem to have less AIUB while having urine pushing and micturition pain correlate with an AIUB diagnosis.

Clustering and rule-based analysis. To create homogeneous groups of patients, one method consists of performing clustering. The optimal number of clusters was 11, based on the silhouette scores in Table 2. Table 3 shows the rules defined by Skope Rules to describe each cluster. Rules have a median of 2.5

Table 1: **Population characteristics.** Mean and standard deviation are presented for quantitative attributes, and numbers and proportions for binary qualitative attributes. P-values were adjusted using Bonferroni correction to control family-wise error rate.

		Total	Non-AIUB	AIUB	p-value
	Nb patients	120	61 (50.8)	59 (49.2)	
Quanti.	Temperature	38.72 (± 1.8)	39.15 (± 1.9)	38.29 (± 1.7)	0.0552
Quali.	Nausea	29 (24.2)	10 (16.4)	19 (32.2)	0.4224
	Lumbar pain	70 (58.3)	51 (83.6)	19 (32.2)	<0.01
	Urine pushing	80 (66.7)	21 (34.4)	59 (100.0)	<0.01
	Micturition pain	59 (49.2)	10 (16.4)	49 (83.1)	<0.01
	Urethra Burning	50 (41.7)	21 (34.4)	29 (49.2)	0.8814

Table 2: Silhouette Score for multiple numbers of clusters for Raw data.

K	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Raw	0.56	0.44	0.37	0.42	0.46	0.51	0.54	0.54	0.56	0.57	0.56	0.56	0.56	0.56

attributes per rule. All rules have perfect precision and recall with a maximum of three attributes, which is a small enough number of attributes to facilitate the interpretation of each rule. The most used attributes are urethra burning and temperature with six distinct occurrences, both previously defined as not significantly discriminating for AIUB diagnosis in Table 1. Only one cluster, Cluster 2, uses only significantly discriminating attributes. Also, having eleven clusters makes it challenging to easily understand the rules and clusters.

3.2 XAI analysis

Local post-hoc explanations. An XGBoost model was also trained and explained through SHAP method [13]. The model had an accuracy of 98.33%, a sensitivity of 96.72%, a specificity of 100% and an AUC ROC Score of 99.06%. Figure 1 shows the SHAP mean absolute influences and the distributions of influences based on the attribute value. The three most important attributes were Micturition pain, Urine pushing and Temperature. Micturition pain and Urine pushing increases the risk of having AIUB. On the contrary, a higher temperature decreases the probability of having AIUB. In particular, having urine pushing also seems to have less impact on the prediction than not having urine pushing. In contrast, Nausea and Urethra burning have little to no impact on the predictions. For nausea, SHAP describes that having them increases the risk of AIUB for some patients and a subgroup of patients is identified.

Figure 2 shows the distribution of influences only for patients having Nausea. Looking in details at these patients, they all suffer from lumbar pain, micturition pain and temperature above 40°C (which is higher than the dataset mean). There seems to be a subgroup of patients with a strong relationship between

Table 3: Decision Rules for clusters based on raw data, with the number of patient per cluster and the mean percentage of AIUB-risk.

	Rules	Nb	Mean %
1	Nausea = 1 & Urine pushing = 0	10	45.6
2	Lumbar pain = 0 & Urine pushing = 0	10	10.7
3	Nausea = 1 & Urethra burning = 1	9	72.2
4	Temperature < 39.85 & Micturition pain = 0 & Urethra burning = 1	10	13.0
5	Lumbar pain = 0 & Urethra burning = 1	20	97.1
6	Temperature < 38.95 & Temperature > 36.65 & Urine pushing = 0	13	11.0
7	Temperature < 38.95 & Lumbar pain = 0 & Micturition pain = 0	10	59.9
8	Nausea = 1 & Urine pushing = 1 & Urethra burning = 0	10	73.6
9	Temperature > 39.85 & Nausea = 0 & Urethra burning = 1	11	11.2
10	Lumbar pain = 0 & Micturition pain = 1 & Urethra burning = 0	10	97.1
11	Temperature < 36.65 & Urethra burning = 0	7	11.2

these four attributes. Moreover, for this subgroup of patients, there is a strong correlation between the attribute Urine Pushing and the presence of AIUB: when a patient has urine pushing, they have an AIUB; when they do not have urine pushing, there is no AIUB. This subgroup is probably best to study, as the nausea attribute may create a real-world bias due to its strong association with other attributes in the dataset.

Clustering and rule-based analysis. As one subgroup is already discovered, clustering can help to find other subgroups of interest. For clustering on SHAP influences, the optimal number of clustering is set as 7, based on the silhouette score in Table 4. Table 5 shows rules defined by SkopeRules for clusters based on influences. These rules have a median of two attributes per rule and focus mainly on statistically relevant attributes. Only one rule consists of three attributes, and the most used attribute is Urine Pushing, with five occurrences. As shown before for the "Nausea subgroups", this attribute is the most important for patients with Nausea (clusters 4 and 6) and also for patients with lumbar pain (clusters 3 and 5). Urine Pushing does not appear in rules only for clusters 2 and 7, the two biggest clusters, where AIUB-risk is respectively very low and very high. These clusters may be interesting to study from a medical point of view to understand patients characteristics and why the Urine Pushing variable is not the most relevant variable to distinguish them from other clusters. Also, although Micturition pain is the most influential attribute for SHAP, it is not very present in the rules, mainly because this attribute seems replaced by the attribute Nausea in the clusters since there is a strong link between having Nausea and Micturition pain.

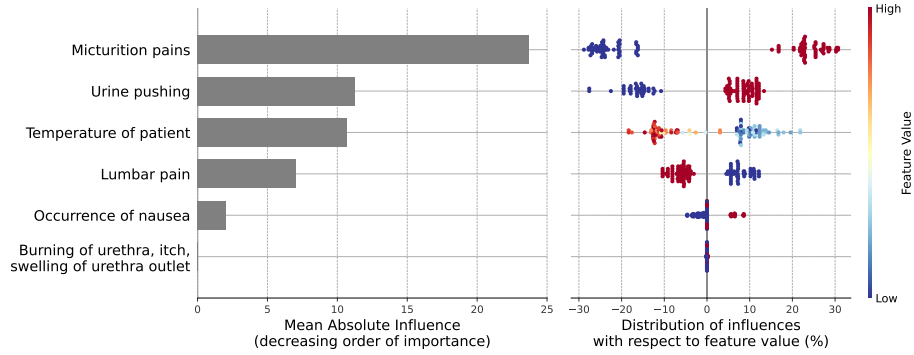


Fig. 1: SHAP mean absolute influences and Distribution of influences for the trained modelling.

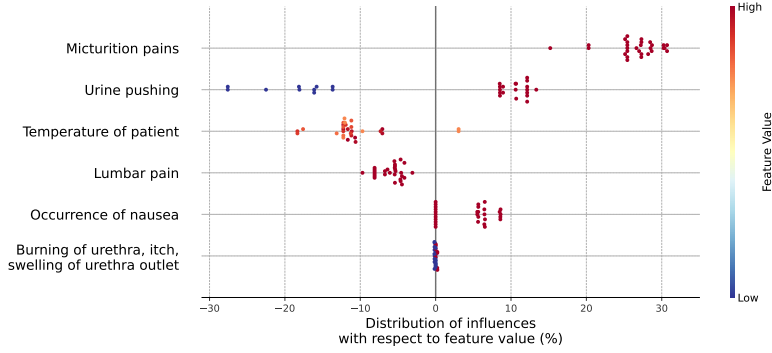


Fig. 2: Distribution of SHAP influences for patients with nausea.

Table 4: Silhouette Score for multiple numbers of clusters for XAI data.

K	2	3	4	5	6	7	8	9	10	11	12	13	14	15
XAI	0.59	0.59	0.52	0.62	0.69	0.76	0.74	0.69	0.63	0.61	0.61	0.61	0.62	0.67

Table 5: Decision Rules for clusters based on influences, with the number of patient per cluster and the mean percentage of AIUB-risk.

Rules	Nb	Mean %
1 Temperature ≤ 38.89 & Urine pushing = 0	20	11.0
2 Micturition pain = 0 & Urethra burning = 1	21	12.0
3 Lumbar pain = 0 & Urine pushing = 0	10	10.7
4 Nausea = 1 & Urine pushing = 0	10	45.6
5 Lumbar pain = 0 & Urine pushing = 1 & Micturition pain = 0	10	59.9
6 Nausea = 1 & Urine pushing = 1	19	73.0
7 Lumbar pain = 0 & Micturition pain = 1	30	97.2

4 Discussion & Perspectives

In this study, both raw data and explainability methods detect patterns in the data, subgroups of patients and information about the relationship between the AIUB disease and patients' symptoms. In addition to the information known in the literature [3] and found in the raw data analysis, the explanation-based data analysis allowed risk and protective factors to be identified more concisely. Rules are mainly based on statistically significant attributes, adding interactions between attributes, and with the target class, compared to raw data analysis. The smaller number of clusters and attributes in each rule also simplifies the understanding of patient subgroups and the relationship of each attribute to the AIUB risk. With raw data, multiple clusters have similar mean percentages of AIUB risk and almost identical patients. The differences between these clusters are often based on attributes not important for detecting AIUB. This behaviour can be beneficial to study the dataset in-depth, less for discovering the attributes that truly impact the diagnosis of the disease and for capturing concise knowledge. The conciseness provided by influences also makes it easier to assign a new patient to a subgroup of patients to study their disease and risk factors. This advantage comes from the ability of ML modelling to capture more complex relationships than traditional statistical methods. Finally, the explanation data allowed the discovery of relevant subgroups of patients, including those with nausea. This subgroup has strong relationships between several attributes, and the presence of AIUB is based solely on the attribute Urine Pushing, making its study interesting for understanding the mechanisms of the disease in some patients. Finding this type of subgroup can help to investigate biases in the dataset, especially around the attribute Nausea.

However, the proposed method should be applied and tested in more complex medical contexts, with datasets having different characteristics, such as more observations, more attributes, more variability leading to lower model performances. Therefore, this is the principal axis of future work to identify the main improvement points so that the proposed method can be tested with practitioners in the loop and fully and reliably adopted by them.

Acknowledgement We gratefully acknowledge the French ANRT and Kaduceo company for providing us with PhD grants (no. 2020/0964), and the Programme d'Investissements d'Avenir and the French ANR for the support of the national infrastructure ECELLFrance (PIA-ANR-11-INBS-005) and the PhD grant EUR CARe N°ANR-18-EURE-0003.

References

1. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794 (2016). <https://doi.org/10.1145/2939672.2939785>
2. Cooper, A., Doyle, O., Bourke, A.: Supervised clustering for subgroup discovery: An application to covid-19 symptomatology. In: ECML-PKDD Proceedings (2021)

3. Czerniak, J., Zarzycki, H.: Application of rough sets in the presumptive diagnosis of urinary system diseases. In: *AI and Security in Computing Systems* (2003)
4. Doumard, E., Aligon, J., Escriva, E., Excoffier, J.B., Monsarrat, P., Soulé-Dupuy, C.: A quantitative approach for the comparison of additive local explanation methods. *Information Systems* 114 (2023). <https://doi.org/10.1016/j.is.2022.102162>
5. Excoffier, J.B., Escriva, E., Aligon, J., Ortala, M.: Local Explanation-Based Method for Healthcare Risk Stratification. In: *Medical Informatics Europe 2022. Studies in Health Technology and Informatics* (2022)
6. Excoffier, J.B., Salaün-Penquer, N., Ortala, M., Raphaël-Rousseau, M., Chouaid, C., Jung, C.: Analysis of covid-19 in patients in France during first lockdown of 2020 using explainability methods. *Medical & Biological Engineering & Computing* 60 (2022). <https://doi.org/10.1007/s11517-022-02540-0>
7. Ferretini, G., Escriva, E., Aligon, J., Excoffier, J.B., Soulé-Dupuy, C.: Coalitional strategies for efficient individual prediction explanation. *Information Systems Frontiers* (2021). <https://doi.org/10.1007/s10796-021-10141-9>
8. Gardin, F., Gautiern, R., Goix, N., Ndiaye, B., Schertzer, J.M.: *Skope-rules* (2019), <https://github.com/scikit-learn-contrib/skope-rules>
9. Hoofnagle, C.J., van der Sloot, B., Borgesius, F.Z.: The european union general data protection regulation: what it is and what it means. *Information & Communications Technology Law* 28 (2019)
10. Lee, K., Ayyasamy, M.V., Ji, Y., Balachandran, P.V.: A comparison of explainable artificial intelligence methods in the phase classification of multi-principal element alloys. *Scientific Reports* 12 (2022)
11. Liu, Y., Liu, Z., Luo, X., Zhao, H.: Diagnosis of parkinson’s disease based on shap value feature selection. *Biocybernetics and Biomedical Engineering* 42 (2022)
12. Lundberg, S.M., Erion, G.G., Lee, S.I.: Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888* (2018)
13. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: *NeurIPS Proceedings* (2017)
14. Monsarrat, P., Bernard, D., Marty, M., Cecchin-Albertoni, C., Doumard, E., Gez, L., Aligon, J., Vergnes, J.N., Casteilla, L., Kemoun, P.: Systemic periodontal risk score using an innovative machine learning strategy: An observational study. *Journal of Personalized Medicine* 12 (2022). <https://doi.org/10.3390/jpm12020217>
15. Morgenthaler, S.: Exploratory data analysis. *WIREs Comp Stats* 1 (2009)
16. Ribeiro, M., Singh, S., Guestrin, C.: “why should i trust you?”: Explaining the predictions of any classifier. In: *KDD Proceedings* (2016)
17. Wang, H., Doumard, E., Soulé-Dupuy, C., Kémoun, P., Aligon, J., Monsarrat, P.: Explanations as a new metric for feature selection: a systematic approach. *IEEE Journal of Biomedical and Health Informatics* (2023)
18. Wirsch, A.: Analysis of a top-down bottom-up data analysis framework and software architecture design. Ph.D. thesis, MIT (USA) (2014)