Multilevel Clustering Explainer: An Explainable Approach to Electronic Health Records

José M. Clementino Jr*, Bruno S. Faiçal*, Christian C. Bones*,
Caetano Traina Jr.*, Marco A. Gutierrez† and Agma J. M. Traina*
*Institute of Mathematics and Computer Sciences, University of São Paulo (USP)- São Carlos, SP,BR 13566-590
†Heart Institute Clinical Hospital, Faculty of Medicine, University of São Paulo (HCFMUSP)- São Paulo, SP,BR 05403-900
Email: juniorclementino@usp.br

2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS) | 978-1-6654-4121-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/CBMS52027.2021.00073

***Abstract*—Machine learning (ML) algorithms have been used in many areas of activity, and their results can often be applied without further human intervention. The ML algorithms have also been widely used in medical contexts, but in this area, the result needs to be thoroughly confirmed by a specialist, who needs explanatory information on how the results were obtained. Aimed at such scenarios, we propose the Multilevel Clustering Explainer (MCE), a method capable of providing explanatory information to health professionals about the knowledge discovery process. The MCE was developed for the analysis of medical data, providing a synthesis of explanatory information for the specialist to quickly and clearly understand how the results were obtained.**

***Index Terms*—Explainable Artificial Intelligence, XAI, Electronic Health Records**

I. Introduction

Machine Learning (ML) algorithms are popular in several areas, helping the decision-making. In the medical context, they can assist, for example, in making early and accurate diagnoses [1, 2]. On the other hand, it is common that professionals with specific knowledge about machine learning concepts are required to best use those tools. Often, this knowledge is not within the physician's competence, leading to a loss of confidence in the result and doubts about the tool's suggestions, weakening the specialist's belief about a possible diagnosis. [3].

Researchers in the *Explainable Artificial Intelligence* (XAI) area have put great efforts to improve processing and responses explanability related to Machine Learning algorithms [4]. However, there is a gap in studies on the explanability of ML algorithms in the medical context using unsupervised tasks.

This work proposes a new method of explanability for unsupervised collaborative tasks in the knowledge discovery on Electronic Health Records (EHR). This novel method allows a multilevel analysis of details (Local and Global) directed to the medical context, allowing users to personalize the information of interest and the way it is presented. Henceforth, we refer to the EHR as all the patient's hospitalization records. A patient can have more than one EHR due to multiple hospitalizations.

The authors would like to thank Coordination for the Improvement of Higher Education Personnel (CAPES), grant PROEX-11357281/M; the Sao Paulo Research Foundation (FAPESP), grants 2016/17078-0, 2018/06228-7, 2019/04660-1, 2018/06074-0, 2020/07200-9; and the National Council for Scientific and Technological Development (CNPq).

Table I compares our method with related works. The well-known *LIME* approach [5] provides only local explanations. The *SHAP* [6] method provides local and global explanations, but it is not suitable for sparse data and as *LIME*, *SHAP* works only with supervised tasks. Automatic detection of Parkinson's disease is proposed in [7] based on genetic programming, but it does not present visual results. An explanatory model for the prevention of hospital respiratory failure in patients diagnosed with Coronavirus (COVID-19) is present in[8], however, it uses *SHAP* exclusively for explanations. It is possible to see that there is a lack of explainable methods for specific contexts (such as in health) in order to provide brief, clear and useful information. We did not find other explainable method in the scientific literature for unsupervised tasks in the EHR's context. To fill this gap, we propose a new method that satisfies the aforementioned requirements and provide multilevel explanatory information at Local and Global levels.

The remaining of this paper is organized as follows. Section II shows the details of the proposed method. Section III describes the experiments performed and the data used. Following, section IV discusses the results. Finally, section V presents our conclusions and future work.

II. The Multilevel Clustering Explainer (MCE)

We propose the *Multilevel Clustering Explainer* (MCE), which has three main phases. Phase: (1) perform the data pre-processing; (2) build the explainable structure at local and global levels; and (3) performs the data visual mapping and build the results presentation, as can be seen in Figure 1.

The MCE method receives as input the data already clustered by an extgernal algorithm and the parameters chosen by the user. This allows the use of agnostic clustering algorithms.

A. Phase 1 – Data pre-processings

The input is the dataset already clustered. Pre-processing involves to ajust the data formating different inputs, adding object identifiers (OIds) to the data elements that do not already haveit, as well as identifying the cluster where the each element belongs.

A new column is created in the dataset to identify the cluster where each element belogs. It will store the identifier IDs to link each object to its respective EHR, allowing retrieving additional information for future display. A second column

2372-9198/21/$31.00 ©2021 IEEE
DOI 10.1109/CBMS52027.2021.00073

TABLE I: Comparisons of MCE to existing methods.

Methods	Local Explanation	Global Explanation	General or Specific Context	Multilevel Explanation	Proposed for Unsupervised Tasks
LIME [5]	Yes	No	General	No	No
Unamed [6]	Yes	Yes	General	No	No
GE [7]	Yes	No	General	No	No
SHAP [8]	Yes	Yes	General	No	No
EXPLAIN-IT [9]	Yes	No	General	No	Yes
MCE (proposal)	**Yes**	**Yes**	**Specific**	**Yes**	**Yes**

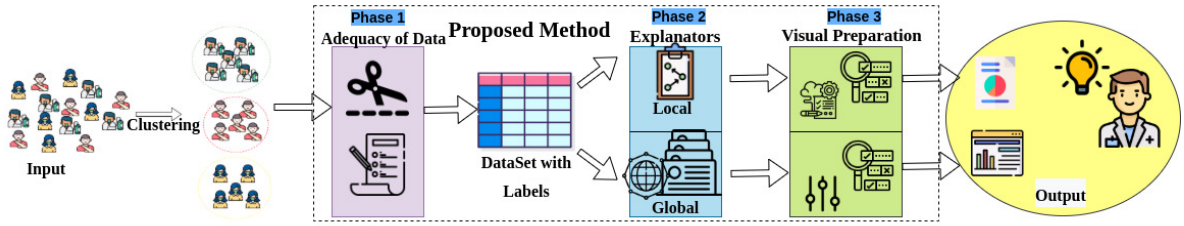


Fig. 1: Multilevel Clustering Explainer (MCE) workflow. Phase 1: Clean data. Phase 2: Construct explanatory information (with different levels of detail, as Local or Global). Phase 3: information presentation

is insert to store the cluster Ids (labels) where each object was inserted.

B. Building the Explainable architecture - Phase 2

Phase 2 constructs the explanatory information in two levels: Local and Global. At the local level, explanatory information is built referring to a specific EHR. At the global level, the explanatory information refers to the most influential attributes to form a group of EHR considered similar during the grouping process.

1) Local: The explanatory information describes the attributes that most influenced the decision to insert an object in a group at this level. For this work, the term "attribute" refers to the medical procedures recorded in the EHR.

Figure 2 exemplifies the calculation of the attribute influences for the object indicated by the specialist. Indications are made specifying the line number of the object in the dataset or indicating the EHR's ID (Figure 2-Input).

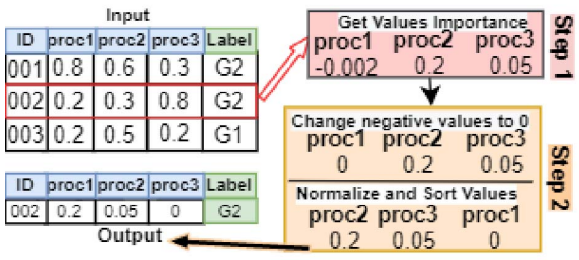


Fig. 2: Workflow for building the back-end for the explanatory information at the Local level - Phase 2

The attribute values of each object are retrieved from the dataset to calculate the influences of each attribute. MCE proposes to be flexible with the process of calculating influences, which can be carried out with a specific approach from the specialist or from existing methods in the scientific literature (Figure 2-Step 1).

The attributes influences are calculated, generating a new dataset with the exchanged samples and the respective group Id where the object was inserted during the grouping process. In this new dataset, an interpretable model (for example, a decision tree) is trained and employed to assess learning, which must perform an accurate classification of the sampled data . The learned function is used to map the most significant attributes for the prediction, estimating the respective influences. Thus, the learning of the interpretable model allows to infer the magnitude of the influence that each attribute presents to classify each object in the appropriate group. For example, an attribute will have its influence set to 0 when it is not used in the interpretable model. In contrast, an attribute will have a large magnitude of its influence when it indicates the correct classification of the object under a specific condition, typically greater than a threshold.

After calculating the influences, the values are adjusted to be objective and succinct for the medical context. It is necessary because calculating the influence can generate large amounts of information, making its application to practical situations both time-consuming and bearing unnecessary complexity. Negative values spot attributes that penalize placing the object in the group where it was inserted. MCE directs the information built for explanability to a direct approach, which favors its practical application. Therefore, negative values are converted to zero, which corresponds to attributes that did not influence inserting the object in that group (Figure 2-Step 2). This conversion avoids the large variability of negative values and fosters the explicability of why the object was inserted in that group.

Step 2 normalizes and sorts the values in descending order. Positive values are normalized, thus attributes with positive

influence clearly express its magnitude. A descending ordering is performed to pass the values for Phase-3 (information presentation). Algorithm 1 creates the explainable information for Phase 3 at the Local level.

Algorithm 1 Local Explainer

Input: Id_Object, DS_Clust
Output: Inf_Local
1: $Object \leftarrow Recovery(Id\_Object, DS\_Clust)$
2: $Inf\_Local \leftarrow Met\_Leverage(Object)$
3: $Inf\_Local \leftarrow Met\_RemoveNegative(Inf\_Local)$
4: $Inf\_Local \leftarrow Met\_Normalization(Inf\_Local)$
5: $Inf\_Local \leftarrow Met\_Organize(Inf\_Local)$

Initially, the object indicated by the specialist is retrieved from the data set generated by Phase 1, as shown in algorithm 1 line 1. Line 2 calls the method that calculates the influence of the object's attributes. In line 3 and 4, negative influence values are replaced by 0 and normalized. Finally, the attributes are sorted in descending order and the process finishes.

2) Global: The main difference among the Local and Global levels is the granularity of the information estimated. At the Global level, every object in the group defined by the specialist is processed, which generates an overview of how the analyzed cluster was built. At the Local level, the information refers to the insert operation of each record in a single cluster.

Figure 3 shows the process for calculating the most influential attributes for building the respective group. Initially, the specialist informs the group Id requiring explainable information, and all the elements that make up the respective group are retrieved from the data set generated in Phase 1.

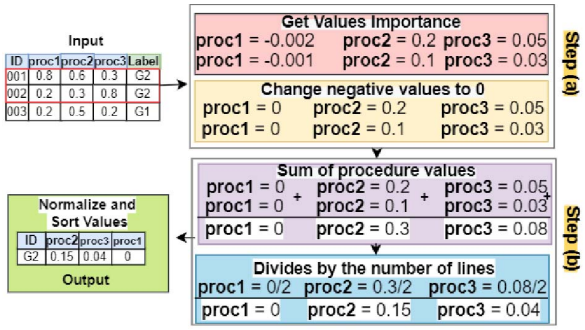


Fig. 3: Workflow to build global explanatory information

Thereafter, the attribute values of all objects are calculated and the negative values are zeroed, as show in Figure 3-Step a. Its execution is the same as for the Local level.

The impact of the individual influences are calculated at the Global level as shown in Figure 3-Step b. The analyst can choose which expression must be used by indicating the initialization parameter. Several expressions were implemented. Here we detailed the calculation using the arithmetic average as it is easy to understand. Moreover, it has shown promising results to indicate the most influential attributes for the grouping process.

The average calculated for each attribute considering the objects that belong to the group under analysis is presented in equation 1, where m_j is the total number of objects in group G_j, and A_{ij} is the value of the attribute in each object belonging to group j. The final result is a vector with the value average of the influences of each attribute.

$$Avg(G_j) = \frac{1}{m_j} \sum_{i=1}^{m_j} A_{ij} \quad (1)$$

Finally, the average influence values of the attributes are normalized based on the objects of the respective group. These calculations are also similar to those performed for the Local level.

Algorithm 2 describes how the influences of the most influential attributes are calculated to create the respective group at Global level. In line 1, the objects belonging to the group indicated by the specialist are retrieved from the data set generated by Phase 1. In lines 2-6, the influence of each object's attributes in the group is calculated, the zeroing negative values. DS_Tem is a temporary dataset with the non-negative influences values of each attribute of all objects. Lines 7-10, is calculate the arithmetic average for each attribute in the temporary data set (DS_Tem) , resulting in the array Inf_Globa with the average values of the influences of the objects in the group. Finally, the array is normalized and sorted in a decreasing order, as for the Local level.

Algorithm 2 Global Explainer

Input: $Id\_Cluster$, DS_Clust
Output: Inf_Global
1: $DS\_Objects \leftarrow Recovery(Id\_Cluster, DS\_Clust)$
2: **for each** $Object \in DS\_Objects$ **do**
3: $Object \leftarrow Met\_Leverage(Object)$
4: $Object \leftarrow Met\_RemoveNegative(Object)$
5: $DS\_Temp.insert(Object)$
6: **end for**
7: **for each** $Attribute \in DS\_Temp$ **do**
8: $Aver\_Temp \leftarrow Met\_Average(Attribute)$
9: $Inf\_Global.insert(Aver\_Temp)$
10: **end for**
11: $Inf\_Global \leftarrow Met\_Normalization\_Organize(Inf\_Global)$

C. Vizual Presentation - Phase 3

Phase 3 needs to insert complementary information to the admission record, including: the Id of the patient admission record and the Diagnosis admission at Local level, and the total amount of EHR for the investigated group at Global level. This phase organizes the explainable information generated in Phase 2 in a chart, for visual purpose. Visually summarizing the information provides the specialist with a clear and quick understanding of the information needed to support decision making processes.

It is possible to customize the visual presentation in the following characteristics: (i) type of graph, (ii) number of attributes to be displayed (attributes of interest), (iii) absolute or percentage values, (iv) whether the sum of the values not belonging to the attributes of interest should be displayed. The customization parameters supported are the same for both

levels. An example of the MCE visual presentations can be seen in Figure 5 for the Local level and in Figure 4 for the Global level. They are detailed in section IV.

Algorithm 3 presents the details of Phase 3 execution. As explained earlier, this phase uses the Phase 2 output according to the request level. In lines 1 - 4 the data for a specific EHR are retrieved by ID_Visit. Proper execution at the level defined by the $Recovery\_Data$ method is defined by the $Level$ parameter, which can indicate *Local* or *Global*. The $Plot$ method builds the graph using the Local level explainability information (Inf_Local), the recovered data (DS_Visu), the configuration parameters indicated by the parameter vector Vec_Param and the execution is driven by the $Level$ parameter. Lines 5 - 9 process the data recovered from the EHR from the group identified by the Id_Clust, always guided by the $Level$ parameter. The $Plot$ method builds the graph using the explainability information for the Global level (Inf_Global), the recovered data (DS_Visu) and the configuration parameters indicated by the parameter vector Vec_Param. The execution is again driven by the $Level$ parameter. Finally, in line 10, the $Save$ method generates a file with the graphics and the complementary information retrieved.

Algorithm 3 Build the visual presentation

Input: $Inf\_Local, Inf\_Global, Id\_Visit$
Input: $Id\_Clust, Level, EHR, Vec\_Param$
1: **if** $Level = Local$ **then**
2: $DS\_Visu \leftarrow Recovery\_Data(Id\_Visit, Level, EHR)$
3: $GF\_Visu \leftarrow Plot(Inf\_Local, DS\_Visu, Level, Vec\_Param)$
4: **else**
5: **if** $Level = Global$ **then**
6: $DS\_Visu \leftarrow Recovery\_Data(Id\_Clust, Level, EHR)$
7: $GF\_Visu \leftarrow Plot(Inf\_Global, DS\_Visu, LevelVec\_Param)$
8: **end if**
9: **end if**
10: $Save(GF\_Visu)$

III. Experiments

A. Materials

Our tool uses EHRs represented in the OMOP data. We use data provided by the *Heart Institute* (InCor) of the University of São Paulo, which contains real, anonymized data approved by its ethics committee. The grouping of the hospitalized patients was performed using the vector representation *Bag-of-Attributes* and the *SK-Means* algorithm [10].

B. Setup

The experiments were organized in two parts, regarding Local and Global views.The results evaluation was carried out in two parts and it seeks to identify relationships between the information displayed at the Local and Global levels. The relationships are understood as similarities between the attributes indicated as the most influential and predominant in diagnoses for that group (understood as a standard).

The data has 10,291 EHRs, which makes it difficult to visualize the descriptive information for each group. Therefore, we selected the EHRs from the four larges groups built by the *SK-Means* algorithm. The selected groups have the ids 7, 5, 9 and 45, which $2,185$, $1,184$, 356 and 330 objects respectively. We focus our experiments by the number of objects in the groups because we believe that larger groups have more expressive patterns, although their larger data volume also makes them are more difficult to analyze.

We emphasize that qualitative approaches are widely used in the scientific literature [5, 6], because a method of explainability may be more suitable for a given area than for others. We emphasize this premise as a motivation for the our work because there are no explanatory methods in the scientific literature aimed at the medical context. Therefore, the results are analyzed qualitatively, with the support of health professionals.

IV. Results and discussion

In this section, the results of the proposed method are described and detailed. The results are presented at their respective levels (Local and Global) for a clearer visualization. However, they can be retrieved between subsections and analyzed by observing the complementary characteristics between levels.

It is important to clarify that some images may not be clearly legible due to the size necessary for a good visualization. The results for all groups (at the Global level) and the records used for this analysis (at the Local level) are available in a public repository[1].

A. Global Level

At the global level, the MCE presents general explanatory information for one group and the most influential attributes for objects to be inserted in that group. Figure 4 demonstrates the information received by the specialist at this level of explainability.

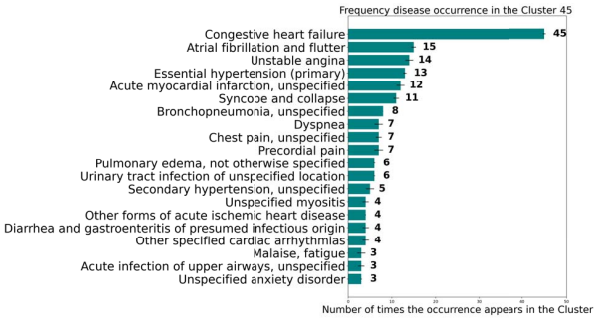
Figure 4(a) shows the group Id and the number of EHRs records that compose the group. This experiment shows results for the Cluster Id 45 with 245 EHRs inserted. Figure 4(b) shows new horizontal bar graph is presented with the frequency of diseases occurrences (attributes) in the analyzed records. This group's predominant disease occurrence is "Congestive heart failure" with 45 occurrences from the 246 records analyzed. There may be more than one disease occurrence for each EHR in the data, but the repetition of some occurrence in a EHR is not considered consistent.

In Figure 4(c), a horizontal bar graph shows the frequencies procedures performed in the Cluster 45. The predominant procedures of this group are "Creatinine/sorology" with 246 tests performed and "Urea/sorology" with 241 tests. With these data, it is possible to identify the procedures commonly requested for patients presenting similar complaints. An important feature in this data is the independence between the number of EHRs and how many times the procedures were performed. The procedures can be performed more than once during a hospitalization, and it is also possible to exist EHRs where some procedures have not been performed. The

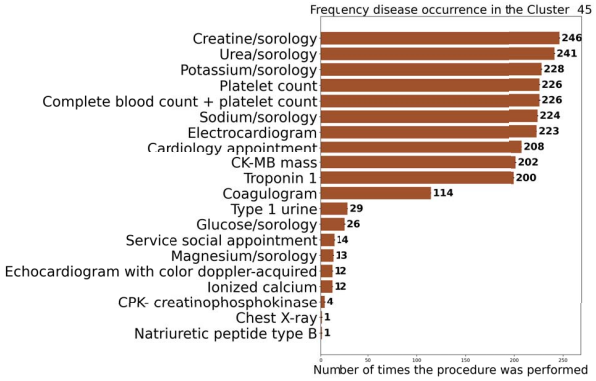
[1]https://github.com/clementinojr/Multilevel-Clustering-Explainer-an-explainable-Approach-to-EHR

Cluster Label	Quantity of EHR
45	245

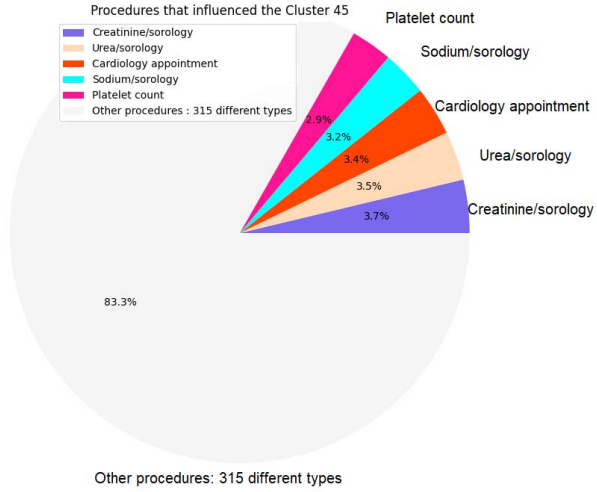
(a)



(b)



(c)



(d)

Fig. 4: Explanatory information at the Global level

case of procedures performed more than once in a EHR can be seen when comparing the number of EHRs (245) and how many times the procedure "Creatinine/sorology" was performed (246). On the other hand, the case where procedures were not performed in all EHRs can be noted in procedures that have a frequency lower than the number of EHRs. As an example, it is possible to mention procedure "Urea/sorology", which was performed 241 times in 245 EHRs. Finally, it is important to note that when there is a procedure with the same number of EHRs, it is not a proof that this procedure has been performed in all EHRs as it may exist repetitions in some EHRs but not in others.

Figure 4(d) presents the procedures (attributes) considered most influential for a record to be inserted in the group with Id 45. Thus, the five most influential procedures for inserting a record in the group 45 were "Creatinine/sorology" with 3.7%, "Urea/sorology" with 3.5%, "Cardiology appointment" with 3.4%, "Sodium/sorology" with 3.2% and "Platelet count" with 2.9%. Also, it is possible to note that another 315 procedures were performed on the 245 EHRs and share the complementary 83% of influence. As we consider many values with low influence, they are presented in the graph in a condensed way under the description "Other procedures".

B. Local Level

We selected the explanatory information for a random record inserted in the Cluster Id 45. Although the admission record was selected at random, the group was chosen in a specific way to allow an analysis linked to the Global level. This approach allows a *top-down* view of explanatory information.

The explanatory information is show in Figure 5. Figure 5(a) is where some descriptive information for the record is located, such as the group Id to which the record was assigned to, the EHR Id and the name of the disease occurrence. This information shows the specialist that the record in Figure 5 belongs to group 45, the ID of this EHR record is 8903852871656482.0 and the name of the occurrence is "Secondary hypertension, unspecified" . Crossing this information with the information of the Global level (Figure 4), it is possible to notice that this record is one of the five existing records in the group with Id 45 that are linked to the occurrence "Secondary hypertension, unspecified".

Figure 5(b) shows the influence rate of the five most influential procedures (attributes). The most influential procedures in decreasing order of the respective levels are "Creatinine/sorology" with 0.10037,"Cardiology appointment" with 0.09870, "Urea/sorology" with 0.09737, "Sodium/sorology" with 0.08826 and "Complete blood count + platelet count" with 0.07476.

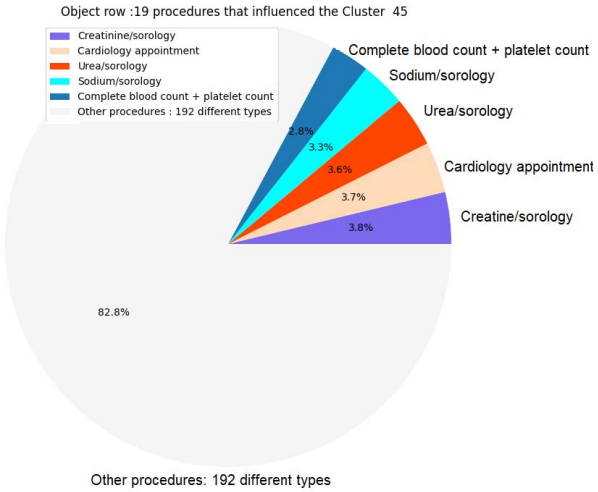
The Four procedures indicated as the most influential for the insertion of this record (local level) in a group are also the most influential procedures for building the group 45 (global level). Only the fifth most influential procedure at both levels is the same. In addition, the second and third position procedures at both levels occur in reversed positions in the opposite level. Through the visualization of the similarity between the procedures with the most significant influence at both levels, it is possible to better understand why the record with Id 8903852871656482, 0 was inserted in the group 45, as it shows the variables that most influenced the grouping. In addition, this explanatory information exposes the coherence

Cluster Label	EHR ID	Occurence Name
45	8903852871656482.0	Secundary hypertension, unspecified

(a)

Attribute Name	Value
Other procedures: 192 different types	2.2132235019222573
Creatinine/sorology	0.10037760766764824
Cardiology appointment	0.0987036308436467
Urea/sorology	0.09737369995322803
Sodium/sorology	0.08826610759338882
Complete blood count + platelet count	0.07476231160293385

(b)



(c)

Fig. 5: Local level output for the MCE method.

between the two levels (Local and Global) for understanding the clustering process. Finally, another 192 procedures share an influence rate with a value of 2.21322 condensed to the term "Other procedures". We emphasize that the individual influence values are lower than those presented as more influential.

Figure 5(c) presents a pie chart with percentage values referring to the procedures' influence. It also shows the summation of the procedures influences not shown as the most influential under the term "Other procedures".

V. Conclusion and Future Work

This paper presents the Multilevel Clustering Explainer (MCE) method, which was designed to build and presenting explanatory information on the pattern detection process using patients's EHR. The proposed method's objective is to gather and present explanatory multilevel information (local and global) regarding the process carried out in the detection of patterns (such as grouping) to support the decision-making of health specialists.

The results showed that MCE can gather explanatory information about the data used to detect patterns as well as to organize and describe information to complement the explanation in the medical context. The proposed method also allows the specialist to customize the way results are presented with the insertion of parameters other than the default ones. Finally, the method proved to be robust in constructing explanatory information and allowed a *top-down* analysis to foster the understanding of possible cases of inconsistencies when analyzed by the specialists.

References

[1] Y. Xie, X. A. Chen, and G. Gao, "Outlining the design space of explainable intelligent systems for medical diagnosis," in *Joint Proceedings of the ACM IUI 2019, Los Angeles, USA, March 20*, 2019.

[2] A. Barredo Arrieta *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020. [Online]. Available: https://doi.org/10.1016/j.inffus.2019.12.012

[3] R. Guidotti *et al.*, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 93:1–93:42, Aug. 2018. [Online]. Available: http://doi.acm.org/10.1145/3236009

[4] M. Ribera and À. Lapedriza, "Can we do better explanations? a proposal of user-centered explainable ai." in *IUI Workshops*, 2019.

[5] 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, San Francisco, CA, USA, August 13-17, 2016*, 2016, pp. 1135–1144.

[6] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 4768–4777.

[7] F. Cavaliere *et al.*, "Parkinson's disease diagnosis: Towards grammar-based explainable artificial intelligence," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, 2020, pp. 1–6.

[8] A. D. Haimovich *et al.*, "Development and validation of the quick covid-19 severity index: a prognostic tool for early clinical decompensation," *Annals of emergency medicine*, vol. 76, no. 4, pp. 442–453, 2020.

[9] A. Morichetta, P. Casas, and M. Mellia, "Explain-it: Towards explainable ai for unsupervised network traffic analysis." New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3359992.3366639

[10] J. M. Clementino *et al.*, "Bag-of-attributes representation: A vector space model for electronic health records analysis in omop," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 2020, pp. 197–202.