

Hierarchical Method for Classifying Latent Traumatic States (CAH-ET)

Kpinna Tiekoura Coulibaly¹, Abdoul Maïga², Kamagaté Beman Hamidja¹, Diaby Moustapha¹

¹LASTIC, African Higher School of Information and Communication Technologies (ESATIC), Abidjan, Ivory Coast

²UFRMI, Felix Houphouët Boigny University (UFHB), Abidjan, Ivory Coast

Email: tiekoura77@yahoo.fr

How to cite this paper: Coulibaly, K.T., Maïga, A., Hamidja, K.B. and Moustapha, D. (2024) Hierarchical Method for Classifying Latent Traumatic States (CAH-ET). *Open Journal of Applied Sciences*, 14, 3500-3515. <https://doi.org/10.4236/ojapps.2024.1412229>

Received: November 11, 2024

Accepted: December 8, 2024

Published: December 11, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This article presents a hybrid method of automatic classification of latent traumatic states adapted to the analysis of social resilience processes. Our approach combines the Hierarchical Ascending Classification (CAH) technique with decision tree. It is primarily aimed at improving the identification and categorization of traumatic states by integrating the strengths of both methods. CAH is used to cluster data, allowing the detection of underlying patterns in traumatic states. Then, decision trees are applied to classify these clusters, providing a clear and accessible interpretation of the results. This combination not only provides a better understanding of the data structure, but also provides accurate and actionable classifications. This work highlights the importance of this hybrid method in the field of social resilience processes, particularly in psychology and psychiatry, where early detection and classification of trauma can have a significant impact on the patient's follow-up. Experimental results show an improvement in the classification accuracy of our approach compared to a classification method for the same domain using a hybridization between the partitioning technique and genetic algorithms. This opens promising prospects for the application of this approach in clinical settings of social resilience.

Keywords

Social Resilience, Classification, Machine Learning

1. Introduction

The current global environment with its corollary of successive crises of all kinds (military-political, socio-economic crisis, natural disasters, terrorism, etc.) is becoming increasingly traumatic for the humans we are. According to the World

Health Organization (WHO) [1], between 2004 and 2030, the relative share of mortality due to trauma will increase in the world (for example, transportation accidents will go from 9th to 5th in the cause of death, suicides from 16th to 12th and interpersonal violence from 22nd to 16th). In addition, according to the same organization, intentional injury is the leading cause of death among people under 45 years of age and up to almost 40% of causes of death among youth under 15 years of age. Thus, in such an environment, the predominant question is not how to eliminate these crises, but rather how to overcome them for a better life. In other words, we are looking for ways to adapt to these recurring traumas and move forward in life. This concern raises the thorny question of research around the term “resilience” which has captivated many governments and organizations around the world. Resilience, beyond its many definitions, generally refers to the ability of an entity (individual, household, community, and so on) to bounce back or adapt in the face of a traumatic shock. In 2005, following the publication of the proceedings of the MIT symposium entitled “The resilient city: trauma, recovery and remembrance” and the adoption of the Hyogo Declaration [2] by countries, urging the various states and organizations to develop resilience policies for their populations, the concept of resilience, now enthroned as a new paradigm of risk science, will be highly publicized and will trigger several research projects to operationalize the concept.

Literature on resilience allows us to observe, besides a few operational approaches, a large cross-disciplinary panel of theoretical approaches often elusive due to the polysemic nature of the term. One of the most important aspects in the field of social resilience is the study of social stratifications or sociological typologies to help decision-making. However, to our knowledge, to date, existing typology methods are too general to consider the specificities of resilience and are difficult to use for non-specialists in modelling.

Indeed, the particularity of a classification model is the measure of similarity or dissimilarity used to establish typologies. In a polysemic field such as social resilience, similarity measures to find existing similarities between traumatized individuals must depend on several abstract qualitative variables, including adaptability, altruism, incoherence, logic, spirituality, coherence and so on. This is not the case with conventional classification methods, which do not take these specific features into account. Thus, the problem of developing classification methods adapted to the context of social resilience processes is acute.

In this work, we propose a hybrid model for classifying the typologies of individuals from social resilience processes to allow users to easily analyze social resilience processes and optimize policies aimed at increasing resilience of individuals. Our model is based on a hierarchical classification approach and decision tree technique and uses the SimCT semantic similarity measure [3] that we developed in previous work on social resilience.

In the remainder of this article, we first present a brief state-of-the-art of some classification methods. Next, we set out the problem and the objective of this

work, before presenting our contribution with an experiment on real data from a survey. Finally, we discuss the experimental results before concluding.

2. State of the Art of Some Automatic Classification Methods

Hierarchical classification methods and partitioning methods are the two main families of automatic classification, also called unsupervised classification. For several years, a third classification family combining these two approaches has been added. These are the mixed classification models. In general, automatic classification encompasses all statistical techniques for identifying groups, often referred to as classes, within a range of objects. Unlike the supervised classification (or discriminant analysis) which, from a set of already identified classes, allows to find the best class to which a given object belongs, Unsupervised classification aims to constitute unidentified classes that group the different objects.

There is a wide range of research in the literature on automatic classification methods (hierarchical and non-hierarchical). Can be found in the work of (Xiao, 2024) [4] and (Ntayagabiri *et al.*, 2023) [5].

In general, automatic classification methods are based on four (4) main techniques [6] namely: classification hierarchy, partitioning, grid and density. Depending on the classification technique discussed, five categories of collection of algorithms have developed [7].

In this work, we will be interested in classification methods based on hierarchy, of which some recent works are presented below. Two types of hierarchical classification are distinguished: the hierarchical ascending classification (CAH) or agglomerative and the hierarchical descending classification (CDH) or divisive. The bottom-up approach starts by treating each object as a separate class. At each step, the two closest classes are merged until a single class is obtained or until a termination condition is met. In contrast, the top-down approach starts with all objects grouped into a single class. At each iteration, a class is divided into smaller sub-classes until only one object per class remains, or until a stop condition is reached.

(Coulibaly *et al.*, 2017) [8] have developed a hybrid partitioning classification algorithm for identifying significant groups based on resilience levels. This generates from a traditional partitioning method of partitions which are then optimized using the genetic algorithms technique, to give the best possible partition: the one that minimizes most intra-class inertia and favors more distinct classes while eliminating single-element classes. The results of the simulations showed that the algorithm converged after 150 iterations to provide a solution that meets the expected objective. The Rand index (0.89) obtained, no doubt, reflects the good performance of the algorithm. However, this approach, although it has the merit of operationalizing the analysis of resilience processes, has some limitations, in particular the fact that results are highly dependent on the number of initial centers, and it presents a counter performance for small-size data.

(Bellanger *et al.* 2021) proposed a top-down compromise classification method called hclust-compro [9]. In their work, the authors consider two sources of in-

formation associated with the same objects. Thus, their method allows a compromise between the hierarchies obtained from each source taken separately. They also use a convex combination of the dissimilarities associated with each source to modify the dissimilarity measurement in the classical CAH algorithm. Their objective function is a minimization function based on the absolute difference of correlations between initial dissimilarities and cophenetic distances, as well as a resampling procedure to ensure robustness in the choice of the mixing parameter. One of the major drawbacks of this method, which is adapted to archaeological data, is that it can only consider two sources of information associated with the same objects.

(Hakami and L. Chang, 2023) provide a classification model to improve the classification of cardiac arrhythmias which are characterized by irregular heart rhythms and pose a considerable challenge to medical diagnosis because of their diverse and subtle manifestations [10]. This study proposes a method that combines sophisticated feature engineering and hierarchical classification. In order to ensure a solid classification, the study used the algorithms Random Forest and XGBoost. A two-tier hierarchical classification model has been put in place, starting with grouping heart rates into different classes, followed by a second model that sharpens the distinctions between specific classes. The results showed significant improvements in model performance through the addition of new features, which allowed us to increase accuracy and F1 scores. This approach allows to obtain interesting results with a certain speed, especially thanks to XGBoost, but it is nevertheless that it has some limitations including: the complexity to implement, Effectively enable and configure XGBoost which causes time loss; overfitting of training data that is likely to skew predictions of new data; the difficult interpretation and understanding of predictions obtained from XGBoost, all that can make it difficult to troubleshoot and adjust and finally, the fact that the algorithm is a little bit demanding in using memory for large data, Causing difficult execution on machines with limited capacity.

Finally, (Moiraghi *et al.* 2024) proposed a “zero-shot” hierarchical classification on the common vocabulary of public procurement [11]. This classification of public tenders is a useful task both for the companies invited to participate in and for the control of fraudulent activities. The authors propose this method to overcome difficulties of classification with respect to a real taxonomy, including the fact that some fine-grain classes have an insufficient number of observations throughout the formation, while other classes are much more frequent (or thousands of times) than the average. Their zero-shot approach is based on a pre-trained language model that relies solely on label description and respects the taxonomy of labels. The test results show that the proposed model performs better in classifying infrequent classes compared to three different baselines, and that it is also possible to predict classes never seen before. However, one of the disadvantages of this approach remains on the one hand, the challenge related to the physical infrastructure and associated costs due to the size of the re-training and inference from

the basic model. On the other hand, the modification of all layers of the model can cause harmful edge effects in the results. Added to this is the fact that poor data quality could lead to errors.

After the presentation of some recent classification methods in literature and their limitations, what is the interest of our present proposal on the hybridization of the hierarchical classification method and the decision tree technique, in the context of the analysis of social resilience processes? This question is answered in the following section.

3. Problem and Objective

An important aspect in operationalizing the concept of social resilience is the study of social stratifications or sociological typologies to aid decision-making. However, existing typology methods are often too general to consider the specificities of resilience and are difficult to use for non-modeling specialists. In addition, most traditional partition search methods have limitations, including their inability to effectively exploit the search space. How can we build an efficient stratification method adapted to social resilience processes? This is the problem that our present work addresses.

The aim of this paper is to respond to the great need for operational tools for analyzing social resilience processes, particularly for classifying and interpreting data.

4. Contribution

Our contribution is a classification algorithm adapted to resilience processes that is a hybrid between hierarchical ascending classification (HAC) and the decision tree technique. This choice combines the strengths of the CAH and decision trees. Thus, unlike partitioning algorithms and several traditional classification methods, our approach will allow a better understanding and analysis of social resilience process data. Thanks to the decision tree technique, our proposal is also resource-efficient and robust to outliers as well as noise in the data. Like hierarchical classification methods, it ensures a certain stability.

Indeed, this type of algorithm, unlike some partitioning algorithms such as k-means, does not require any initialization. Thus, run twice in a row, the algorithm will give the same result twice on the same dataset. Such stability is necessary for the analysis of data from a social resilience process given its sensitive nature. In the field of classification, however, one of the difficulties of any classification method is the choice of the number of classes. In most cases, this number is set arbitrarily at the beginning of the algorithm by the user. This arbitrary fixation is problematic, as it does not allow to identify the optimal number of classes. For CAH methods, dendrogram representation gives a global view of the topology of observations and allows the user to have an idea of the appropriate number of classes to use. Thus, in our algorithm the successive partitions are nested, and it is not necessary to specify a priori the number of classes that we want to have. This

choice can be made from the dendrogram. Anything that guarantees natural groupings and an optimal number of classes. Finally, because of its flexible nature, allowing the use of digital and categorical data, our approach is well suited to the analysis of social resilience processes.

However, one of the main disadvantages of CAH algorithms is the high cost in time to calculate distances at the initial stages on large data sets. In practice, some solve this problem by hybridizing the CAH algorithm with a partitioning method such as k-means, to reduce considerably the initial number of individuals and then perform the CAH on the obtained classes. This technique, for our part, causes other difficulties, since k-means does not allow to develop an optimal set of clusters and leads to a certain inconsistency by giving variable results on different executions of the algorithm. In addition, k-means is limited for non-numeric data.

Before presenting our hybrid classification algorithm, we will present in the following, modeling our approach by first defining the principle of classification of individuals in classes and then the technique of characterization by decision trees.

4.1. Modeling Our Approach

4.1.1. Basic Principles of Individual Classification

To overcome the fundamental problem of CAH, methods related to the temporal complexity of distance calculations, we initially use a representative sample of the total set of individuals. This reduces the size of the data and therefore the number of distance calculations between pairs of individuals at the initial stage.

In addition, for the aggregation of classes, to allow qualitative data to be considered, we use the notion of “close resilience boundary” rather than “geometric distance” as is the case in many methods, such as the CURE algorithm [12]. Thus, two classes are aggregated and if at iteration t , these two classes have the largest number of pairs of individuals with close resilience boundaries (*i.e.* the global resilience boundary).

The second step of our classification method concerns the characterization of the classes obtained. We use the decision tree technique here because of their ability to provide simple and readable visualization as well as an easy interpretation of results.

We define the concept of resilience boundary below.

Definition 1: The **elementary resilience threshold** (β_{ij}^t) of an individual (i) at time (t) relative to a given capacity (j) represents the minimum threshold below which the individual is vulnerable to this capacity.

Definition 2: The **overall resilience threshold** (β_i^t) of an individual (i) at time (t) represents the minimum threshold, below which the individual is vulnerable to all abilities. It is obtained by averaging the elementary resilience thresholds of all capacities.

$$\beta_i^t = \frac{\sum_{j=1}^{card(D)} \beta_{ij}^t}{card(D)} \quad (1)$$

Where $\text{card}(D)$ is the number of dimensions considered.

Note

$\Omega = \{x_1, x_2, \dots, x_n\}$, the set of individuals who suffered the traumatic shock to be classified.

$P^m = \{C_1, C_2, \dots, C_k\}$, the set of individual partitions at the iteration m .

C_i , the i^{th} class of individuals.

β_i^t , the resilience boundary of individual i , at time t ;

$|\Delta\beta_i^j|$, the positive difference in individual resilience thresholds x_i and x_j ;

θ , the threshold of difference.

$N_{Br}(C_i, C_j)$, the number of pairs of individuals belonging to classes C_i and C_j having close resilience boundaries.

Two individuals x_i and x_j have close resiliency thresholds, if the positive difference in their resilience thresholds is less than or equal to the threshold set θ .

Thus, for x_1, x_2 and x_3 , individuals belonging to different classes:

$$|\Delta\beta_1^3| \leq \theta \Rightarrow \beta_1 = \beta_3 \quad (2)$$

$$|\Delta\beta_1^2| > \theta \Rightarrow \beta_1 > \beta_2 \text{ or } \beta_1 < \beta_2 \quad (3)$$

The classes C_a and C_b are grouped, if they have the largest number of pairs of individuals with close resiliency boundaries. In this case, all individuals of the class C_b are added to the class C_a :

$$\text{If } N_{Br}(C_a, C_b) \leftarrow \underset{\forall C_i, C_j \in P^m}{\text{Max}} N_{Br}(C_i, C_j) \text{ then } C_a \leftarrow x_i, \forall x_i \in C_b \quad (4)$$

4.1.2. Variable Characterization Using the Decision Tree Technique

Decision trees [13] are a supervised learning method used for data exploration and prediction. The technique consists in iteratively constructing homogeneous classes of individuals by asking a succession of binary questions on the different attributes. It allows to predict the value of a qualitative variable from decryptable variables qualitative and/or quantitative. One of the main advantages of decision trees is their graphical representation. It provides a simple and readable visualization of anything that facilitates the interpretation of results. This graphical representation is in the form of a tree consisting of terminal sheets (the classes of individuals) and nodes corresponding to binary questions on a variable of the data set.

Tree construction process:

The effectiveness of a decision tree depends heavily on the purity of the leaves generated by the different nodes. The objective is to create for each node, two sheets that are more homogeneous than the preceding node. This requires that the questions chosen for the construction of the tree be as discriminating as possible. Two mathematical tools are generally used to test the purity of a leaf: the Gini index [14] and probabilistic entropy [15]. The first one is used by the CART algorithm while the C4.5 algorithm uses probabilistic entropy.

If we designate by X_i the i^{th} discrete variable describing an individual,

x_{i1}, \dots, x_{in} , the n modalities that can take this variable and p_{i1}, \dots, p_{in} , the respective probabilities of obtaining these modalities, the entropy $H(X_i)$ of the variable X_i is calculated in base b as follows :

$$H_b(X_i) = -\sum_{j=1}^n p_j \log_b(p_j) \quad (5)$$

Entropy, like the Gini index, is also described as a measure of disorder. More precisely, entropy is used to measure the disorder generated by the choice of the variable X_i as the cut-off variable of the node α and the x_{ik} modality as the criteria for the distribution of individuals in the different classes.

Entropy gives a value of zero when the variable X_i is deterministic and returns only one value x_0 and it is maximum (positive) when X_i is uniformly distributed, i.e., $\forall j, p_j = 1/n$.

In the process of constructing the tree, it happens that one ends up in a situation of over-learning or over-adjustment, characterized by an «exaggerated» description of the data set. This situation results in too many terminal sheets in the decision tree, which becomes de facto long and illegible. This is a dimensioning problem which can be solved by the so-called tree pruning technique. It allows to keep a correct level of generality by proceeding by cross validation. In other words, the dataset is divided into sub-sets, each serving as a learning sample and then as a validation sample.

Principle and stopping criteria:

Like all learning algorithms, the decision tree algorithm has the following stopping criteria:

- When the terminal leaves contain a minimum number of individuals so that further division would result in leaves smaller than the set threshold.
- Reaching a maximum number of leaves.
- When further division of the tree does not improve its discriminating quality.

The decision tree is constructed according to the following principle:

- a) Check whether the present level meets one of the stop criteria listed above.
- b) If yes, stop the process.
- c) If not,
 - Choose the question variable that maximizes information gain.
 - Partition the individuals according to the chosen question and create the corresponding node.
 - repeat steps a, b and c, starting from the new node.

4.2. Proposed Algorithm

The classification algorithm we propose here has the advantage of providing both a categorization of individuals victims of a traumatic shock and a characterization of the partitions generated using a decision tree. It is therefore a hybrid method between unsupervised and supervised learning.

Note by $\text{mad}(\cdot)$ the characterization of classes by the decision tree technique as used in the CART algorithm. The decision tree will be represented by \mathcal{A} .

The hierarchical algorithm for classification of latent traumatic states is as follows:

Hierarchical algorithm for classification of traumatic states (ClustResi)

Input: Ω , a set of traumatized individuals to be classified.

K , the number of classes desired.

θ , the threshold of difference.

Ω_{Desc} , the set of class descriptors

Output: $\Xi(P(C_l | l \in [1, k]), \Lambda)$, the final partition grouping the classes by level of trauma and the decision tree characterizing these classes.

BEGIN

// Procedure for obtaining the k fixed classes

1. Extract from the set of individuals (Ω), a representative sample (Γ_m) of size m
2. Calculate the similarity matrix of Γ_m
3. Calculate positive differences in resilience limits for pairs of individuals (x_i, x_j) and Γ_m
4. Subdivide Γ_m into h subsets of size m/h
5. Divide each sub-set into m/hq sub-classes (With $q > 1$)
6. Delete classes with very small numbers (1 or 2 elements)
7. Determine for each subclass a set of well-distributed points representative of the subclass.
8. Hierarchically merge subclasses with the highest number of nearby resilience bounds:

$$\text{If } N_{Br}(C_a, C_b) \leftarrow \text{Max}_{\forall C_i, C_j \in \mathcal{P}^m} N_{Br}(C_i, C_j) \text{ then } C_a \leftarrow x_i, \forall x_i \in C_b$$

9. Repeat step 6 until you have the desired k classes.

// Procedure for classifying the n individuals in the set Ω

10. Identify a representative for each class
11. Assign each individual to the class whose representative is closest to it
12. Repeat step 9. Until all individuals belong to a class
13. Return the set $P(C_l | l \in [1, k])$ of partitions.

// Class characterization procedure

14. $\Theta(\Lambda) \leftarrow \text{mad}(\Omega_{desc} | C_l)$

15. Return final partition and decision tree

$$\Xi(P(C_l | l \in [1, k]), \Lambda)$$

END

5. Experimentation

One of the thorny problems a researcher usually faces in applying theories is obtaining real and reliable data. This was our case. We were indeed confronted with a lack of data on social resilience processes. The limited data available does not meet the requirements of the models developed. In this work, we use data from

one of our research studies on social resilience (Coulibaly and Maïga, 2019) [16]. These data were obtained from a survey of a sample of 100 people on their resilience process in the face of the trauma caused to them by the 2010 military-political crisis in Côte d'Ivoire. Our present hybrid classification algorithm (CAH-ET) therefore uses the resilience scores of individuals that were calculated using our dynamic model for measuring social resilience presented in the said work. The variables characterizing individuals are made up of a number of indicators and three dimensions characteristic of social resilience defined in (Coulibaly *et al.*, 2015) [17]. These are: sense of coherence (COH), awareness (CONSC) and sense of humor (HUM).

For the implementation of our hybrid automatic classification algorithm, we used the R language. We conducted the experiment on a Surface Pro 9, 64-bit with an Intel(R) Core(TM) i7-1065G7 processor @ 1.30GHz 1.50 GHz, 16.0 GB RAM.

In the next section, we present the results of applying our algorithm to the resilience data of this sample of 100 people, in order to find the different typologies that emerge. These typologies will be particularly useful for decision-making.

6. Results and Discussion

6.1. Class Identification Step

We present here the application of our hybrid hierarchical classification model to the data collected in the survey. The execution of our algorithm allows to obtain the different significant classes in which are distributed the individuals composing the target population but also their characterization. We refer here to `rclass.cah`, the function implementing our hierarchical classification algorithm and `res.class_cah`, the object containing the results of the classification. The main command line of our algorithm is as follows:

```
Res.class_ag<-rclass.cah(mbase.stat,dist="Simsymb",agglom="borneResilience",dendo=TRUE, list.rep=TRUE, arbre.d=TRUE)
```

Figure 1 below shows the dendrogram corresponding to the hierarchical classification of 50 people from our samples. As a reminder, a dendrogram is a

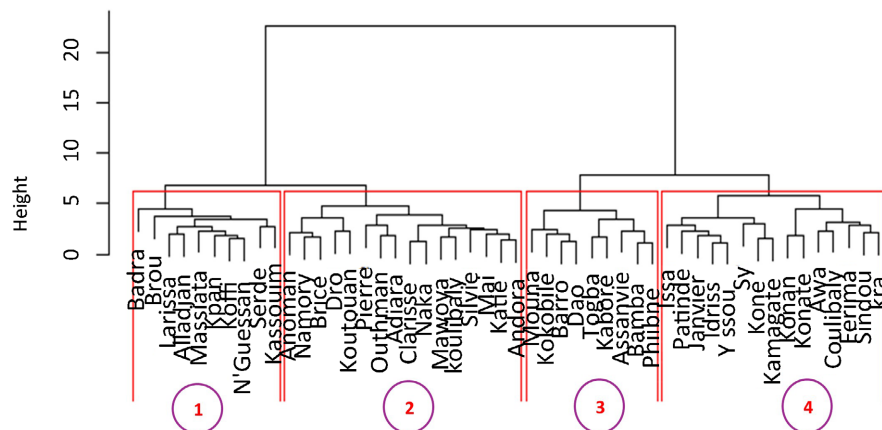


Figure 1. Classification of 50 individuals based on their resilience capacity.

hierarchical grouping diagram, enabling data to be organized in a tree structure according to their similarities.

As can be seen, **Figure 1** shows the existence of four (4) classes among the individuals presented. Each class groups similar individuals according to a common characteristic. The dendrogram is therefore limited to showing the number of groupings that can logically emerge from the sample, without saying more.

The characterization of these classes using the decision tree technique will therefore provide more information on these groupings, for eventual decision-making.

6.2. Stage of Characterization of Variables

After the identification of classes, the next step in our algorithm concerns their characterization or the identification of discriminant variables by the decision tree technique. For this, we use the initial database (base.stat). Generally, for a better exploration of a data set, there is much interest in the characterization of certain variables. The aim is to understand the structuring of the data set by highlighting the existing relationships between certain variables and to find the most discriminating variables.

The execution of our basic function `rclass.cah`, allows to obtain the following characterization based on the psychosocial state of individuals belonging to different classes (**Figure 2**).

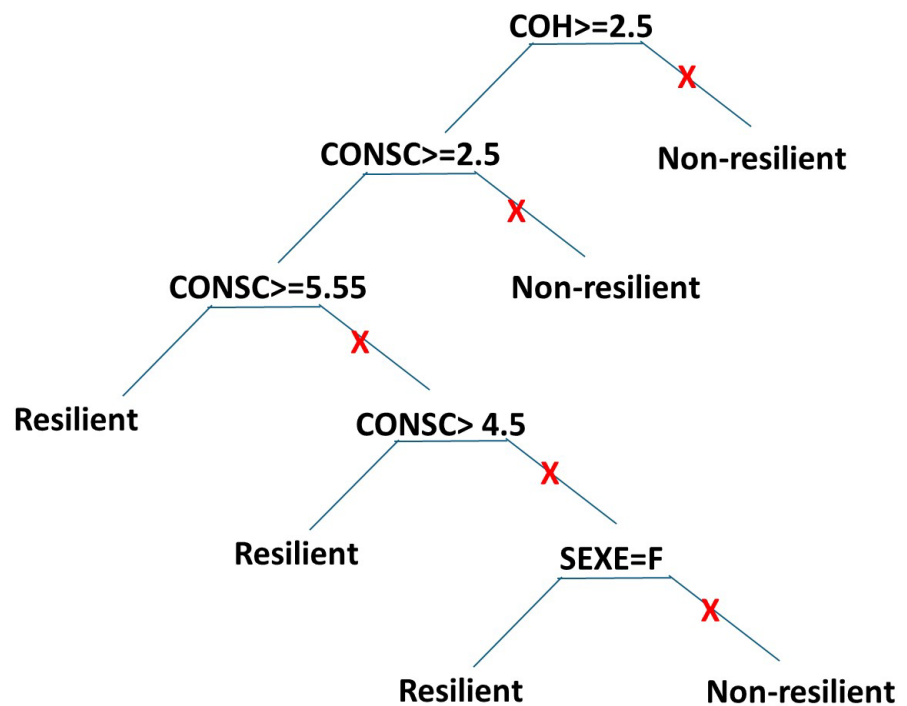


Figure 2. Decision tree characterizing the psychosocial state of individuals.

Based on the information collected, reading the tree above (see the leaves of the tree) allows us to distinguish 4 categories or classes of people, according to their level of resilience of which 3 groups are resilient and the last is not resilient. This result of the decision tree is therefore in phase with the number of classes identi-

fied in the first part of our algorithm. The categories identified are:

- the class of resilient people with a consciousness gradient greater than or equal to 5.5,
- the class of resilient people with a gradient of consciousness strictly below 5.5 and above 4.5,
- the class of resilient people with a gradient of consciousness below 4.5 and above 2.5 and non-feminine.
- Finally, the class of non-resilient persons who do not meet one of the resilience criteria above.

Moreover, it can be seen that the degree of consistency of individuals (CONSC) was considered by the algorithm as the most discriminating variable among the different descriptors used.

6.3. Validation of Our Approach

To validate our approach, we will compare its performance with that of some methods in the literature. However, it is important to remember that the performance of a classification method depends heavily on the field of application of its measure of similarity or dissimilarity and the type of data used. Based on this assumption, it would be more appropriate to compare our approach with classification models adapted to the analysis of social resilience processes. Thus, only those of (Coulibaly *et al.*, 2017) [8] meet this requirement.

In the field, the performance of machine learning algorithms is usually evaluated using the classical precision, recall and F-measure indices. These are calculated from the four (4) response possibilities when a classification query is made (see **Table 1** below).

Table 1. The possibilities of responding to a classification request.

True Positive (TP)	The classifier correctly finds the element as belonging to the class
False positive (FP)	Classifier mistakenly finds the element as belonging to class
True Negative (TN)	The classifier correctly finds the element as not belonging to the class
False Negative (FN)	The classifier mistakenly finds the element as not belonging to the class

In our context, relative to a given class:

- Precision represents the proportion of elements assigned to the class that are relevant. It measures the ability of a classifier to reject irrelevant elements.

$$Precision_{C_i} = \frac{TP_{C_i}}{TP_{C_i} + FP_{C_i}} \quad (6)$$

- Recall is the proportion of relevant items that are assigned to the class. Measures the classifier's ability to give all relevant elements.

$$Recall_{C_i} = \frac{TP_{C_i}}{TP_{C_i} + FN_{C_i}} \quad (7)$$

- The F-Measure (or F1 Score) represents the harmonic mean of precision and recall. It measures the ability of the classifier to give all relevant elements and to refuse others.

$$F - Measure_{C_i} = 2 * \frac{Precision_{C_i} * Recall_{C_i}}{Precision_{C_i} + Recall_{C_i}} \quad (8)$$

In this section, to calculate these three indices, we will use the results from the implementation of the two classification approaches: [Coulibaly *et al.*, 2017) [8] and our present approach], with the data mentioned in section 5.

The results from the calculation of indices are given in **Table 2** below. The values reported are the averages obtained from 8 runs of each approach under the same initial conditions.

Table 2. Comparison of the performance of the two classification approaches.

Number of classes	Precision		Recall		F-Measure	
	(Coul <i>et al.</i> , 2017) [8]	CAH-ET	(Coul <i>et al.</i> , 2017) [8]	CAH-ET	(Coul <i>et al.</i> , 2017) [8]	CAH-ET
K = 2	0.381	0.384	0.950	0.950	0.543	0.546
K = 3	0.490	0.518	0.851	0.850	0.621	0.643
K = 4	0.483	0.550	0.724	0.725	0.579	0.625
K = 5	0.510	0.562	0.703	0.701	0.591	0.623

The following **Figures 3-5** respectively show the results of **Table 2** above according to Precision, Recall and F-Measure.

The analysis of these results allows us to make the following observations:

- When increasing the number of classes, accuracy increases while recall decreases. This confirms the principle of compromise between these two metrics, hence the reason for their joint study in the evaluation of models. Indeed, mathematically, the progressive increase in the classification threshold also induces an increase in positives (VP and FP), resulting de facto in an increase in precision [18]. The number of PS is decreasing faster than the number of VPs, so there will be a decrease in the re-call. Conversely, when the classification threshold is lowered, the number of false negatives decreases and the number of true positives increases, resulting in an increase in recall and a decrease in precision.
- On the three metrics, our approach has a global advantage over the approach (Coul *et al.*, 2017) [8], although both approaches have almost similar results for the Recall (with a slight advantage of the approach (Coul *et al.*, 2017) [8] at k = 3 and k = 5). This result shows, compared to the other approach, the performance of our approach. In other words, our classifier makes good predictions of social resilience data with a higher level of accuracy or quality of classes (Precision) than (Coul *et al.*, 2017) [8] and a level of completeness or quantity of classes (Recall) more or less equivalent to the above approach.

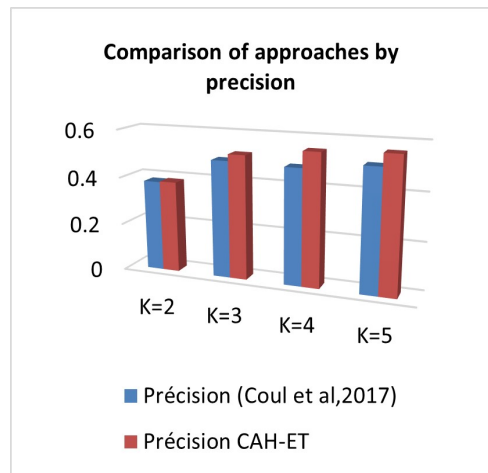


Figure 3. Comparative results of the Precision between CAH-ET and (Coul *et al.*, 2017 [8]).

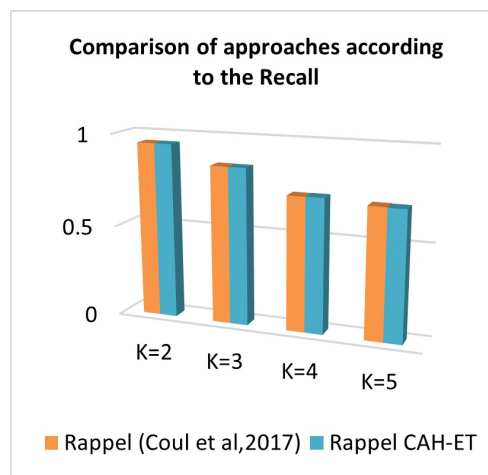


Figure 4. Recall comparative results between CAH-ET and (Coul *et al.*, 2017 [8]).

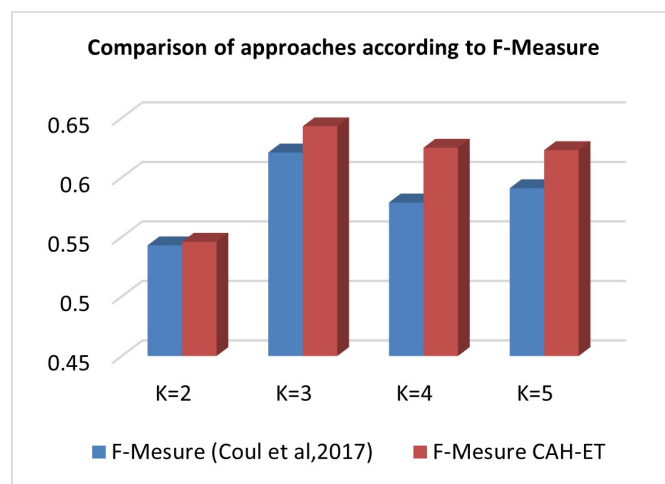


Figure 5. Comparative results of the F-Measure between the approaches CAH-ET and (Coul *et al.*, 2017 [8]).

7. Conclusions

Ultimately, our proposal for a hybrid automatic classification method for latent traumatic states, combining the techniques of hierarchical ascending classification (HAC) and decision trees, represents a significant advance in the field of social resilience process data analysis. This integrated approach not only makes it possible to identify homogeneous groups of individuals presenting similar traumatic states, but also provides clear, interpretable decision-making tools for practitioners. The results obtained demonstrate the effectiveness of this hybrid method, which combines the robustness of CAH for data segmentation with the predictive power of decision trees. By facilitating understanding of the mechanisms underlying traumatic states, this method paves the way for more targeted and personalized interventions, contributing to better care for those affected.

However, the robustness of this approach needs to be strengthened, especially in terms of execution time, for better analysis of large-scale data, and its application to other areas in addition to social resilience processes needs to be explored.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] World Health Statistics (2008) Part I: Ten Highlights in Health Statistics, Future Trends in Global Mortality: Major Shifts in Cause of Death Patterns. 30.
- [2] Nation Unites (2005) Déclaration de Hyogo. Conférence mondiale sur la prévention des catastrophes.
- [3] Tiekoura, C.K., Marcellin, B.K., Odilon, A., Michel, B. and Boko, A. (2016) SimCT: A Measure of Semantic Similarity Adapted to Hierarchies of Concepts.
- [4] Xiao, E. (2024) Comprehensive K-Means Clustering. *Journal of Computer and Communications*, **12**, 146-159. <https://doi.org/10.4236/jcc.2024.123009>
- [5] Ntayagabiri, J.P., Ndikumagenge, J., Ndayisaba, L. and Philippe, B.K. (2023) Study on the Development and Implementation of Different Big Data Clustering Methods. *Open Journal of Applied Sciences*, **13**, 1163-1177. <https://doi.org/10.4236/ojapps.2023.137092>
- [6] Shukran, M.A.M., Mohd Yunus, M.S.F., Abdullah, M.N., Isa, M.R.M., Khairuddin, M.A., Maskat, K., *et al.* (2022) Survey on Clustering Techniques for Image Categorization Dataset. *Journal of Computer and Communications*, **10**, 177-185. <https://doi.org/10.4236/jcc.2022.106014>
- [7] Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A.Y., *et al.* (2014) A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis. *IEEE Transactions on Emerging Topics in Computing*, **2**, 267-279. <https://doi.org/10.1109/tetc.2014.2330519>
- [8] Kpinna, C., Konan, B., Michel, B. and Oumtanaga, S. (2017) Evolutionary Method of Population Classification According to Level of Social Resilience. *International Journal of Advanced Computer Science and Applications*, **8**, 139-144. <https://doi.org/10.14569/ijacsa.2017.080119>
- [9] Bellanger, L., Coulon, A. and Husi, P. (2021) Une méthode de classification as-

- cehdante hiérarchique par compromis: hclustcompro. *9e Conférence Internationale Francophone sur la Science des Données (CIFSD)*, Marseille, June 2021, 49-60.
- [10] Hakami, F. and Chang, L. (2023) Enhanced Arrhythmia Classification through Feature Engineering and Hierarchical Modeling. *2023 IEEE International Conference on Big Data (BigData)*, Sorrento, 15-18 December 2023, 6155-6158. <https://doi.org/10.1109/bigdata59044.2023.10386806>
- [11] Moiraghi, F., Palmonari, M., Allavena, D. and Morando, F. (2024) Zero-Shot Hierarchical Classification on the Common Procurement Vocabulary Taxonomy. *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, Osaka, 2-4 July 2024, 273-278. <https://doi.org/10.1109/compsac61105.2024.00045>
- [12] Guha, S., Rastogi, R. and Shim, K. (1998) Cure: An Efficient Clustering Algorithm for Large Databases. *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, Seattle, 2-4 June 1998, 73-84. <https://doi.org/10.1145/276304.276312>
- [13] Rakotomalala, R. (2005) Arbres de Décision. *Revue Modulad*, No. 33, 163-187.
- [14] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [15] (2007) Éric Parent et Jacques Bernier: Le raisonnement bayésien. Springer-Verlag.
- [16] Tiekoura, C.K. and Abdou, M. (2019) Dynamic Models for Measuring Social Resilience. *International Journal of Computer Science and Information Security (IJCSIS)*, **17**, No. 10.
- [17] Tiekoura, C.K., Yapou, A.O., Konan, M.B. and Michel, B. (2015) Resilometrical Modeling of Interactions in Social Resilience Dimensions. *International Journal of Computer Science Issues (IJCSI)*, **12**, 149.
- [18] Tremblay, C. (2024) Precision Recall Curve|Performance en Classification Machine Learning. Kobia.