An Effective Prediction Approach for the Management of Children Victims of Road Accidents

F. Saadi[1]*, B. Atmani[2], F. Henni[3], H.Benfriha[4], Z.Addou[5], R. Guerbouz[6]

[1] Laboratoire d'Informatique d'Oran (LIO), University Oran 1 (Algeria)
[2] Laboratoire d'Informatique d'Oran (LIO), University of Mostaganem (Algeria)
[3] Computer Science and new Technologies Lab (CSTL), University of Mostaganem (Algeria)
[4] Laboratoire d'Informatique d'Oran (LIO),Institut teccart (Canada)
[5] Réanimation Polyvalente Pédiatrique EHS Canastel, University of Oran 1 (Algeria)
[6] Service Neurochirurgie, Etablissement Hospitalo Universitaire Oran (Algeria)

* Corresponding author: saadi_fatima@hotmail.fr

Received 28 October 2022 | Accepted 23 December 2022 | Published 5 February 2024

Abstract

Road traffic generates a considerable number of accidents each year. The management of injuries caused by these accidents is becoming a real public health problem. Faced with this latter, we propose a new clinical decision making approach based on case-based reasoning (CBR) and data mining (DM) techniques to speed up and improve the care of an injured child. The main idea is to preprocess the dataset before using K Nearest Neighbor (KNN) Classification Model. In this paper, an efficient predictive model is developed to predict the admission procedure of a child victim of a traffic accident in pediatric intensive care units. The evaluation of the proposed model is conducted on a real dataset elaborated by the authors and validated by statistical analysis. This novel model executes a selection of relevant attributes using data mining technique and integrates a CBR system to retrieve similar cases from an archive of cases of patients successfully treated with the proposed treatment plan. The results revealed that the proposed approach outperformed other models and the results of previous studies by achieving an accuracy of 91.66%.

Keywords

Case Based Reasoning, Data Mining, Decision Tree, Medical Decision Making Approach, Predictive Model, Road Accident, Selection of Relevant Attributes, Traumatic Brain Injuries.

DOI: 10.9781/ijimai.2024.02.001

I. Introduction

Road accidents can have catastrophic physical and psychological effect on individuals, their families, and the community. According to the World Health Organization, road accidents cause 1.35 million fatalities and 20 to 50 million illnesses every year. Eighty percent of road traffic accident victims are children with traumatic brain injury, severe brain damage, and associated chest or abdominal disorders https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. Providing prompt and appropriate first aid, medical diagnosis and treatment to road traffic accident victims in the critical first few hours after the incident significantly increases their chances of survival and reduces the severity of injuries. Doctors may face critical problems when making quick decisions about diagnosing, especially when a child's life is at stake. This study aims to accelerate and improve the management of an injured child by predicting the course of actions for admission to the services of Pediatric Traumatic Brain Injury (PTBI) following the ABCDE protocol that consists of the following: Airway (Voice, Breath sounds); Breathing (Respiratory rate, Chest wall movements, Chest percussion...); Circulation (Skin color, sweating, Capillary refill time...), Disability (Alert, Voice responsive, Pain responsive), and Exposure (Expose skin) [1].

To achieve our objective, we propose a new medical decision-making approach guided by CBR and data mining techniques. The development of the database was entirely carried out by the pilot team of the University Training Research Project entitled "Implementation of a Medical Decision Support System adapted for the care of child victims of accidents". The collection of patients' records were done at three medical institutions in Oran, Algeria: the pediatric surgery clinic and the intensive care unit of the University Hospital Center; and the intensive care unit of Canastel Pediatric Hospital.

The proposed approach includes a multi-step procedure: Collection, preprocessing, selection of relevant attributes by data mining technique, and Case-Based Reasoning (CBR). The effectiveness of CBR relies heavily on the quality of its case-base content [2]. Before incorporating the data collected into Clinical Decision Support Systems' (CDSS) knowledge base [3], it becomes imperative to assess and enhance the data's quality. Consequently, data preprocessing procedures assume primary importance as they lay to improve the accuracy of CBR

Please cite this article in press as:
F. Saadi, B. Atmani, F. Henni, H.Benfriha, Z.Addou, R. Guerbouz. An Effective Prediction Approach for the Management of Children Victims of Road Accidents, International Journal of Interactive Multimedia and Artificial Intelligence, vol. 9, no. 3, pp. 104-114, 2025, http://dx.doi.org/10.9781/ijimai.2024.02.001

systems [4]. Preprocessing the dataset before prediction is the main concept in this work. This step itself consists of four steps called cleaning, discretization, statistical analysis and classification.

Case-based reasoning (CBR) stands as an essential component of artificial intelligence, involving the resolution of novel problems by drawing upon existing solutions from similar cases [5]. This methodology aligns seamlessly with the problem-solving approach adopted by healthcare professionals when confronted with new cases, making its integration into clinical settings a natural fit. The allure of CBR in medical domains lies in the existence of a comprehensive case base, housing a wealth of patient-specific information, including symptoms, diagnoses, treatments, and outcomes. Consequently, numerous CBR systems have been developed to aid in medical diagnosis and decision support [2].

Designing an effective CBR system hinges on two crucial steps: selection of relevant attributes and case retrieval. The choice of appropriate features for case representation and the effective retrieval method are pivotal in ensuring the quality of CBR [6]. As a result, much of the research on CBR primarily revolves around addressing these specific issues [7].To improve our CBR system, DM techniques are used. The concept "data mining" has become more often employed in the literature on medicine during the past several years [8]. Its use in the processing of medical data, however, has only lately become somewhat more widespread. This is especially true for real-world applications in clinical medicine, which may profit from particular data mining techniques that can perform predictive models, take advantage of the clinical domain's knowledge, and describe proposed decisions after the models have been used to support clinical decisions [8]. The decision tree is integrated in the approach as a discriminating tool that selects the most relevant attributes, removes less significant attributes, and thereby reduces the number of attributes utilized in the similarity calculation, hence speeding up the retrieval phase [9].

The paper is organized as follows: Section II summarizes the relevant works available in the literature. The proposed approach is described in detail in section III. Section IV includes the experimental results obtained and a comparison with other selected research works. Finally, conclusions are provided in Section V.

II. Background

The volume, complexity and dynamics of clinical information are a challenge for doctors. Since the 1960s, researchers have envisioned the day when computers could help build a system that processes medical data like a doctor, offers insight into the nature of medical issue resolution, and permits the building of formal clinical reasoning models, and, most importantly, decreases medical error and misdiagnosis. Any technology that can improve the ability to correctly diagnose human disease is a necessary advancement for the well-being of humanity.

Information technologies have been used in the medical field since computers were invented, and various types of computer applications have been created, including Clinical Decision Making Approaches (CDMA) which have been increasingly popular in several medical fields [10]. According to these authors [10], the CDMA improves healthcare quality by providing more precise diagnoses and treatments that are more effective and reliable, while also lowering healthcare costs and avoiding errors caused by doctors' lack knowledge. The reasoning is a crucial function performed by the CDMA inference engine, which combines medical knowledge with patient-specific data and delivers relevant decisions. Several methods for representing medical reasoning have been developed in Artificial Intelligence (AI) and are employed in the construction of decision support systems. These include rule based reasoning [10], [11], Artificial Neural Network [12], KNN [13], ontology [14], association rules [15] case-based reasoning [16]–[18], decision trees [19], random forest [20], fuzzy logic [21], [22], ...etc. Table I lists some Clinical Decision Support Systems (CDSS), their reasoning techniques, and their application area. CBR has demonstrated to be highly promising in using intelligent systems in the healthcare business, among the various AI reasoning techniques. It is a significant methodology and an area of machine learning [5] used in the creation and enhancement of CDSS in a number of projects involving diverse medical applications [9].The first CDSS used pure CBR, i.e. they used CBR alone as a reasoning methodology, and these systems were successful. But their use in practice remains limited due to the complexity of the medical field which cannot be treated with pure CBR.

Many successful hybridization techniques of CBR with other computing methods have been developed in Artificial Intelligence to try to solve the limitations of classical CBR and to achieve the mission of a CDSS which is the improvement of time and accuracy of medical diagnosis [19].

Retrieve-Reuse-Revise-Retain are the four main steps that make up the CBR cycle, the first step (Retrieve) is the most important and expensive phase in the cycle. In this phase the CBR system searches between past cases the most similar case, or more precisely the case that has the problem part similar to the target case. This research is based on the similarity calculus among cases [9]. To develop an efficient CBR-based system, the complexity of the retrieval step must be optimized and improved. Several research works have experimented the integration of techniques issued from DM to improve the efficiency of this step. Guo and Wu [23] developed a Bayesian Network model and CBR hybrid system for use in the healthcare industry. The goal of this system is to increase case retrieval accuracy in the context of big data. Mansoul & Atmani [24] proposed a new approach destined for the medical field that combines CBR and multi-criteria analysis (MCA). This combination aims to facilitate the choice of a solution among several solutions proposed in the retrieval step; which improves this phase. Benfriha et al. [25] suggested an approach for the initial health care of a child's PTBI in the traffic accident. The approach's goal is enhance the retrieval phase in the CBR cycle by using multi-label text categorization which allows reducing the search space and keeping only the most relevant cases and therefore speeding up the retrieval step. The retrieval is related to the representation of the data and to the similarity measures used. Indeed, two aspects of retrieval can quickly become time-consuming, particularly when it comes to solving real problems: the case's basic size and the large number of features. Many works on CBR has proposed improving the retrieval phase through the reorganization of the case base. Mansoul & Atmani [26] proposed a CBR system for predicting the best presumptive diagnosis of orthopedic illnesses. During the retrieval process, the suggested systems include clustering to decrease the base case. Malviya et al. [27] created a medical image retrieval system named (CBMIR) for finding and recovering lung computed tomography (CT) images from a large medical image dataset; this system combines CBR with k-means clustering-based segmentation. Saadi et al. [28] integrated fuzzy clustering in the retrieval phase of CBR. The goal of this work was to minimize the case base and improve the retrieval step by optimizing the similarity computation time. Benamina et al. [29] combined a fuzzy decision tree and CBR to realize a Fuzzy CBR medical system for diabetic patients. The objective of this combination is to improve the retrieval phase by reducing the complexity of the similarity calculation.

The amount of attributes clearly influences the similarity calculus. The more attributes, the higher the processing time is. To reduce the number of attributes, several DM techniques were proposed for

TABLE I. Medical Systems Integrated Data Mining Techniques in CBR Cycle

Reference	System	Technique(s) used	Application field
[10]	-	Rule-Based Reasoning CBR	Management of Drugs Intoxications in Childhood
[11]	-	Rule-Based Reasoning	Asthma diagnosis
[12]	-	ANN	Classification of causes for non- adherence with medication
[13]	-	K-Nearest Neighbor (KNN	Prediction the diabetes disease
[14]	-	Ontology	Depression
[15]	-	Bayesian Network association rules	Diagnosis Vaccination: Detection of lost ones and abundance causes in relation to the mother's socio-economic characteristics.
[16]	Protos	CBR	Classification and Diagnosis Hearing disorders
[17]	Electronic Medical Record system (ArdoCare)	CBR	Electronic healthcare suspicion of aortic pathology
[18]	CBR-Nursing Care Plans System (CNCPS)	CBR	Formulation of effective nursing care plans
[19]	-	Decision trees	Classification of diabetes, hepatitis and heart diseases
[20]	-	Random Forests Classifier	Diagnosis of Heart Arrhythmia
[21]	CKD (Chronic Kidney Disease) diagnostic fuzzy expert system	Fuzzy logic	Prediction chronic kidney disease
[22]	-	Fuzzy C-Means (FCM) clustering	Prediction of Chronic Kidney Disease Progression

selecting relevant variables for the similarity calculation in order to speed up the retrieval step. As a result, the relevant variables are chosen with the goal of reducing the number of attributes used to measure similarity, allowing decreasing the computation time and accelerating and improving the retrieval phase and model performance, such as predictive accuracy [30]. Feuillâtre et al. [31] developed a clinical decision tree for the selection of attributes in order to optimize the retrieval step in CBR. Jarmulak et al. [32] applied the genetic algorithms for the selection of relevant features. Ayed et al. [33] eliminated noisy and redundant attributes and left only the relevant ones for the application of similarity measures by using the Relational Evidential C-Means (RECM). Addisu et al. [34] designed a CBR diagnostic framework for malaria diagnosis and applied an information gain algorithm to select relevant attributes. Henni et al. [35] enhance the retrieval step by reducing the number of attributes using the random forest algorithm. Saadi et al. [9] proposed an approach guided by CBR and decision tree. This work integrates a decision tree to eliminate irrelevant attributes used in the similarity calculation.

III. The Proposed Predictive Approach (TBI-CDMA)

In this article, we propose a new Clinical Decision Making Approach (CDMA) based on CBR, intended for the management of children victims of traffic accidents and who have suffered a traumatic brain injury. This approach is adopted for developing an efficient predictive model (TBI-CDMA) to predict the admission procedure of pediatric traffic accidents victims in pediatric intensive care units. We apply this approach to a real dataset that we have collected following the ABCDE protocol. ABCDE is a protocol for assessing and treating critically ill or wounded people in a systematic way [1]. The Protocol can be used in any clinical emergency, in the street without any equipment or, in a more sophisticated form, in emergency rooms, general wards of hospitals and critical care units. It was developed by the American College of Surgeons in the USA to improve the management of polytrauma victims by early detecting physiological changes that put them in danger of mortality [36]. Since the PTBI dataset collected represents only the signs and symptoms existing in ABCD, the experts preferred to remove the gestures associated with Exposure (E) and followed the ABCD protocol. Fig. 1 summarizes the essential phases of the proposed system.

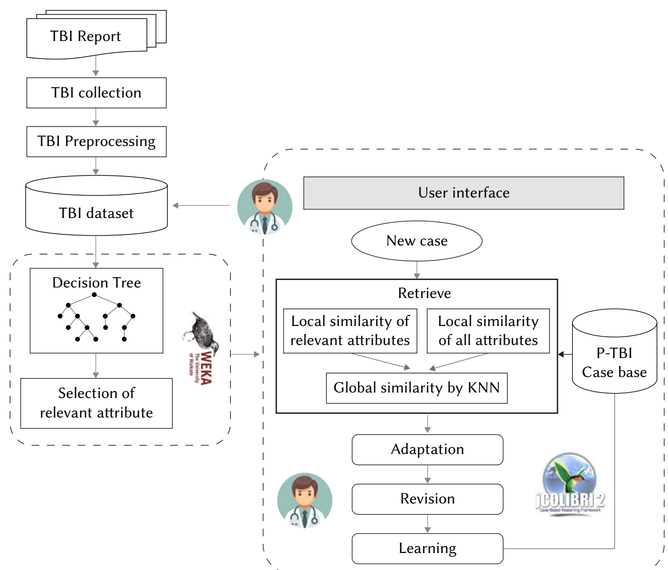

Fig. 1. Overview of TBI-CDMA.

This section presents the steps described in Fig. 1.

A. PTBI Collection

The term "data collection" refers to a systematic approach that entails gathering and measuring many types of information in order to provide a comprehensive and precise picture of an area of interest. The collection of data allows us to answer pertinent questions, evaluate results, and better predict future probabilities and trends. In the intensive care unit at the study site, most of the clinical and administrative information circulating in all the health services is still kept on paper. For this study, we used clinical reports from child traffic accident victims. Since this is the first study in this area, we have selected the records of patients (children) who present a head injury.

B. PTBI Preprocessing

Prior to constructing a model, the data requires preprocessing to derive usable variables from raw data. This crucial step involves extracting meaningful variables and values from the data, and the expertise of healthcare professionals plays a significant role in this process. While removing missing values and outliers generally enhances algorithm accuracy, it is essential to recognize that missing data can carry valuable information, a discernment only achievable by a healthcare professional. For this fact, the main idea in our work is preprocessing the dataset before making the prediction. The preprocessing stage is divided into four sections: cleaning, discretization, statistical analysis and classification. The management of noise or missing values will be performed on the collected dataset as a first preprocessing step. Then, a step of discretization of continuous values into a categorical representation to preserve the learning efficiency and the relevance of the model to build [37], knowing that this step was done with the help of experts who gave us the cut-off points for the attributes. The next step consists in a statistical analysis on the variables by applying a descriptive analysis for each variable and evaluating the correlation and relation between features (attributes) using the Mann-Whitney U test (or Wilcoxon rank-sum test) [38], and finally an evaluation of some classifiers on this dataset is mandatory to validate PTBI. The confusion matrix in Table II is used to compute the performance metrics that are employed in the evaluation.

- Accuracy: is the rate of correct classification, it's defined by (1):

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{1}$$

- Precision: the frequency with which the classifier correctly classifies all of the predictions, as shown in (2) (1):

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

- Recall: Proportion of actual positive results correctly identified, as seen in (3):

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

- F-Measure: is calculated based on precision and recall, as shown in (4):

$$F - Measure = \frac{2 * TP}{2TP + FP + FN} \tag{4}$$

- Kappa Statistic: is a measure of inter-rater agreement for categorical or nominal data. It quantifies the level of agreement between two or more annotators (observers) in the context of classification or labeling tasks. Kappa values range between -1 and 1, where 1 indicates perfect agreement, 0 indicates agreement equivalent to random chance, and values less than 0 signify less agreement than expected by chance..
- ROC: is the ratio between the recall (True Positive Rate TPR (3)) and the FPR (False Positive Rate defined in (5)):

$$FPR = \frac{FP}{FP + TN} \tag{5}$$

- Specificity: is the rate of true negatives defined by (6):

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

TABLE II. Format of Confusion Table

	Predicted Class			
Actual class	True (TP)	Positives	False (FP)	Positives
	True (TN)	Negatives	False (FN)	Negatives

C. The Selection of Relevant Attributes

The success of CBR (Case-Based Reasoning) relies on the quality of data (cases) and the speed of the retrieval process, which can be time-consuming, especially when dealing with a large number of cases. Consequently, a case may contain a significant number of attributes [39]. Some attributes are crucial for case reasoning, while others may be unnecessary and result in increased complexity during the retrieval phase. To ensure a robust CBR system performance, it becomes essential to manage the content of the case base. One of the proposed solutions to address these issues is attribute selection. There are several methods of selection of relevant attributes like Chi-square, Euclidean distance, T-test, information gain, correlation-based feature selection, decision tree, ...etc.

Some classification algorithms have earned a reputation for being able to focus on relevant attributes while ignoring others that aren't. The decision tree is a widely used technique for attribute selection in machine learning and data mining [40]. This technique involves building a simple structure where non-terminal nodes represent tests on attributes, and terminal nodes reflect decision outcomes. The test attribute at each node is determined using the information gain measure, which tends to favor attributes with a larger number of values. This algorithm which is an extension of the ID3 algorithm (the basic decision tree induction algorithm) was further improved by C4.5 [41], calculates the rate of information gain (Gain Ratio) at the place of information gain. The C4.5 tree utilizes the gain ratio to determine splits and select the most important attributes. Consequently, the C4.5 decision tree excels at focusing on relevant features while disregarding irrelevant ones. Compared to other strategies, C4.5 generates fewer rules, resulting in reduced error rates and higher precision in the result set. Additionally, C4.5 constructs a trimmed tree, ensuring faster results compared to alternative techniques. The Gain Ratio decreases and corrects the information gain bias by taking the intrinsic information of a split into account. It is therefore weighted by a function that penalizes tests that split the elements into too many sub-classes. This distribution measure is named SplitInfo. The equations (7, 8) refers to the Gain Ratio [42]:

$$GainRatio = \frac{InformationGain}{SplitInfo} \tag{7}$$

or

$$GainRatio = \frac{Entropy(before) - \sum_{j=1}^{k} Entropy(j, after)}{\sum_{j=1}^{k} w_j \, log_2 w_j} \tag{8}$$

Where *before* is the dataset before splitting the data. *K* is the number of subsets generated by the splitting, (*j*, *after*) subset *j* after splitting the data.

$$Entropy(S) = \sum_{i=1}^{c} -p_i log_2 p_i \tag{9}$$

S is the current state, p_i is the probability of an event *i* of state *S*.

The attribute with the highest gain ratio is chosen as the splitting attribute [43]. In the decision tree generated, non-leaf nodes are deemed relevant attributes. By employing the C4.5 algorithm, the attributes resulting from the decision tree become the most significant ones, and these will be used in the similarity calculation, instead of using all attributes. This approach ensures that only the most informative attributes are considered for accurate similarity calculation.

D. Case Based Reasoning

The central notion in a CBR system is the case that represents in our dataset a patient. A record in the PTBI dataset is composed of descriptors divided into a problem part and a solution part. At this stage, we have a case base ready to use in the CBR system. The cycle begins with the retrieval step. In order to succeed in this phase, which is based on the similarity calculation, we need to identify the best similarity measures that allow us to produce good results. Similarity refers to a measure that indicates the strength of a relationship between two features or objects. The purpose of conducting a similarity analysis is to compare two lists of components and calculate a single numerical value representing their evaluation [44]. The similarity measurements are instrumental in retrieving similar cases from the case-based approach.

Among the various distance measurement methods, Euclidean distance is the most commonly used, relying on the spatial location of objects. This method computes a distance as the square root of the sum of squares of numerical differences between two analogous objects. The standard Euclidean distance serves as the fundamental approach for describing the relationship between two cases, forming the neighboring figures of an arbitrary case [45]. In CBR, The Euclidean distance stands as the most prevalent distance metric used to measure the similarity of numerical data [46], and the equality distance which is a simple function for categorical data that have fixed values [47]. Thus, we have exploited these local similarity functions (between the attributes of the cases) and implemented them in the jCOLIBRI platform. The equations for these functions are:

Euclidean Distance: the formula of the euclidean distance is defined by (10):

$$D(x,y) = \frac{\sqrt{x^2 - y^2}}{x+y} \tag{10}$$

Equal: if two attributes (x,y) are equal it returns 0, otherwise it returns 1.

After computing the similarity measures between attributes, the global similarity calculates the distance between two cases using the K-Nearest Neighbor (KNN) technique, which is widely employed in Case-Based Reasoning systems. The algorithm computes the similarity between a new case and a stored case by performing a weighted summation of the pairwise attribute similarities.

With the normalization of local similarity functions, the resulting global similarity value lies within the range of 0 to 1, where 0 indicates complete similarity between the cases.

IV. Experimentation and Results

The PTBI dataset collected from the services mentioned before will be used for the prediction of the course of action to be taken on admission to the services concerned. Most collected clinical information was stored in paper format and recorded in chronological order in an archived register. PTBI was collected in the period between 2017 and 2021. This was done using a Java application that captures the information of patients who have traumatic brain injury. Four parts make up the preprocessing step:

- Cleaning: The data cleaning process was essential to ensure the dataset's quality and reliability. Various issues, such as empty, duplicated, unreadable files, and missing values, were identified during a thorough examination. Empty files were promptly addressed by consulting with doctors and deleting them to avoid any misleading results. Duplicated files were carefully managed, keeping only one representative instance to maintain efficiency and consistency in the analysis.
- Discretization: During the discretization process, several variables were transformed into discrete categories or intervals. This was done for variables such as systolic pressure, diastolic pressure, heart rate, Glasgow score, and more. The decision to discretize these variables was based on the recommendations provided by doctors who were actively involved in the research project. They provided valuable insights and expertise in determining the appropriate cut points or thresholds for discretization. In addition to the input from doctors, international protocols used in Algeria were also consulted to guide the discretization process.
- Statistical analysis: Before deciding if the dataset is ready for use, we applied a statistical analysis under SPSS software. A descriptive analysis is performed on every attribute and the correlation between attributes was calculated. This statistical analysis allowed us to validate the PTBI and determine the attributes that we will be using in TBI-CDMA. Our dataset must be in the form of a case base, the case represents in our dataset a patient. PTBI comports 174 cases. A case of PTBI consists of 36 descriptors divided into a problem and a solution presented according to the ABCD Protocol. The problem part is composed of five categories. Fig. 2 shows the description of a case.

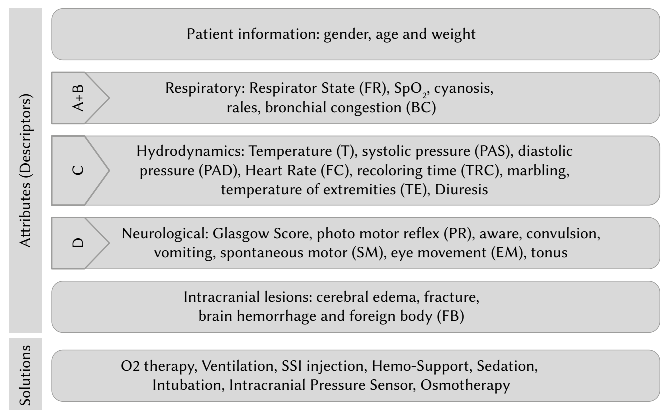

Fig. 2. Description of Case Base.

The solution part of a case encompasses the course of actions to be executed by the doctor, which is presented in the form of plans. To organize and structure these plans effectively, the ABCD protocol is employed.

However, certain data mining techniques, such as the decision tree, require that the class variable be represented with a single column. In collaboration with health experts, a coding system was devised to codify the different plans into six distinct classes. These classes are described in detail in Table III, with each class listed in descending order based on the severity of the patient's condition upon admission to the pediatric intensive care unit. The organization of these classes allows for effective utilization of DM techniques, such as decision trees, to aid in clinical decision-making processes.

- C5: it is the conduct to hold for a seriously affected case, it includes oxygenation with a nasal probe, non-invasive ventilation, application of an osmotherapy and hemo-support, it must be injected with an isotonic saline, intubated and sedated, and monitored their intracranial pressure.
- C4: designates the actions to be done for patients a little less serious compared to the previous class, these actions are represented by oxygenation with a nasal probe, invasive ventilation, isotonic saline injection, intubation and sedation.
- C3: includes actions to be taken for a patient who does not have a serious neurological problem, so it is enough to give him oxygenation with a mask, non-invasive ventilation, and isotonic saline injection.

TABLE III. Coding of Classes

A+B		C		D				Class
O2 therapy	Ventilation	SSI injection	Hemo-Support	Sedation	Intubation	Intracranial Pressure Sensor	Osmotherapy	
nasal probe	invasive ventilation	✓	✓	✓	✓	✓	✓	C5
nasal probe	invasive ventilation	✓		✓	✓			C4
Mask	non-invasive ventilation	✓						C3
Mask	non-invasive ventilation							C2
		✓						C1
								C0

TABLE IV. Results of the Application of Wilcoxon Rank-sum Correlation Test Between Each Feature and the Target Feature

Attribute	Age	Gender	Weight	PAS	PAD	FC	TE	Diuresis	Marbling	TRC	SPO_2	FR	Cyanosis	RALES	BC	Glasgow	Tonus	Aware	Convulsion	Vomiting	SM	EM	PR	Fracture	CE	FB	BH
P_value	0,31	0,05	0,19	0	0,79	0,45	0,42	0	0	0	0,03	0,91	0	0,62	0	0	0,04	0	0	0,1	0	0	0	0	0	0	0

- C2: designates the course of actions for patients who ot have a neurological disability or a hemodynamic problem but solely have respiration problems, oxygenation using a mask and non-invasive ventilation are the adequate actions.
- C1: represents the solution for a patient who has a small circulation problem, an injection of SSI is enough for him.
- C0: denotes that the patient is in good health and does not require treatment.

The authors utilized a widely used univariate statistical test called the Wilcoxon rank-sum test. This test generates a p-value, which represents the probability of a feature being associated with the target variable. The resulting p-value serves as a score that facilitates the creation of a ranking, highlighting the attributes (features) based on their level of correlation with the target variable.

A low p-value, close to 0, indicates a strong relationship between the examined attribute and the class variable. Consequently, such attributes are considered highly relevant in the context of the analysis. On the other hand, a high p-value, close to 1, suggests that the attribute is not correlated with the class variable, making it less relevant for the study (Table IV).

By employing the Wilcoxon rank-sum test and analyzing the resulting p-values, the authors were able to identify and rank the attributes based on their degree of linkage (correlation) with the target variable. This approach helped to identify significant attributes that are closely related to the class variable and those that have less relevance in the analysis [48]. Upon referring to Table IV, it becomes evident that the variables highlighted in grey are strongly correlated (p-value=0) with the class variable, which represents the action plan implemented by the doctor. The shading of the grey color varies, with a darker shade indicating a higher correlation coefficient, while a lighter shade suggests a decrease in the correlation strength. In contrast, cells without any color signify variables that are not correlated with the class variable.

- Classification: The dataset is evaluated using Bagging (REPTree), NB, PART, J48, RF, SVM, and KNN classification techniques. Table V lists the results of the different evaluation measures.

  According to the data presented in Table V, the evaluation measures in terms of Accuracy, F-Measure, Precision, Recall, Kappa, and ROC Area exceeded 89%. These impressive results provide encouraging evidence, validate the reliability of our dataset, and provide strong support for the success of our research study.

TABLE V. Evaluation Measure of Dataset

Classifier	Accuracy	Precision	Recall	FMeasure	Kappa	ROC Area
Bagging	92.52	0.92	0.92	0.91	0.89	0.99
BN	93.10	0.93	0.93	0.93	0.89	0.99
PART	94.82	0.94	0.94	0.95	0.92	0.99
SVM	98.27	0.98	0.98	0.98	0.97	0.99
J48	94.25	0.93	0.94	0.93	0.91	0.99
KNN (IBK)	100.00	1.00	1.00	1.00	1.00	1.00
RF	100.00	1.00	1.00	1.00	1.00	1.00

PTBI case base was randomly split into train and test sets (70%, 30%). In order to obtain better results, several decision trees are generated under the Weka, Orange, Tanagra, Sipina and R platforms. The decision trees obtained are similar. They have four levels, except the Weka platform which generated a tree of 5 levels. We opted for the decision tree (Fig. 3) implemented by the algorithm C4.5 (J48) under the WEKA platform because the attributes it generated are the most logical. C4.5 select the attribute having the maximum informational gain rate value for splitting the node.

The results obtained from the decision tree analysis (Fig. 3) demonstrated the significant benefits of utilizing a DM technique. By employing the decision tree, a substantial reduction in attributes was achieved. Specifically, the decision tree identified and listed seven highly relevant attributes with the highest informational gain rate.

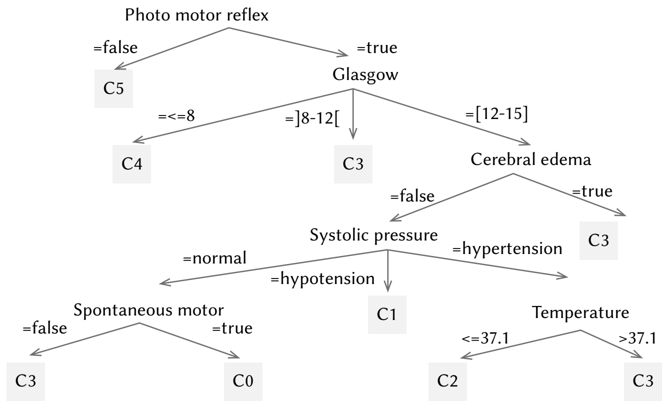

Fig. 3. Decision tree generated by WEKA.

As mentioned previously, conducting a statistical study on PTBI was crucial to assess the correlation between various attributes and the target class. Interestingly, the attributes generated by the decision tree were found to be among those exhibiting strong correlations with the class variable. These attributes proved to be instrumental in predicting the class, thereby confirming the effectiveness of our model. The validation of these results was further supported and endorsed by our collaborating doctors. Their expertise and insights reinforced the credibility and reliability of the model's outcomes, validating the relevance and importance of the selected attributes for predicting PTBI. Overall, the implementation of the decision tree and the statistical study offered valuable insights into the significance of certain attributes in the context of PTBI prediction.

After the generation of the case base and the selection of the most relevant attributes, it is now possible to model the TBI-CDMA solution in the framework jCOLIBRI.

- The case representation: In the context of CBR systems, there is no standardized representation for a case, leading us to adopt a vector form representation to effectively capture and organize the information collected and available in the files. The case is structured into two distinct parts: the problem part, which pertains to the PTBI situation. it encompasses crucial details such as the clinical signs, symptoms, and intracranial lesions related to PTBI, providing essential insights into the patient's condition and medical history. On the other hand, the solution part outlines the specific physical gestures applied as a course of action to address the identified PTBI situation. Within these parts, a total of 29 descriptors are included to comprehensively account for the relevant information.
- The retrieval phase is the cornerstone of a CBR system, where the system identifies the most relevant cases by calculating their similarity to the target case stored in the case base. To compute local similarities of the attributes, two similarity measures are employed based on the attributes revealed by the decision tree (refer to Fig. 3). For numerical attributes, the Euclidean distance is used to measure similarity, enabling the system to quantify the similarity between the target case and other cases stored in the case base. For categorical attributes with fixed values, the equal function is applied to assess their similarity. This function facilitates the comparison of categorical attributes by determining whether their values are the same or not.

  The global similarity calculates the distance between two cases in the jCOLIBRI platform. This platform follows the principle of the K-Nearest Neighbor (KNN) method. In this context, the value of k for the k-nearest neighbor is set to 3, meaning that the system considers the three most similar cases to the target case. After identifying the three nearest cases, a majority vote is applied to determine the appropriate gesture that should be taken to save the child's life. This voting process involves considering the actions taken in the three nearest cases and selecting the gesture with the highest frequency as the recommended course of action for the current case.
- Adaptation: After retrieval, the system evaluates the degree of similarity of the selected cases with the current case. The degree of similarity determines whether an adaptation is necessary or if the solution can be used as it is. If the solution requires adaptation, the doctors are involved in adapting the therapy for the child victim of a road accident. Concerning this phase, it's common in the majority of CBR-based systems to entrust this task to domain experts due to the absence of predefined adaptation rules. Furthermore, the field of medical diagnosis, which revolves around the well-being of patients, is highly sensitive. In such domains, manual adaptation is viewed as a positive rather than a negative aspect. Additionally, as emphasized in [49], the authors assert that "when the knowledge domain is ambiguous or lacks clarity, the development of automatic adaptation mechanisms becomes challenging or is discouraged."
- Revision: The doctor validates the adapted solution.
- Learning: If the generated solution does not exist in the case base and is considered successful in the review state, this new case is added to the case base allowing the system to learn.

A. Performance Evaluation

Our TBI-CDMA system is designed to assist doctors in intensive care services by predicting the appropriate course of action for managing children who have been victims of road accidents. To assess the system's performance, we applied it to the test cases, which constituted 30% of the case base. We selected five real examples (target cases) from the test set, as outlined in Table VI. These target cases were then subjected to the TBI-CDMA analysis to facilitate a comparison between the results obtained from the retrieval process using all attributes and the retrieval process using only the relevant attributes derived from the decision tree. By conducting this comparison, we aim to evaluate the impact of attribute selection on the system's performance and determine the effectiveness of the decision tree in identifying the most crucial attributes for making accurate predictions. This analysis will provide valuable insights into the system's ability to deliver precise and relevant recommendations, enhancing its overall utility in real-world medical scenarios.

Table VII illustrate the result of applying the TBI-CDMA with all attributes and only the relevant attributes for all target cases given in Table VI.

As demonstrated in Table VII, the TBI-CDMA system achieved successful predictions of the best plan of action for case 3 when utilizing all attributes. However, when employing only the relevant attributes derived from the decision tree, the system achieved successful predictions of the best course of actions for 3 out of 5 cases. The performance of the TBI-CDMA system in accurately predicting the best course of actions highlights the effectiveness of the decision tree in identifying important attributes, leading to improved decision support and more informed medical management for children victims of road accidents. According to the doctors' assessment, TBI-CDMA exhibits tolerable errors, with most incorrectly classified cases still providing applicable plans (gestures that can be performed). For instance, in case 4, the predicted class (C3) includes additional actions like an injection of isotonic saline, despite the actual class being C2, denoting non-invasive ventilation. Similarly, in case 5, the predicted class (C2) includes additional preventive actions like oxygenation and isotonic saline injection, while the actual class is C1, representing only

TABLE VI. Target Cases

Case	Age	Gender	Weight	T	PAS	PAD	FC	TE	Diuresis	Marbling	TRC	SPO_2	FR	cyanosis	RALES	BC	Glasgow	Tonus	Aware	Convulsion	Vomiting	SM	EM	PR	Fracture	CE	FB	BH	Class
1	48	M	20	36.9	Normal	Normal	Bradycardie	Cold	Preserved	false	3	Normal	Normal	false	Normal	false	[12,15]	Normal	true	false	false	true	Normal	true	false	false	false	false	C3
2	164	M	50	37.5	Normal	Normal	Bradycardie	Cold	Preserved	false	3	Normal	Normal	false	Normal	false	< = 8	Hypo	true	false	false	false	false	true	false	false	false	false	C4
3	156	M	50	38.7	Hypertesion	Normal	Bradycardie	Hot	Preserved	false	ã3	Normal	Normal	false	Normal	false	< = 8	Hypo	False	False	False	false	false	False	True	True	False	True	C5
4	160	M	50	36.4	Hypertesion	Normal	Bradycardie	Normal	Preserved	false	ã3	Normal	DR-H	false	Normal	false	[12 - 15]	Normal	True	False	False	True	Normal	True	False	False	False	False	C2
5	64	F	18	37	Normal	HVS-H	Bradycardie	Normal	Preserved	false	ã3	Normal	Normal	false	Normal	false	[12 - 15]	Normal	True	False	True	True	Normal	True	False	True	False	False	C1

TABLE VII. Results Obtained by TBI-CDMA

Actual class of cases		Predicted class with relevant attributes	Predicted class with all attributes
Case 1	C3	C3	C4
Case 2	C4	C4	C5
Case 3	C5	C5	C5
Case 4	C2	C3	C0
Case 5	C1	C2	C3

isotonic saline injection. The doctors view these additional actions as preventive measures that can be applied to any patient admitted to the pediatric intensive care unit, regardless of their health status. Furthermore, the system empowers doctors to evaluate the proposed plan's suitability for a new case. If the plan is deemed appropriate, it can be directly applied; otherwise, modifications can be made as needed. This flexibility and control provided to the doctors enhance the system's usability and efficacy in real-world medical scenarios. Overall, the TBI-CDMA system has delivered encouraging results, with plans validated and approved by the doctors. Its ability to provide relevant and applicable plans, along with the option for doctor intervention and modification, underscores the potential of the system to enhance medical decision-making and improve patient care in pediatric intensive care units. Performance measurements can be used to assess the efficacy of our TBI-CDMA, we calculated the well-known performance and evaluation measures: the accuracy (equation 1), the precision (equation 2), the sensitivity (recall) (equation 3), and the specificity (equation 7) which are the common statistics to evaluate a model's performance. A comparison of the proposed approach (Approach 1) with other approaches was done (Table VIII):

- Approach 2: The authors of this approach [25] used the multi-label techniques to enhance the retrieval step. Their experiments were conducted using PTBI case base.
- Approach 3: in this approach, Informational gain is used for the selection of relevant attributes. Nine attributes are selected as relevant attributes for the similarity calculation, and then retrieval step by KNN has done.
- Approach 4: retrieval by KNN using all attributes.

TABLE VIII. Evaluation of the Proposed Approach

Measure	Approach 1	Approach 2	Approach 3	Approach 4
Precision	75.00	71.15	69.23	65.38
Sensitivity	75.00	71.15	69.23	65.38
Accuracy	91.66	90.38	89.74	88.46
Specificity	95.00	94.23	93.84	93.07

As depicted in Table VIII, the proposed approach (Approach 1) exhibits superior performance compared to the other three approaches. Approach 1 achieves an accuracy score of 91.66, precision and sensitivity of 75.00, and specificity of 95.00, outperforming the alternatives. Approaches 3 and 4 also demonstrate good performances, while Approach 1 fares the worst for all measures.

The performance of the proposed model was extensively evaluated using seven medical datasets obtained from the UCI Machine Learning Repository [50]. These datasets are well-regarded for their application in benchmarking and real-world data collection within the CBR and data mining community. Our model was applied to these datasets to validate its effectiveness. Each medical dataset was specified with its attributes and instances. Table IX presents comprehensive details about the datasets, including accuracy and Roc area results, showcasing the performance of our model when applied to these datasets.

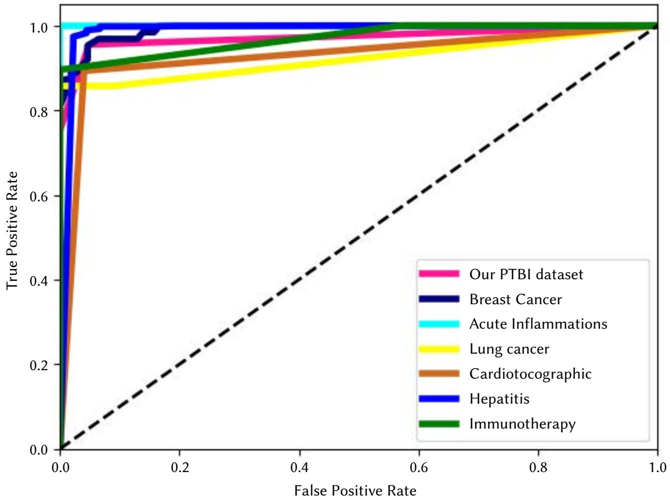

Fig. 4. ROC Curves.

As demonstrated in Table IX, our model that integrates the decision tree in the retrieval step to ameliorate this step and the CBR process has achieved remarkable levels of accuracy, surpassing 90%, when applied to the medical datasets. Furthermore, the ROC area for all datasets is notably high, as depicted in Fig. 4. Our model exhibits the highest classifier performance not only for our specific dataset but also for other publicly available datasets, thus validating its efficacy and reliability.

These findings suggest that employing pure CBR alone may not be sufficient to design a Clinical Decision Support System (CDSS) with optimal performance. Instead, integrating data mining (DM) techniques in the retrieval phase can enhance the overall performance of a CBR-based system. The incorporation of relevant attributes revealed by the decision tree in TBI-CDMA contributes to improved retrieval and surpasses the traditional CBR approach, enhancing its potential as an effective tool for assisting doctors in making critical decisions regarding the management of pediatric patients involved in road accidents.

V. Conclusion

In this work, the authors proposed a novel Clinical Decision Making Approach (CDMA) known as TBI-CDMA, specifically designed for the care of children victims of a road accident, from the site of the accident until their arrival in the pediatric intensive care unit. The authors applied their approach to a real dataset (PTBI) elaborated by them. The proposed TBI-CDMA consisted of a series of multistep procedures, including data collection, preprocessing, selection of relevant attributes, and Case-Based Reasoning (CBR). Real-world datasets often contain incomplete, noisy, and inconsistent data, which can obscure valuable patterns. Hence, the preprocessing step plays a crucial role in generating high-quality data and enhancing the accuracy and efficiency of the models [51]. Within TBI-CDMA, data preprocessing assumed a fundamental role with a significant impact on its accuracy. it involves various processes such as data cleaning, discretization, statistical analysis, and classification. These steps aimed to prove the PTBI dataset, making it more suitable for the subsequent stages of the decision-making process, and ensuring the system's overall effectiveness in providing appropriate and timely care for the child victim of road accident. In the statistical analysis step, an evaluation of the correlation of the features (attributes) with the target was done; and in the classification step, we evaluated some classifiers on our dataset. The subsequent step involved integrating DM techniques with CBR, especially in the retrieval step. Medical databases are known to be voluminous, and each case can contain numerous attributes. While some attributes are crucial, others may be unnecessary and lead to increased complexity during case retrieval. To ensure optimal performance of a CBR system, maintaining the content of the case base is essential. One proposed solution to address these issues is attribute selection. By identifying and retaining only relevant attributes, the selection of relevant attributes aims to reduce the size of the case base and minimize retrieval time which improved the performance of this step. This approach was tested using other datasets and yielded good results.

Based on the obtained results, the preprocessing step successfully validated the reliability of our PTBI dataset and identified attributes strongly correlated with the treatment plan applied by doctors. This finding further confirms the significant reduction of attributes achieved through the decision tree. Also, our approach demonstrates that utilizing only relevant attributes in similarity calculations generates accurate plans endorsed by doctors, resulting in superior performance compared to other techniques. Overall, the innovative TBI-CDMA approach proved highly effective in preprocessing and validating the PTBI dataset, resulting in faster and enhanced retrieval and medical management, vital for time-sensitive decision-making scenarios. This significant advancement holds great promise for improving the care of child victims of road accidents.

A limitation of the CDMA in relation to the PTBI dataset is that the number of cases collected is reduced, resulting in the dataset not representing all of the signs, symptoms, and gestures applied to a child who has been injured. Once we have more cases, we may be able to improve the model's predictive performance. On another level, there is still work to be done at the core of the model. Indeed, this project did not take into consideration the adaptation of the solutions and was satisfied with a manual adaptation made by the doctors. It is legitimate to think about an automatic or semi-automatic adaptation once the solution is deployed. knowing that in this step, the system evaluates the degree of similarity between the selected cases and the current case, providing the doctors with the most relevant cases along with their respective degrees of similarity. The doctor then chooses the case that they find most appropriate or may make modifications to their solution if they deem it necessary. The idea is to record the modifications made by the expert on the proposed solution in order to generate rules that capitalize on this expertise. This can only be done when the solution is used over a sufficiently long period.

References

[1] T. Thim, N. H. V. Krarup, E. L. Grove, C. V. Rohde, B. Løfgren, "Initial assessment and treatment with the airway, breathing, circulation, disability, exposure (abcde) approach," *International journal of general medicine*, vol. 5, p. 117, 2012.

[2] P. Andritsos, I. Jurisica, J. I. Glasgow, "Case-based reasoning for biomedical informatics and medicine," *Springer Handbook of Bio-/Neuroinformatics*, pp. 207–221, 2014.

[3] S. S. R. Abidi, S. Manickam, "Leveraging xml-based electronic medical records to extract experiential clinical knowledge: An automated approach to generate cases for medical case-based reasoning systems," *International Journal of Medical Informatics*, vol. 68, no. 1-3, pp. 187–203, 2002.

[4] K. C. A. D. Borges, I. F. de Barcelos Tronto, R. de Aquino Lopes, J. D. S. da Silva, "A methodology for preprocessing data for application of case based reasoning," in *2012 XXXVIII Conferencia Latinoamericana En Informatica (CLEI)*, 2012, pp. 1–8, IEEE.

[5] N. Choudhury, S. A. Begum, "A survey on case- based reasoning in medicine," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 8, pp. 136–144, 2016.

[6] F. Saadi, B. Atmani, F. Henni, "Improving retrieval performance of case based reasoning systems by fuzzy clustering," *International Journal of*

Interactive Multimedia and Artificial Intelligence, (2023), doi: http://dx.doi.org/10.9781/ijimai.2023.07.002.

[7] Y. Guo, J. Hu, Y. Peng, “Research of new strategies for improving cbr system,” *Artificial Intelligence Review*, vol. 42, pp. 1–20, 2014.

[8] R. Bellazzi, B. Zupan, “Predictive data mining in clinical medicine: current issues and guidelines,” *International journal of medical informatics*, vol. 77, no. 2, pp. 81–97, 2008.

[9] F. Saadi, B. Atmani, F. Henni, “Integration of datamining techniques into the cbr cycle to predict the result of immunotherapy treatment,” in *2019 International Conference on Computer and Information Sciences (ICCIS)*, 2019, pp. 1–5, IEEE.

[10] B. N. Barigou, F. Barigou, C. Benchehida, B. Atmani, G. Belalem, “The design of a cloud-based clinical decision support system prototype: management of drugs intoxications in childhood,” in *Research Anthology on Decision Support Systems and Decision Management in Healthcare, Business, and Engineering*, IGI Global, 2021, pp. 387–409.

[11] L. J. Hoeksema, A. Bazzy-Asaad, E. A. Lomotan, D. E. Edmonds, G. Ramírez-Garnica, R. N. Shiffman, L. I. Horwitz, “Accuracy of a computerized clinical decision-support system for asthma assessment and management,” *Journal of the American Medical Informatics Association*, vol. 18, no. 3, pp. 243–250, 2011.

[12] S. Walczak, S. R. Okuboyejo, “An artificial neural network classification of prescription nonadherence,” *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, vol. 12, no. 1, pp. 1–13, 2017.

[13] E. K. Hashi, M. S. U. Zaman, M. R. Hasan, “An expert clinical decision support system to predict disease using classification techniques,” in *2017 International conference on electrical, computer and communication engineering (ECCE)*, 2017, pp. 396–400, IEEE.

[14] Y.-S. Chang, C.-T. Fan, W.-T. Lo, W.-C. Hung, S.-M. Yuan, “Mobile cloud-based depression diagnosis using an ontology and a bayesian network,” *Future Generation Computer Systems*, vol. 43, pp. 87–98, 2015.

[15] F. Z. Benhacine, B. Atmani, F. Z. Abdelouhab, “Contribution to the association rules visualization for decision support: A combined use between boolean modeling and the colored 2d matrix,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 38–48, 2019.

[16] E. R. Bareiss, B. W. Porter, C. C. Wier, “Protos: An exemplar-based learning apprentice,” in *Machine learning*, Elsevier, 1990, pp. 112–127.

[17] C. Cândea, G. Cândea, Z. B. Constantin, “Ardocare - a collaborative medical decision support system,” *Procedia Computer Science*, vol. 162, pp. 762–769, 2019.

[18] V. Tang, P. K. Y. Siu, K. L. Choy, H. Y. Lam, G. T. S. Ho, C. K. M. Lee, Y. P. Tsang, “An adaptive clinical decision support system for serving the elderly with chronic diseases in healthcare industry,” *Expert Systems*, vol. 36, no. 2, p. e12369, 2019.

[19] D. S. Kumar, G. Sathyadevi, S. Sivanesh, “Decision support system for medical diagnosis using data mining,” *International Journal of Computer Science Issues (IJCSI)*, vol. 8, no. 3, p. 147, 2011.

[20] E. Alickovic, A. Subasi, “Medical decision support system for diagnosis of heart arrhythmia using dwt and random forests classifier,” *Journal of medical systems*, vol. 40, no. 4, pp. 1–12, 2016.

[21] A. Yadollahpour, J. Nourozi, S. A. Mirbagheri, E. Simancas-Acevedo, F. R. Trejo-Macotela, “Designing and implementing an anfis based medical decision support system to predict chronic kidney disease progression,” *Frontiers in physiology*, p. 1753, 2018.

[22] F. Hamedan, A. Orooji, H. Sanadgol, A. Sheikhtaheri, “Clinical decision support system to predict chronic kidney disease: A fuzzy expert system approach,” *International journal of medical informatics*, vol. 138, p. 104134, 2020.

[23] Y. Guo, K. Wu, “Research on case retrieval of bayesian network under big data,” *Data & Knowledge Engineering*, vol. 118, pp. 1–13, 2018.

[24] A. Mansoul, B. Atmani, “Combining multi-criteria analysis with cbr for medical decision support,” *Journal of Information Processing Systems*, vol. 13, no. 6, pp. 1496– 1515, 2017.

[25] H. Benfriha, B. Atmani, F. Barigou, F. Henni, B. Khemliche, S. Fatima, A. Douah, Z. Z. Addou, “Improving cbr retrieval process through multilabel text categorization for health care of childhood traumatic brain injuries in road accident,” in *Proceedings of Sixth International Congress on Information and Communication Technology*, 2022, pp. 721–731, Springer.

[26] A. Mansoul, B. Atmani, “Clustering to enhance case- based reasoning,” in *Modelling and Implementation of Complex Systems*, Springer, 2016, pp. 137–151.

[27] N. Malviya, N. Choudhary, K. Jain, “Content based medical image retrieval and clustering based segmentation to diagnose lung cancer,” *Advances in Computational Sciences and Technology*, vol. 10, no. 6, pp. 1577–1594, 2017.

[28] F. Saadi, B. Atmani, F. Henni, “Integration of fuzzy clustering into the case base reasoning for the prediction of response to immunotherapy treatment,” in *Mediterranean Conference on Pattern Recognition and Artificial Intelligence*, 2019, pp. 192–206, Springer.

[29] M. Benamina, B. Atmani, S. Benbelkacem, “Diabetes diagnosis by case-based reasoning and fuzzy logic,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 3, pp. 72–81, 2018.

[30] M. A. Hall, “Correlation-based feature selection for machine learning,” 1999.

[31] H. Feuillâtre, V. Auffret, M. Castro, H. Le Breton, M. Garreau, P. Haigron, “Study of similarity measures for case-based reasoning in transcatheter aortic valve implantation,” in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4, IEEE.

[32] J. Jarmulak, S. Craw, R. Rowe, “Genetic algorithms to optimise cbr retrieval,” in *European Workshop on Advances in Case-Based Reasoning*, 2000, pp. 136–147, Springer.

[33] S. B. Ayed, Z. Elouedi, E. Lefevre, “Managing uncertainty during cbr systems vocabulary maintenance using relational evidential c-means,” in *2020 International Multi-Conference on:“Organization of Knowledge and Advanced Technologies”(OCTA)*, 2020, pp. 1–10, IEEE.

[34] E. G. Addisu, A. S. Boltena, S. Y. Amare, “Case-based reasoning framework for malaria diagnosis,” *I. J. Information Technology and Computer Science*, vol. 6, pp. 31-48, 2020.

[35] F. Henni, B. Atmani, F. Atmani, F. Saadi, “Improving coronary artery disease prediction: Use of random forest, feature importance and case-based reasoning,” *International Journal of Decision Support System Technology (IJDSST)*, vol. 15, no. 1, pp. 1–17, 2023.

[36] M. D. Von Atzingen, D. R. C. Schmidt, E. A. P. M. Nonino, “Elaboration and application of an evaluation instrument in the immediate postoperative period, based on the advanced trauma life support protocol,” *Acta Paulista de Enfermagem*, vol. 21, pp. 616–623, 2008.

[37] U. Stańczyk, B. Zielosko, G. Baron, “Discretisation of conditions in decision rules induced for continuous data,” *PloS one*, vol. 15, no. 4, p. e0231788, 2020.

[38] F. Wilcoxon, “Individual comparisons by ranking methods,” in *Breakthroughs in statistics*, Springer, 1992, pp. 196–202.

[39] A. Smiti, Z. Elouedi, “Using clustering for maintaining case based reasoning systems,” in *2013 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO)*, 2013, pp. 1–6, IEEE.

[40] H. Sun, X. Hu, “Attribute selection for decision tree learning with class constraint,” *Chemometrics and Intelligent Laboratory Systems*, vol. 163, pp. 16–23, 2017.

[41] J. R. Quinlan, “Induction of decision trees,” *Machine learning*, vol. 1, pp. 81–106, 1986.

[42] A. G. Karegowda, A. Manjunath, M. Jayaram, “Comparative study of attribute selection using gain ratio and correlation based feature selection,” *International Journal of Information Technology and Knowledge Management*, vol. 2, no. 2, pp. 271–277, 2010.

[43] J. K. Han, M. Kamber, “Data mining: Concepts and techniques,” Morgan Kaufmann, 2001.

[44] I. Ragnemalm, “The euclidean distance transform in arbitrary dimensions,” *Pattern Recognition Letters*, vol. 14, no. 11, pp. 883–888, 1993.

[45] J. Ahn, S.-H. Ji, S. J. Ahn, M. Park, H.-S. Lee, N. Kwon, E.-B. Lee, Y. Kim, “Performance evaluation of normalization-based cbr models for improving construction cost estimation,” *Automation in Construction*, vol. 119, p. 103329, 2020.

[46] M. T. Rezvan, A. Z. Hamadani, A. Shalbafzadeh, “Case-based reasoning for classification in the mixed data sets employing the compound distance methods,” *Engineering Applications of Artificial Intelligence*, vol. 26, no. 9, pp. 2001–2009, 2013.

[47] A. Idri, A. Abran, T. Khoshgoftaar, “Fuzzy analogy: A new approach for software cost estimation,” in *International Workshop on Software Measurement*, 2001, pp. 28–29, Citeseer.

[48] D. Chicco, G. Jurman, “Machine learning can predict survival of patients

with heart failure from serum creatinine and ejection fraction alone," *BMC medical informatics and decision making*, vol. 20, no. 1, pp. 1–16, 2020.

[49] F. Henni, *Composition Dynamique de Services Web par Apprentissage Artificiel*. PhD dissertation, Université Oran 1, 2015.

[50] C. B. D. Newman, C. Merz, "UCI repository of machine learning databases," 1998. [Online]. Available: http://www.ics.uci.edu/„mlearn/MLRepository.html.

[51] S. Zhang, C. Zhang, Q. Yang, "Data preparation for data mining," *Applied artificial intelligence*, vol. 17, no. 5- 6, pp. 375–381, 2003.

Fatima Saadi

Fatima Saadi is currently a PhD candidate at the University of Oran 1 and affiliated researcher in Laboratoire d'Informatique d'Oran, Algeria. Her research interests include data mining, case-based reasoning,information retrieval, medical decision support systems, machine learning.

Baghdad Atmani

Baghdad Atmani is currently a Full Professor in Computer Science. His field of interest is artificial intelligence and machine learning. His research is based on knowledge representation, knowledge-based systems, CBR, data mining, expert systems, decision support systems and fuzzy logic.

Fouad Henni

Fouad Henni is a teaching researcher in computer science at Mostaganem university, Algeria. His research interests are in the areas of semantic Web services, artificial intelligence, case-based reasoning and deep learning, with a particular emphasis on applications in medical diagnosis. He is member of the Computer Science and new Technologies Lab (CSTL).

Hichem Benfriha

Hichem Benfriha is a computer science teacher in the Department of Technical Sciences, University of Mascara Mustapha Stambouli, Algeria. He is currently a Research Member of Laboratory of Computer Science of Oran. He is currently a PhD candidate in the Computer Science Department of Oran 1 University (Algeria). He received his Master of Science degree in 2012 from the same university. His research interests focus on CBR, data Mining, text mining, information extraction, information retrieval, natural language processing, machine learning and Multi-label classification areas.

Zakaria Zoheir Addou

Zakaria Zoheir Addou is a pediatric anesthesiologist and critical care physician since 2002 at pediatric hospital in Oran, Algeria. He received his PhD in 2017 at the university of Ahmed Ben bella 1 Oran in medical science. He is also a member of laboratory of pediatric injury in the University of Oran. His field of research is safety of anesthesia outside operating room in children.

Rabah Guerbouz

Rabah Guerbouz is a neurosurgeon at the University Hospital of Oran, Algeria since 2007. He obtained his doctorate in 2018 at the University Ahmed Ben Bella 1 Oran in medical sciences. He is a teacher at the national paramedical school.