



Retrieving similar cases for clinical decision support in the context of revascularization of lower limbs

Margaux Roux ^{a,*}, Rafaëlle Spear ^{a,b}, Céline Fouard ^{a,*,*}, Pascal Haigron ^c

^a Université Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, Grenoble, 38000, France

^b Centre Hospitalier Universitaire Grenoble Alpes, Grenoble, 38000, France

^c Univ Rennes, Inserm, LTSI – UMR 1099, Rennes, F-35000, France

ARTICLE INFO

Keywords:

Case-based reasoning
Revascularization of lower limbs
Clinical case description
Similarity measure

ABSTRACT

Background: Vascular surgeons face complex decisions regarding revascularization of lower limbs of patients, for which there is no decision support solution yet available.

Objective: Relying on Case-Based Reasoning approach, this paper addresses the issue of case retrieving in the context of lower limb revascularization. To this end we propose a similarity measure based on relevant clinical attributes selection and weighting. We analyze its performance for clinical decision support.

Method: Retrieval process is used to anticipate the success or failure of a revascularization strategy for a new patient based on the surgical procedure outcome of similar past patient cases. Using clinical knowledge in Global vascular Guidelines, cases were structured and formalized. They were described by several attributes reflecting the patient risk estimation, limb pathology staging, localization of atherosclerotic lesions, vascular history and revascularization type. To construct the similarity measure for lower limb revascularization CBR, three distance metrics (Euclidean, Cosine and Heterogeneous), three attributes selection scenarios as well as two weighting strategies were identified and compared.

Results: A case base of 17 patients was built and used to assess the relevance of the retrieval process when predicting the long-term outcome of the revascularization procedure. Similarity measure showed good performance and clinical relevance when using a heterogeneous distance with attributes and weights derived from information in official guidelines, vascular history and type of revascularization.

Conclusion: By investigating and validating the similar cases retrieving process, we showed that CBR appears to be a promising approach for clinical decision support in the context of lower extremity arterial disease.

1. Introduction

The Lower Extremities Arterial Disease (LEAD) due to atherosclerosis, may lead to rest pain and ischemic wounds. The major risk of LEAD is limb amputation. To prevent this evolution, revascularization procedures of lower limb arteries may be required. Two main revascularization surgeries are performed: endovascular surgery and open surgery. These two strategies can be combined during the same procedure. The main issue for the vascular surgeon is to perform the best strategy of revascularization for the patient to be treated adapted to its symptomatology. A successful revascularization will reduce pain, aid wounds healing and avoid amputation. LEAD patients are frail because of their advanced age and multiple comorbidities. Consequently, surgeons need to evaluate in their decision-making scenarios the trade-off between the invasiveness of revascularization and the long-term benefit

adapted to each patient. To our knowledge, there is no clinical decision support system dealing with revascularization of lower limbs.

Clinical Decision Support Systems (CDSS) raised interest in the 1990's [1,2] when Electronic Health Records (EHR) appeared with the objective of personalized care. Sutton et al. [1] define CDSS as system intended to improve healthcare delivery by enhancing medical decisions with targeted clinical knowledge, patient demographics and personal health information. Several methodologies have been developed and are currently employed in the field of CDSS [3] such as rule-based, knowledge-based, Bayesian network, and neural network approaches. However, some studies have indicated that clinicians may exhibit limited confidence in the latter method due to its perceived "black box" effect [4]. Additionally, classical machine learning techniques necessitate the availability of a substantial dataset for training purposes. Case Based Reasoning (CBR) is a knowledge-based method that overcomes

* Corresponding author.

E-mail addresses: margaux.roux@univ-grenoble-alpes.fr (M. Roux), celine.fouard@univ-grenoble-alpes.fr (C. Fouard).

<https://doi.org/10.1016/j.ijmedinf.2025.105931>

Received 20 November 2024; Received in revised form 28 February 2025; Accepted 14 April 2025

this challenge as it is close to physician reasoning [5]: the algorithm retrieves the most similar past cases to determine the solution to a new patient. Indeed, it is assumed in CBR that “patients with similar symptoms will likely receive the same treatment with similar outcome”. To identify similar cases, CBR relies on similarity measures, including selected attributes that are weighted and distance metrics between cases. Similarity measures in clinical field have been widely studied [6,7], using one type of data (time-series, images or numerical data). Multiplicity of data types is often encountered with clinical data and it is the case for LEAD context. This problem has been addressed in different clinical domains [8–10]. Using a CBR approach, this paper addresses the issue of case retrieving in the context of lower limb revascularization. To this end we propose a similarity measure based on relevant clinical attributes selection and weighting. We analyze its performance for clinical decision support.

This paper is organized as follows. In the ‘Method’ section the selection of attributes, the estimation of their relative importance and the data formalization to build the case base are described. It also introduces the Similarity Measure used to retrieve similar patients. In the Results section, the evaluation of similarity measure is presented in terms of attribute selection, weighting and distance metrics used. Then, the retrieval of similar cases for three patients is described in details.

2. Methods

To find a solution to a new case, the Case Based Reasoning (CBR) algorithm relies on four steps applied to the case base: Retrieve the most similar case(s), Reuse them, Revise if necessary their solution and Retain the new case. A case, noted as $C = \{A, S\}$, withholds three elements: $A = [a_1, \dots, a_n]$ the list of attributes i.e. descriptors of the problem where n is the number of attributes, S the solution to solve the problem and R the result of the problem resolution (success or failure of S). Similarity between cases can be defined as the degree of resemblance between two problem descriptions. The similarity is measured between the attributes of the new problem, C_q called query case, and those of the solved cases C_c in the case base CB . It can be quantified in a variety of ways [5].

Selection of attributes and the data formalization are first described. In a second part, we define how the similarity between patients can be measured using either a distance metric (homogeneous similarity measures) or a combination of distance metrics (heterogeneous similarity metrics). We then present the evaluation method of similarity measures using different associations and weights of attributes.

2.1. Attributes selection and case formalization

Attribute selection depends deeply on the solution expected from the CBR. In this study, we focus on the outcome after surgery which can be useful to vascular surgeons to decide about the appropriate revascularization type. Accordingly, the user tests a revascularization strategy to predict its clinical outcome.

The similarity measure compares the lists of attributes A_{C_q} and A_{C_c} , C_c being a past solved case from the case base. To select the relevant attributes with regards to decision making in revascularization procedures, we use the official clinical recommendations published by the Society of Vascular Surgery (SVS). In particular, the Global Vascular Guidelines [11] propose a three-step integrated approach to improve decision-making, clinical outcomes and cost-effectiveness in the treatment of Chronic Limb-Threatening Ischaemia (CLTI). The PLAN concept is based on the *Patient risk estimation*, the *Limb staging* and the *ANatomic pattern of disease*. The PLAN concept is organized as shown in Fig. 1.

The ‘Patient Risk Estimation’ step, relies on clinical scales evaluating the frailty of patients. We used in our study the Vascular Quality Initiative (VQI) [12]. This questionnaire assesses the 30 perioperative days survival and up to 2 years survival. The attributes available in the retrospective database corresponding to the ‘Patient Risk Estimation’ step are shown in Table 1.

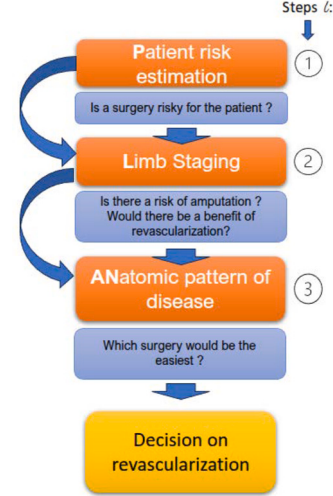


Fig. 1. Representation of the 3 steps of the PLAN concept.

Table 1

Predictors of Survival of the ‘Patient Risk Estimation’ step.

	Attributes Name	Type	Domain
a_1	Age	Numeric	integer
a_2	Rest Pain or Tissue Loss	Binary	[0,1]
a_3	Smoking status	Ordinal	[Never, Former, Current]
a_4	Hypertension	Binary	[0,1]
a_5	Ischemic Cardiopathy	Binary	[0,1]
a_6	Diabetes	Ordinal	[No, Yes, Yes with complexity]
a_7	Chronic Obstructive Pulmonary Disease (COPD)	Binary	[0,1]
a_8	Chronic Kidney Disease stage	Ordinal	[1,2,3,4,5]
a_9	Ambulatory Status	Ordinal	[Autonomous, Cane or Walker, Wheel Chair, Bed bound]
a_{10}	Anti-hypertensive, statines, anticoagulants)	Binary	[0,1]
a_{11}	Statines	Binary	[0,1]
a_{12}	Anticoagulants	Binary	[0,1]

Table 2

Attributes of ‘Limb Staging’.

	Attributes	Type	Domain
a_{13}	Wound grade	Ordinal	[0,1,2,3]
a_{14}	Ischemia grade	Ordinal	[0,1,2,3]
a_{15}	foot Infection grade	Ordinal	[0,1,2,3]

The ‘Limb staging’ step requires the evaluation of the Wifl score [13]. This score entails measurements of the Wound, the severity of the Ischemia, and the stage of the foot Infection. Each of these three elements is assigned a grade from 0 to 3 (grade 3 being the most severe). They are used as attributes in the case base (Table 2).

The ‘ANatomic pattern of disease’ step evaluates the complexity of revascularization specifically in endovascular technique, which relies on anatomic characteristics of the patient and the lesion. As both open and endovascular techniques are evaluated, our case description focuses on the lesion localization (Table 3).

We define L_{PLAN} , a first list of attributes, that includes all attributes given by the PLAN concept and shown in Tables 1, 2 and 3:

$$L_{PLAN} = [a_1, \dots, a_n] \text{ with } n = 18. \quad (1)$$

Existing guidelines may not be exhaustive and we identified attributes based on clinical expertise unlisted previously. They are related

Table 3

Attributes for 'Anatomic Pattern of disease'.

	Attributes	Type	Domain
a_{16}	Femoropopliteal localization	Binary	[0,1]
a_{17}	Infrapopliteal localization	Binary	[0,1]
a_{18}	Upstream of the SFA	Binary	[0,1]

Table 4

Additional attributes.

	Attributes	Type	Domain
a_{19}	Dyslipidemia	Binary	[0,1]
a_{20}	Rhythmic cardiopathy	Binary	[0,1]
a_{21}	Overweight and Obesity	Ordinal	[No, Overweight, Obesity]
a_{22}	Clinical Frailty Score	Ordinal	[1,2,3,4,5,6,7,8,9]
a_{23}	Previous homolateral revascularization	Ordinal	[No, Endovascular, Open Surgery, Both]
a_{24}	Previous controlateral revascularization	Ordinal	[No, Endovascular, Open Surgery, Both]
a_{25}	Previous amputation on homolateral limb	Binary	[0,1]
a_{26}	Previous amputation on controlateral limb	Binary	[0,1]

to patient's previous vascular surgery history, body mass index, dyslipidemia, and general frailty. A survey was conducted among four vascular surgeons of the same hospital with varying levels of experience. The detailed results are presented in Appendix A, Fig. A.1. Table 4 summarizes additional attributes.

A second list of attributes, $L_{PLAN+SUPP}$, is defined as follows:

$$L_{PLAN+SUPP} = [a_1, \dots, a_n] \text{ with } n = 26. \quad (2)$$

The clinical decision is based on the patient and medical history. Since, the technique chosen (endovascular or open surgery) has an impact on the postoperative success, the Revascularization Type (RT) has also to be considered as an attribute. The addition of RT to the second list of attributes creates a third list of attributes $L_{PLAN+SUPP+RT}$:

$$L_{PLAN+SUPP+RT} = [a_1, \dots, a_n] \text{ with } n = 27, \quad (3)$$

with a_{27} the attribute 'Revascularization Type' (endovascular or open surgery).

To sum-up, a past solved case is expressed as follows: $C_c = \{A_c, S_c\}$ and a query case $C_q = \{A_q\}$, with S the outcome of surgery and A the list of attributes which can be one of the three following options:

- $A_1 = L_{PLAN}$
- $A_2 = L_{PLAN+SUPP}$
- $A_3 = L_{PLAN+SUPP+RT}$

The solution S_q i.e. what is to be predicted, is the clinical success of the revascularization procedure. Clinical success is defined as the absence of homolateral reintervention, the absence of a recommendation for an homolateral amputation and perioperative survival. The occurrence of one of these events is considered a failure of the procedure. In this study, the clinical outcome is observed 12 months after surgery. It is represented as a binary variable categorized as either success or failure. S_q is deduced from clinical outcome of retrieved similar patients C_r . At the end of a CBR cycle, the query case is completed $C_q = \{A_q, S_q\}$ and can be included in the case base or not (Retain step).

The case base used in this work is built from a retrospective database of revascularization procedures performed between 2018 and 2020. This dataset was collected from medical records [14] including examination reports and results, consultation letters, hospitalization follow-up reports, and prescriptions. Clinical information was formalized and trans-

formed into structured attributes of Tables 1, 2, 3 and 4, ultimately generating a CSV file. Patients with missing data are excluded from the database, allowing for a complete case analysis.

As a result, the case base includes 17 complete patients, 10 men and 7 women, with a range of ages between 78 and 95 years old. The total of variables is 75, with 34 attributes describing the query patient. Of the 17 patients, 15 underwent endovascular surgery and 2 underwent open surgery. There is a proportion of 52% success 12 months after revascularization (9 patients) in our case base.

2.2. Similarity measure

Similarity is the function used for retrieving cases. If a case C_x is more similar to the new case than a case C_y , then the case C_x is more useful for the solution to apply. Similarity measure is derived from dissimilarity, which is evaluated with weighted distance metrics.

$$Sim(C_q, C_c) = 1 - diss(C_q, C_c) \quad (4)$$

where $Sim(C_q, C_c)$ represents similarity between a query case C_q and a past solved case C_c from the Case Base CB . $diss(C_q, C_c)$ is the dissimilarity between the cases. Distance metrics usually considered in case of homogeneous data are Euclidean (Equation (5)) and Cosine distance (6) expressed as follows:

$$diss_{Euclidean} = \sqrt{\sum_{i=1}^n w_{l_{a_i}} \times (a_{q,i} - a_{c,i})^2} \quad (5)$$

$$diss_{Cosine} = 1 - \frac{\sum_{i=1}^n w_{l_{a_i}} a_{q,i} a_{c,i}}{\sqrt{\sum_{i=1}^n w_{l_{a_i}} a_{q,i}^2} \times \sqrt{\sum_{i=1}^n w_{l_{a_i}} a_{c,i}^2}} \quad (6)$$

with $a_{q,i}$ and $a_{c,i}$ respectively the i^{th} attributes value of the query case C_q and past solved case C_c , n the total number of attributes and $w_{l_{a_i}}$ weight of attributes defined in Equation (12).

In lower limb revascularization context, the clinical data is heterogeneous, encompassing not only numerical data (e.g. age) but also ordinal data (e.g. chronic kidney disease stage) and binary data (e.g. Hypertension). Heterogeneous distance metrics have been developed for this purpose [8,15]. In this work we used an heterogeneous weighted similarity measure constructed from different distance metrics depending on the type of attributes [10]:

$$diss_{Heterogeneous}(C_q, C_c) = \frac{\sum_{i=1}^n w_{l_{a_i}} d(a_{q,i}, a_{c,i})}{\sum_i^n w_{a_i}} \quad (7)$$

where $d(a_{q,i}, a_{c,i})$ the distance between attributes a_i for cases C_q and C_c .

$$d(a_{q,i}, a_{c,i}) = \begin{cases} d_{quantitative}(a_{q,i}, a_{c,i}), & \text{if } a_i \text{ is quantitative} \\ d_{binary}(a_{q,i}, a_{c,i}), & \text{if } a_i \text{ is binary} \\ d_{ordinal}(a_{q,i}, a_{c,i}), & \text{if } a_i \text{ is ordinal} \end{cases} \quad (8)$$

with

$$d_{quantitative}(a_{q,i}, a_{c,i}) = \frac{|a_{q,i} - a_{c,i}|}{range(a_i)} \quad (9)$$

$$d_{binary}(a_{q,i}, a_{c,i}) = \begin{cases} 0, & \text{if } a_{q,i} = a_{c,i} \\ 1, & \text{if } a_{q,i} \neq a_{c,i} \end{cases} \text{ for } a_{q,i}, a_{c,i} \in [0, 1] \quad (10)$$

$$d_{ordinal}(a_{q,i}, a_{c,i}) = O[a_{q,i}][a_{c,i}] \quad (11)$$

$d_{ordinal}$ relies on the definition of a distance matrix for each ordinal attribute. The definition of each distance matrix consists in translating the difference between two ordinal values into a distance (i.e. a numerical value), based on vascular surgeons expertise. The distance matrices for ordinal attributes are reported in the Appendix B.

There are several possibilities, to assign weights to attributes that reflect their relative importance. In this work we consider two weighting

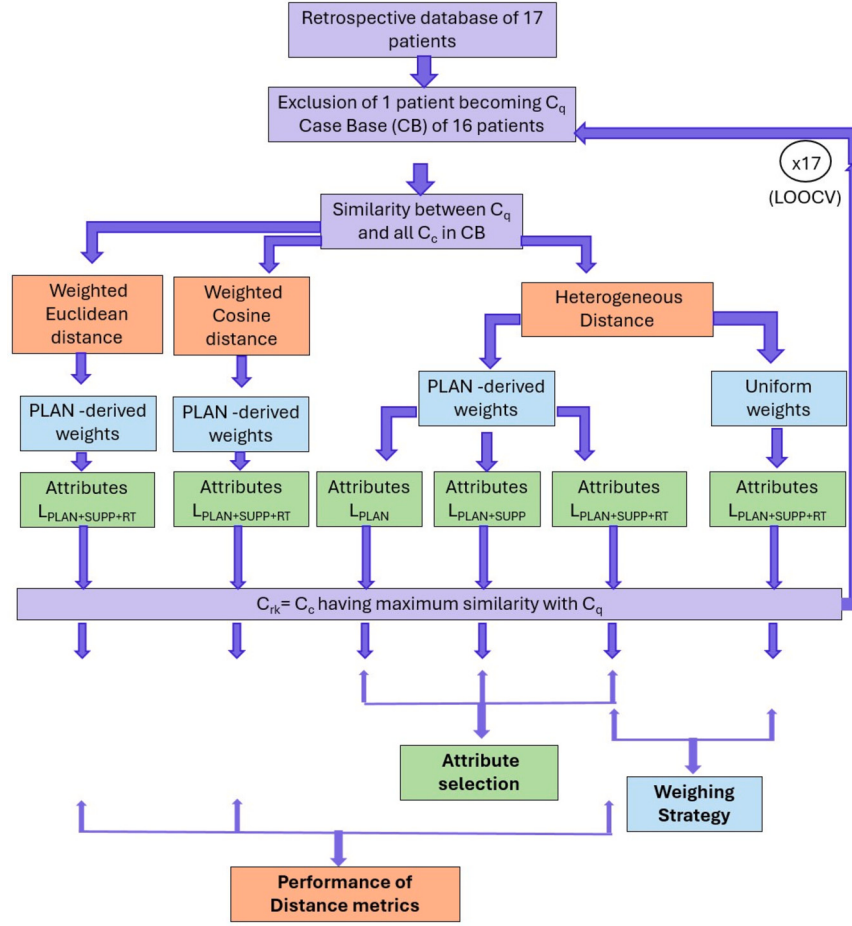


Fig. 2. Flow chart of the evaluation of the different test scenarios.

Algorithm 1 Case Retrieval Algorithm.

Input: *CaseBase* (table containing previous solved cases $C_c = \{A_c, S_c\}$), C_q (vector containing the attributes of the query case $C_q = \{A_q\}$), k (number of similar cases to retrieve), type of dissimilarity (Euclidean, Cosine or Heterogeneous)
Output: Retrieved_Cases: list of the k most similar cases $C_r = \{A_r, S_r\}$ with their solution (surgery outcome)
Initialize an empty table *SimilarityTable*
Initialize an empty list *WeightsList*
Assign *WeightsList* with weights w for each attributes (Equation (12))
for each case $C_c \in \text{CaseBase}$ **do**
 $\text{Dissimilarity} \leftarrow \text{dissimilarity}(C_q, C_c, \text{WeightsList})$ (Equation (5), (6) or 7)
 $\text{SimTable}[C_c] \leftarrow 1 - \text{Dissimilarity}$ (Equation (4))
end for
Sort *SimTable* in descending order based on similarity scores
Retrieved_Patients $\leftarrow$ top k cases from *SimTable*
return Retrieved_Patients

approaches: uniform weights and weights derived from clinical knowledge (PLAN concept). We assume that the importance of an attribute depends on the step in which it is considered in the PLAN concept [11]. Each step l (Fig. 1) is assigned a weight w_l and all attributes within a given step are assigned the step weight. The weight $w_{l_{a_i}}$ (Eq. (12)) of an attribute a_i of the level l of the clinical decision tree is simply defined as follows:

$$w_l = l/L \quad (12)$$

with L the total number steps of the PLAN concept.

As a result, the highest weights were assigned to the complexity of the 'ANatomic pattern of disease' attributes in Table 3 ($w_1 = 1$), followed by the 'Limb staging' attributes in Table 2 ($w_2 = 2/3$) and then

the 'Patient Risk Estimation' attributes in Table 1 ($w_2 = 1/3$). This is consistent with [11] which states that the Anatomical Pattern of Disease is a dominant element for evidence-based revascularization. The additional attributes (Table 4) were considered as part of 'Patient Risk Estimation' step and thus inherited the weights of the attributes of this step. The Revascularization Type is assigned the weight of the 'ANatomic pattern of disease' step.

The retrieval process presented in this section is summarized by the Algorithm 1. The dissimilarity function applied is one of the Equations (5), (6), and (8). This algorithm was implemented in Python 3.6 on a standard Linux PC. The retrieval of most similar cases takes less than a second. It is therefore compatible with clinical use.

2.3. Evaluation

Similarity measures defined using different lists of attributes L_{PLAN} , $L_{PLAN+SUPP}$ and $L_{PLAN+SUPP+RT}$, with uniform or knowledge-derived weights, and with the use of different distance metrics (Equation (4), (5) and (8)) are evaluated. The evaluation flow chart is shown in Fig. 2. The leave-one-out method is employed: at each iteration, a patient is excluded from the case base and considered as the query patient C_q . Even if it is not used for retrieval process, S_q is known and used for evaluation. We assume that each case in the case base can be a query case. For each C_q , the closest patients C_r are retrieved from the case base.

The retrieval process is validated by comparing the solution of retrieved patients (outcome after surgery) S_r with the one of the query patient S_q . To assess the performance of the retrieval process, we use two criteria: the 'Retrieve-based criterion' proposed in [9] and a clinical

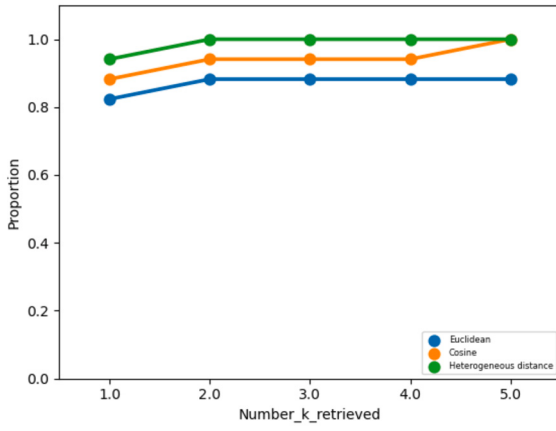


Fig. 3. Proportion of patient with at least one correct solution among the k most similar patients retrieved.

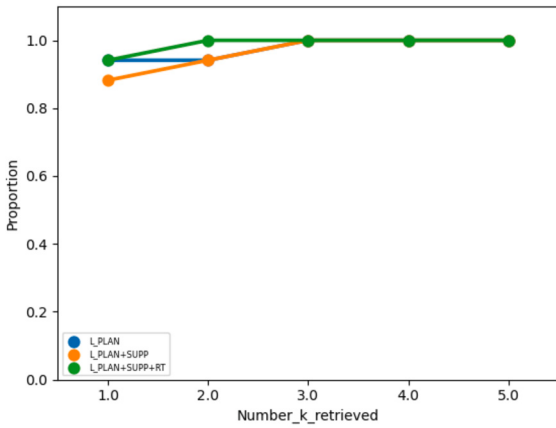


Fig. 4. Proportion of cases with at least one correct outcome among the k most similar cases retrieved.

cal relevance criterion. The 'Retrieve based criterion' method requires selection of most similar cases C_r . The cutoff is set to five similar cases representing 30% of the case base. A similarity measure is considered performant when among the k most similar retrieved patients (where $k \in [1;5]$) the correct solution appears at least once. The clinical relevance is illustrated through three typical retrieved cases. It consists in clinically interpreting the usefulness of the retrieved patients C_{r_k} with respect to the query patient C_q .

3. Results

The results section provides a quantitative evaluation of the retrieval process using the Retrieve-Based Criterion. Then, a clinical observation analyses the retrieval process output for three patients.

3.1. Retrieve-based criterion

The retrieval process is evaluated under the following conditions:

- **Distance metrics:** The three different distance metrics compared are Euclidean distance, Cosine distance and the heterogeneous distance.
- **Attributes selection:** The three lists of attributes compared are L_{PLAN} , $L_{PLAN+SUPP}$ and $L_{PLAN+SUPP+RT}$.
- **Attributes weighing:** The two weighting scenarios compared are uniform weights, and knowledge-derived weights.

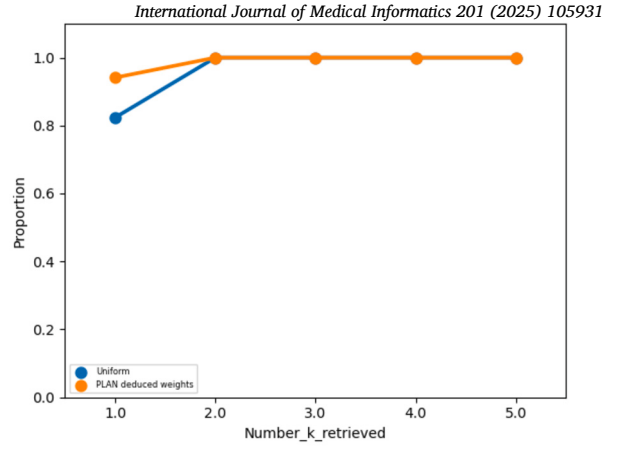


Fig. 5. Proportion of cases with at least one correct outcome among the k most similar cases retrieved.

3.1.1. Distances metric

Similarity measures using three distance metrics with attributes of $L_{PLAN+SUPP+RT}$ and knowledge-derived weights are compared. Fig. 3 illustrates the proportion of cases in which the correct solution S_q (i.e. the real outcome at 12 months post-surgery of C_q) is found at least once among the k most similar patients outcomes S_{r_k} .

The Heterogeneous distance function demonstrated the greatest performance, with 94% of C_q cases presenting with the same solution as the most similar case ($k = 1$). In comparison, Cosine distance function reached 88% C_q cases and Euclidean distance function 82%. In addition, the similarity measure using Heterogeneous distance requires a minimum of two retrieved patients for all query cases C_q to have the correct solution among the retrieved patients. Whereas, the similarity measure using Cosine distance function requires at least five patients to find the correct solution and the similarity measure using Euclidean distance function more than five patients. The higher performance of the heterogeneous distance could be explained by the use of specific distance metrics adapted to the attribute types.

3.1.2. Attributes selection

The analysis of clinical guidelines and the clinical expert survey has yielded three lists of attributes: L_{PLAN} , $L_{PLAN+SUPP}$ and $L_{PLAN+SUPP+RT}$. The objective is here to ascertain whether the addition of attributes enhances the efficacy of the retrieval process. Fig. 4 represents the results obtained with the three lists of attributes considering the heterogeneous distance and knowledge-derived weights.

All three conditions reach 100% of cases with at least the correct outcome S_q once among the three most similar patients. The first list achieving 100% with only two retrieved similar patients, is $L_{PLAN+SUPP+RT}$. The only difference between $L_{PLAN+SUPP+RT}$ and the two other lists is the inclusion of the revascularization procedure type RT . The RT seems to improve the efficacy of the retrieval process. The difference in performance between the L_{PLAN} and $L_{PLAN+SUPP}$ lies in the addition of attributes identified through surveys (see Table 4), predominantly concerning vascular surgical history, in $L_{PLAN+SUPP}$. This list performs slightly less well than L_{PLAN} when $k = 1$.

3.1.3. Attributes weighting

Fig. 5 illustrates the comparison of the similarity measures using two weighting strategies, the heterogeneous distance and the list $L_{PLAN+SUPP+RT}$. The objective of this comparison is to confirm that the application of weights derived from clinical guidelines enhances the efficacy of retrieval.

Both uniform weighting and knowledge-derived weighting reach 100% of C_q cases retrieving the correct solution S_q with a minimum two most similar patients retrieved. Nevertheless, a difference is observed when only the most similar patient is retrieved ($k = 1$). The approach

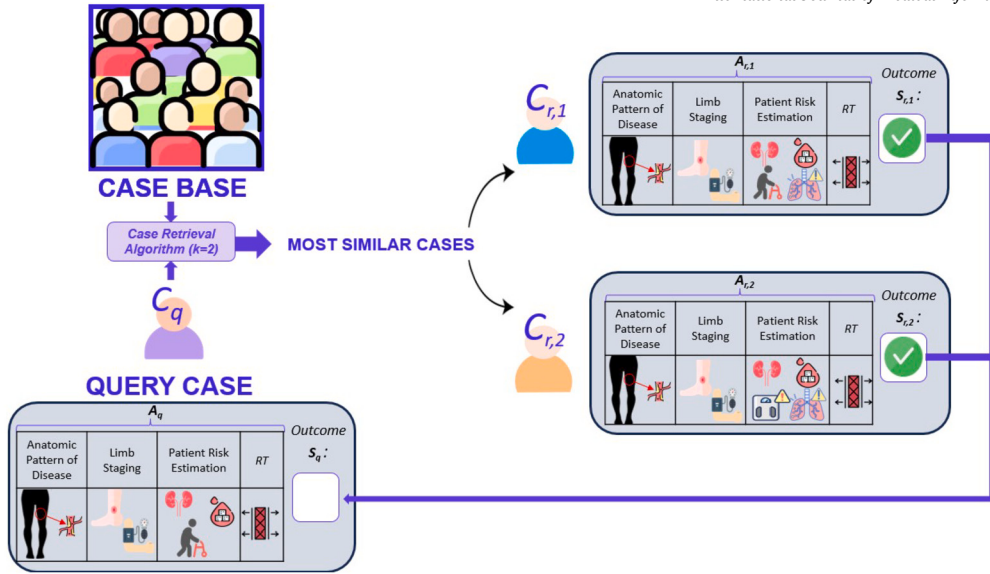


Fig. 6. Illustration of an example of retrieval for a query patient.

based on weights derived from the PLAN concept shows a better efficacy: 94% of C_q cases (i.e. 15 cases) have a retrieved case with the correct outcome compared to 82% for the uniform weighing strategy.

3.2. Clinical analysis of retrieval process

From the Retrieved-Based evaluation above, the similarity measure that performs the best with our case base employs a heterogeneous distance on the list of attributes $L_{PLAN+SUPP+RT}$ and weights derived from the PLAN concept. We propose a clinical analysis of this similarity measure with the observation of three patients. The relevance of the two most similar patients is assessed as the results showed that 100% of C_q cases have the correct solution among the retrieved patients when $k \leq 2$.

The CBR system is queried as follows: “Given the preoperative information A_q of a query patient C_q (see Tables 1, 2, 3 and 4), what would be the outcome 12 months after a revascularization procedure?” as illustrated Fig. 6.

We consider as the first Patient C_{qA} , a 94-year-old woman with a left limb LEAD. The atherosclerotic lesion is located at the femoropopliteal level (above the knee). A comparison of the two most similar patients retrieved reveals a high degree of similarity in a number of preoperative characteristics, including the presence of cardiopathies, absence of diabetes, hypertension, and chronic kidney disease (CKD), as well as the characteristics of wounds, the grade of ischemia, the location of the atherosclerotic lesion, and the estimated procedure. The remaining discrepancies relate to the preoperative weight, COPD and dyslipidemia. The three patients underwent angioplasty at the femoropopliteal level. The solution of the C_{qA} is similar to the two retrieved cases analyzed with a positive outcome 12 months post-surgery.

The patient C_{qB} , aged 81 years, presents with an obstructive lesion at the infrapopliteal level (below the knee). The two patients with the closest characteristics exhibit similarities in age, grade of ischaemia and location, estimated procedure, hypertension, antiplatelet medication, diabetes, wounds and ambulatory status. The major discrepancies relate to CKD, dyslipidemia, and cardiopathy. The 12-month postoperative outcomes for the two most analogous patients were unsuccessful, necessitating re-intervention on the same side. This outcome could be indicative of a potential failure for the query patient B. Indeed, the query patient C_{qB} underwent angioplasty for the infrapopliteal lesion, which also resulted in re-intervention, thus meeting the criteria for failure as defined in this context.

The query patient C_{qC} is a 85 years old male, with an obstructive lesion at the femoropopliteal level on the left limb. A comparison of the two closest patients with our query patient C reveals similarities in the presence of cardiopathies, medications, level of ischemia, overweight status, absence of CKD, hypertension, and wounds. However, there are differences in the estimated procedure: C_{qC} underwent a bypass surgery that necessitated a re-intervention within the 12 months after surgery. This solution S_{qC} is a failure. The most similar patient underwent a bypass surgery, which was successful at one and three months post-surgery but necessitated a re-intervention within the year, with the same outcome as C_{qC} . The second most similar patient underwent an angioplasty to treat the femoropopliteal lesion, and it was a clinical success after 12 months.

This analysis confirmed the clinical relevance of the similarity measure considered performant: the retrieved cases solutions were consistent with the query case solution. Also, the last case C_{qC} gives an indication of the potential impact of the revascularization type RT on the outcome.

In summary, the heterogeneous distance used for similarity measurement showed superior efficacy on retrieving relevant cases with regards to the outcome at 12 months post-surgery. Attribute selection revealed that the PLAN-derived attributes completed with the attributes of vascular history and revascularization type showed the best performance. The clinical consistency of retrieval process of the three cases showed promising relevance for the use of CBR-based decision support system.

4. Discussion

CBR-based CDSS could assist vascular surgeons in their decision making by retrieving similar patients to a new one. In particular, this study focuses on the estimation of the outcome 12 months after surgery. In this context, following the formalization of a case description, a retrieval process was proposed for CBR-based decision support. These preliminary results show that our approach could support clinicians in anticipating long-term surgical outcomes by retrieving cases based on a heterogeneous similarity measure adapted to the diversity of clinical data types in lower limb revascularization.

Similarity measure using a heterogeneous distance seemed more performant in retrieving similar past cases as it considers different attribute types. This result is consistent with literature and confirms that the distances between the different values of an ordinal attribute are non uniform.

Three lists of attributes were compared to analyze the influence of attribute selection. The list comprising the largest number of attributes, including the surgical procedure type, was the most effective in retrieving similar patients. The results of the clinical expert survey showed that surgeons have different intuitions on impacts of additional attributes. For example, the vascular history of the contralateral limb is differently considered among the surgeons. The results obtained with $L_{PLAN+SUPP}$ (Fig. 4) seem to show that the impact of some attributes need to be confirmed through larger database.

This study compared two weighing strategies: uniform weights and PLAN-derived weights. The results indicated that the PLAN-derived weights exhibited superior performance. Indeed, the weights assigned reflect the relative importance of attributes in relation to the question of interest. Uniform weights did not yield inconsistent results neither. One possible reason is that the PLAN-derived weights reflect impact of attributes regarding the surgery itself and not the outcome 12 months after surgery. Distance metric learning relying on a weight optimization process could enhance the performance of the similarity measure. This would necessitate in further works the building of a larger case base.

In the clinical analysis, three cases were detailed and their two most similar patients retrieved. This analysis showed clinical relevance for the proposed similarity measure. The patients C_{qA} et C_{qB} had similar patients retrieved with consistent outcomes. The patient C_{qC} underwent an open surgery procedure, which resulted in a failure after 12 months. The most similar patient retrieved showed the same outcome. The second most similar patient retrieved, underwent an endovascular procedure that resulted with a clinical success after 12 months. Both retrieved patients give an useful information for patient C_{qC} : when considering an open surgery for patient C_{qC} , the retrieved most similar patients could help reconsidering the procedure.

Despite the current dataset size limitations, due to exclusion of missing data and consideration of a single surgical procedure per patient, we managed to show consistency of the retrieval process proposed. It is necessary to extend the present dataset to generalize the findings obtained in this study. A larger case base could also propose a higher diversity of cases (symptoms, revascularization method, descriptors of success). For instance, the procedure outcome attributes could be described with a higher granularity using the occurrence of different events. Moreover, the addition of more precise attributes derived from angioscanners and related to the localization of vascular lesions as well as attributes related to wound evolution could also be considered.

In this preliminary study, we excluded missing data and proceeded to a complete case analysis (CCA). However, missing data is a persistent issue with clinical data. The question of clinical missing data imputation has been extensively addressed in the literature [16–19]. These imputation methods require a larger cohort and to properly identify origin of missing data.

Electronic Health Records (EHR) are utilized on a daily basis by clinicians to organize patients' follow-ups. Consequently, for future development of CBR-based CDSS, it is necessary to ensure the interoperability with EHR systems to enrich the case base. To this end, we have initiated the development of an Electronic Patient File application specifically designed for vascular surgeons' practice. The idea is to populate the common EHR system of the hospital in a structured way while enrich the case base. This approach aims to facilitate the automated patient data collection by clinicians during consultations and hospitalizations. This is a user-centered application that aligns its organization with hierarchy of clinical patients data and aims to reduce missing data when entering patient information.

We considered a single surgical procedure per patient in the case base. However, it is acknowledged that LEAD pathology is a chronic disease, with some patients undergoing multiple revascularization procedures for multiple lesions. The impact of the vascular history of patient regarding the revascularization success should be assessed. It is also the conclusion of the survey presented. Yet, to our knowledge no study seems to propose a specific handling of the 'recurring patients'. It is

therefore necessary to confirm that the similarity measure remains consistent in retrieving similar patients, that is to say, that the similarity will not systematically retrieve the same patient for different symptoms and different procedures.

5. Conclusion

We proposed a definition of a similarity measure adapted to lower limb revascularization decision support, with the selection of distance metrics, the selection of attributes and the estimation of attribute weights. The evaluation of similarity measure showed good performance and clinical relevance when using a heterogeneous distance on the list of attributes $L_{PLAN+SUPP+RT}$ and weights derived from the clinical guidelines. CBR may be a promising method toward decision support in every day practice for clinicians in revascularization of LEAD patients. Future investigations on larger database and more exhaustive description of patients especially concerning the arterial anatomy could help to bridge the gap toward a CBR-based clinical decision support system for vascular surgery.

6. Summary table

What was known on this topic:

- Vascular surgeons face complex decisions regarding revascularization of lower limbs of patients, for which there is no decision support solution yet available.
- Case based reasoning has been shown to make relevant data readily available to support decision in different application fields.
- The performance of a CBR-based decision support system is closely related to defining the appropriate similarity measure depending on the case description and the decisional issue to be addressed.

What this study added to our knowledge:

- The description of patient cases with lower extremity arterial disease is formalized and structured from global vascular guidelines and expert knowledge.
- A case retrieving process adapted to CBR based decision support for lower limb revascularization is proposed.
- The performance of the proposed similarity measure is analyzed and shown based on relevant clinical attributes selection and weighting.

CRedit authorship contribution statement

Margaux Roux: Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rafaëlle Spear:** Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Céline Fouard:** Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Pascal Haigron:** Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Selection of additional attributes

Fig. A.1 presents the results of the clinical survey evaluating the influence of additional attributes.



Fig. A.1. Clinical survey with four experts evaluating the influence of additional attributes on outcome.

Appendix B. Ordinal attributes matrices

Ordinal distance relies on the definition of distance matrices for each ordinal attribute. Vascular surgeons quantified the spread between ordinal values based on their expertise. The ordinal distance matrices as reported in Tables B.1 to B.6. As an example in Table B.1, the attribute a_{21} 'Overweight and Obesity' contains three possible values: 'No', 'Overweight', 'Obesity' (depending on the Body Mass Index (BMI) of the patient). The Distance between two patients without overweight and overweight is more important than the distance between a patient overweight and a patient with obesity.

Table B.1

Distance matrix for 'Overweight' possible values.

X	No	Overweight	Obesity
No	0.0	0.75	1
Overweight	0.75	0.0	0.25
Obesity	1	0.25	0

Table B.2

Distance matrix of 'Smoking status' possible values.

X	No	Deprived	Current smoker
No	0.0	0.75	1
Deprived	0.75	0.0	0.25
Current smoker	1	0.25	0

Table B.3

Distance matrix of 'Chronic Kidney Disease grade' possible values.

X	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Grade 1	0	0.25	0.4	0.8	1
Grade 2	0.25	0	0.25	0.5	0.75
Grade 3	0.5	0.25	0	0.3	0.5
Grade 4	0.8	0.5	0.3	0	0.2
Grade 5	1	0.75	0.5	0.2	0

Table B.4

Distance matrix of 'Ambulatory status' possible values.

X	Autonomous	Cane or Walker	Wheel chair	Bed bound
Autonomous	0	0.4	0.9	1
Cane or Walker	0.4	0	0.5	0.6
Wheel chair	0.9	0.5	0	0.1
Bed bound	1	0.6	0.1	0

Table B.5

Distance matrix of 'Diabetes status' possible values.

X	Not concerned	Diabetes	Diabetes with complications
Not concerned	0	0.8	1
Diabetes	0.8	0	0.2
Diabetes with complications	1	0.2	0

Table B.6

Distance matrix of 'Previous vascular surgery' (homolateral and contralateral) possible values.

X	Not concerned	Endovascular	Open Surgery	Both
Not concerned	0	0.4	0.8	1
Endovascular	0.4	0	0.2	0.6
Open Surgery	0.8	0.2	0	0.2
Both	1	0.6	0.2	0

References

- [1] R.T. Sutton, D. Pincock, D.C. Baumgart, D.C. Sadowski, R.N. Fedorak, K.I. Kroeker, An overview of clinical decision support systems: benefits, risks, and strategies for success, *npj Digit. Med.* 3 (1) (Feb. 2020), <https://doi.org/10.1038/s41746-020-0221-y>.
- [2] R.A. Miller, Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary, *J. Am. Med. Inform. Assoc.* 1 (1) (1994) 8–27, <https://doi.org/10.1136/jamia.1994.95236141>.
- [3] F. Hak, T. Guimarães, M. Santos, Towards effective clinical decision support systems: a systematic review, *PLoS ONE* 17 (8) (2022) e0272846, <https://doi.org/10.1371/journal.pone.0272846>.
- [4] A.L. Fogel, J.C. Kvedar, Artificial intelligence powers digital medicine, *npj Digit. Med.* 1 (1) (Mar. 2018), <https://doi.org/10.1038/s41746-017-0012-2>.
- [5] M.M. Richter, R.O. Weber, *Case-Based Reasoning: A Textbook*, Springer, Berlin Heidelberg, 2013.
- [6] A. Tashkandi, I. Wiese, L. Wiese, Efficient in-database patient similarity analysis for personalized medical decision support systems, *Big Data Res.* 13 (2018) 52–64, <https://doi.org/10.1016/j.bdr.2018.05.001>.
- [7] A. Sharafoddini, J.A. Dubin, J. Lee, Patient similarity in prediction models based on health data: a scoping review, *JMIR Med. Inform.* 5 (1) (2017) e7, <https://doi.org/10.2196/medinform.6730>.
- [8] D. Gu, C. Liang, H. Zhao, A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis, *Artif. Intell. Med.* 77 (2017) 31–47, <https://doi.org/10.1016/j.artmed.2017.02.003>.
- [9] J. Gliozzo, M. Mesiti, M. Notaro, A. Petrini, A. Patak, A. Puertas-Gallardo, A. Paccanaro, G. Valentini, E. Casiraghi, Heterogeneous data integration methods for patient similarity networks, *Brief. Bioinform.* 23 (4) (Jun. 2022), <https://doi.org/10.1093/bib/bbac207>.

- [10] H. Feuillâtre, V. Auffret, M. Castro, F. Lalys, H. Le Breton, M. Garreau, P. Haigron, Similarity measures and attribute selection for case-based reasoning in transcatheter aortic valve implantation, *PLoS ONE* 15 (9) (2020) e0238463, <https://doi.org/10.1371/journal.pone.0238463>.
- [11] M.S. Conte, A.W. Bradbury, P. Kolh, J.V. White, F. Dick, R. Fitridge, J.L. Mills, J.-B. Ricco, K.R. Suresh, M.H. Murad, V. Aboyans, M. Aksoy, V.-A. Alexandrescu, D. Armstrong, N. Azuma, J. Belch, M. Bergoeing, M. Bjorck, N. Chakfé, S. Cheng, J. Dawson, E.S. Debus, A. Dueck, S. Duval, H.H. Eckstein, R. Ferraresi, R. Gambhir, M. Gargiulo, P. Geraghty, S. Goode, B. Gray, W. Guo, P.C. Gupta, R. Hinchliffe, P. Jetty, K. Komori, L. Lavery, W. Liang, R. Lookstein, M. Menard, S. Misra, T. Miyata, G. Moneta, J.A.M. Prado, A. Munoz, J.E. Paolini, M. Patel, F. Pomposelli, R. Powell, P. Robless, L. Rogers, A. Schanzer, P. Schneider, S. Taylor, M. Vega De Ceniga, M. Veller, F. Vermassen, J. Wang, S. Wang, Global vascular guidelines on the management of chronic limb-threatening ischemia, *J. Vasc. Surg.* 69 (6) (2019) 3S–125S.e40, <https://doi.org/10.1016/j.jvs.2019.02.016>.
- [12] J.P. Simons, A. Schanzer, J.M. Flahive, N.H. Osborne, J.L. Mills, A.W. Bradbury, M.S. Conte, Survival prediction in patients with chronic limb-threatening ischemia who undergo infrainguinal revascularization, *Eur. J. Vasc. Endovasc. Surg.* 58 (1) (2019) S120–S134.e3, <https://doi.org/10.1016/j.ejvs.2019.04.009>.
- [13] J.L. Mills, M.S. Conte, D.G. Armstrong, F.B. Pomposelli, A. Schanzer, A.N. Sidawy, G. Andros, The society for vascular surgery lower extremity threatened limb classification system: risk stratification based on wound, ischemia, and foot infection (wifii), *J. Vasc. Surg.* 59 (1) (2014) 220–234.e2, <https://doi.org/10.1016/j.jvs.2013.08.003>.
- [14] G. Le Bourgeois, État de l'art du service de chirurgie vasculaire du CHUGA dans les stratégies de revascularisation de l'artère fémorale des patients de plus de 75 ans dans l'ère endovasculaire, Master's thesis, Jul. 2021, <https://dumas.ccsd.cnrs.fr/dumas-03285699>.
- [15] K. Varpa, K. Iltanen, M. Juhola, Genetic algorithm based approach in attribute weighting for a medical data set, *J. Comput. Med.* 2014 (2014) 1–11, <https://doi.org/10.1155/2014/526801>.
- [16] P.C. Austin, I.R. White, D.S. Lee, S. van Buuren, Missing data in clinical research: a tutorial on multiple imputation, *Can. J. Cardiol.* 37 (9) (2021) 1322–1331, <https://doi.org/10.1016/j.cjca.2020.11.010>.
- [17] M.W. Heymans, J.W. Twisk, Handling missing data in clinical research, *J. Clin. Epidemiol.* 151 (2022) 185–188, <https://doi.org/10.1016/j.jclinepi.2022.08.016>.
- [18] J.C. Jakobsen, C. Gluud, J. Wetterslev, P. Winkel, When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts, *BMC Med. Res. Methodol.* 17 (1) (Dec. 2017), <https://doi.org/10.1186/s12874-017-0442-1>.
- [19] Q. Suo, W. Zhong, F. Ma, Y. Yuan, J. Gao, A. Zhang, Metric learning on healthcare data with incomplete modalities, in: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-2019, International Joint Conferences on Artificial Intelligence Organization*, 2019, pp. 3534–3540.