ISBN 978-989-9121-31-7
2nd International Conference on Challenges in Engineering, Medical, Economics and Education: Research & Solutions (CEMEERS-23b)
Istanbul (Turkiye) Nov. 23-24, 2023

A Maintenance Case Based Reasoning Framework for COVID-19 Diagnosis

Abir Smiti[1]

[1]LARODEC, University of Tunis, Tunisia
ESSEC Tunis - Higher School of Economic and Commercial Sciences Tunis.
ISG Tunis - Higher Institute of Management of Tunis,

Abstract**:** *The rapidly advancing field of artificial intelligence (AI) plays a vital role in healthcare practices and research, significantly enhancing accuracy and achievements in the medical industry. Among the various AI techniques utilized, Case Based Reasoning (CBR) stands out as one of the most widely adopted methods in medicine. The objective of this project is to establish a CBR framework designed to assist patients in diagnosing the COVID-19 pandemic. To ensure the effectiveness, competence, and long-term performance of the CBR system, a maintenance process is imperative. This process is essential for enabling the system to operate seamlessly over an extended period while handling substantial data loads. Consequently, this article introduces a case base maintenance approach known as" Dynamic Soft Case Base Maintenance."*

Keywords: *Covid-19; Case based reasoning (CBR) ; Case base maintenance ; Medical CBR ; Soft computing;*

1. Introduction

The main Machine Learning techniques have been demonstrated to be effective in the medical industry, but there is one drawback: clinicians, doctors, and other medical professionals like to know how and why the system produced a certain recommendation or outcome. The techniques, however, do not deliver such data. However, research is being done to address this issue, often known as the "black box problem," in order to map the behaviour of the system. The "Case Based Reasoning (CBR)" approach [1], which simulates human cognitive thinking and reasoning, was created in the late 20th century, and reveals this weakness. It mimics how humans' reason and think cognitively. It mimics the same process of applying prior knowledge to current issues by extrapolating from them. An intelligent kind of knowledge reuses previously solved problems or cases on the theory that the more similar or near-identical two issues are, the more similar their solutions will be. The case-based reasoning cycle can be continued by looking for prior experiences, applying those experiences to the new situation, testing the answer to see if it works, and saving the new solution into memory if it is effective.

In this paper, we propose a CBR framework for COVID-19 epidemiology with a new approach of maintenance for the CBR named Dynamic Soft Case Base Maintenance, aiming to decrease the search time for solutions, to reduce the storage size and to dynamically manage uncertainty in data to ensure performance and effectiveness of the CBR. The rest of the paper is organized as follows: Section 2 presents Case Based Reasoning. Section 3 shows CBR COVID-19 framework, Section 4 describes and details our new case base maintenance approach, Section 5 provides the experimental results and analysis, finally Section 6 concludes this work and presents our future work.

2. Case Based Reasoning (CBR)

Case-based reasoning (CBR) is a problem-solving approach that relies on previously solved cases to solve new problems by adapting past solutions. The technique involves retrieving a relevant set of similar cases from memory, reusing one or more of the retrieved solutions, revising the solution to fit the current problem, and storing the new solution in memory for future use. CBR finds common use in fields such as engineering, medicine, and law, where there is a wealth of expert knowledge that can be represented in the form of cases. It is a type of artificial intelligence that enables computers to learn from experience and adjust to novel situations (See Figure 1)

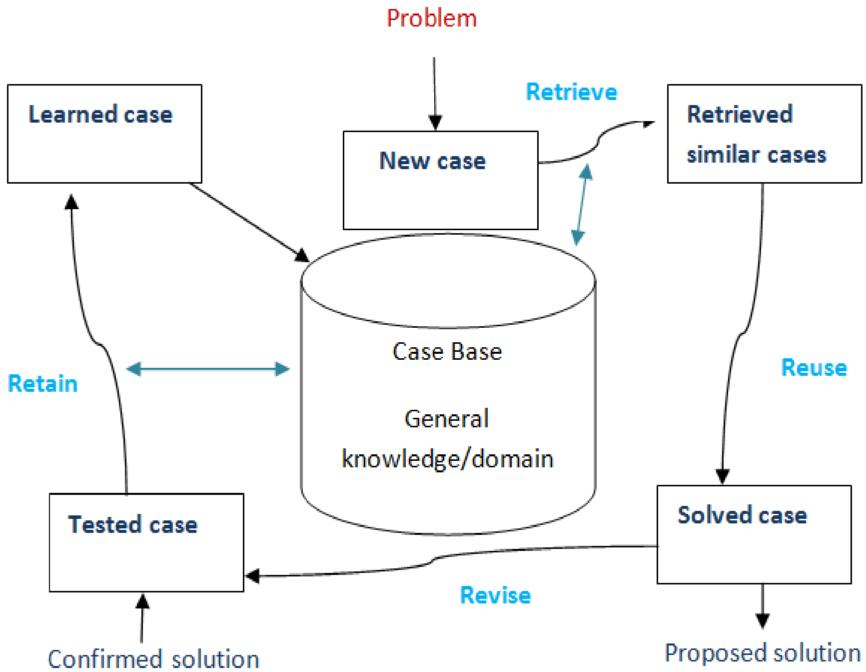

Fig. 1: Case Based Reasoning Cycle

3. CBR-COVID-19 framework

The sickness is known as corona virus disease (COVID-19), and the virus is known as severe acute respiratory syndrome corona virus 2. Corona virus is one of the important infections and an infectious disease that deteriorates the human respiratory system (SARS-CoV-2). Since COVID-19 symptoms are very similar to those of other respiratory illnesses and there is currently no developed vaccine, patients are treated based on past events and viruses that are similar to those they have already experienced. This is similar to the Case Based Reasoning system, which develops solutions to new problems based on past cases presented as a set of related facts and knowledge. As the symptoms of this virus and respiratory diseases like flu or bronchitis are very similar.

The CBR system for COVID-19 prediction [2] can assist users in making decisions regarding whether to check their status when they feel like they have symptoms related to the virus or for clinicians to help ease the diagnosis. When a new virus first appeared, the first insight was to look for earlier viruses that manifested similarly. As a result, the case-based reasoning (CBR) system mimics the same mechanism of reusing past experiences and projecting them on new encountered problems to derive a solution, making CBR one of the best reasoning methodologies in the medical field (See Figure 2).

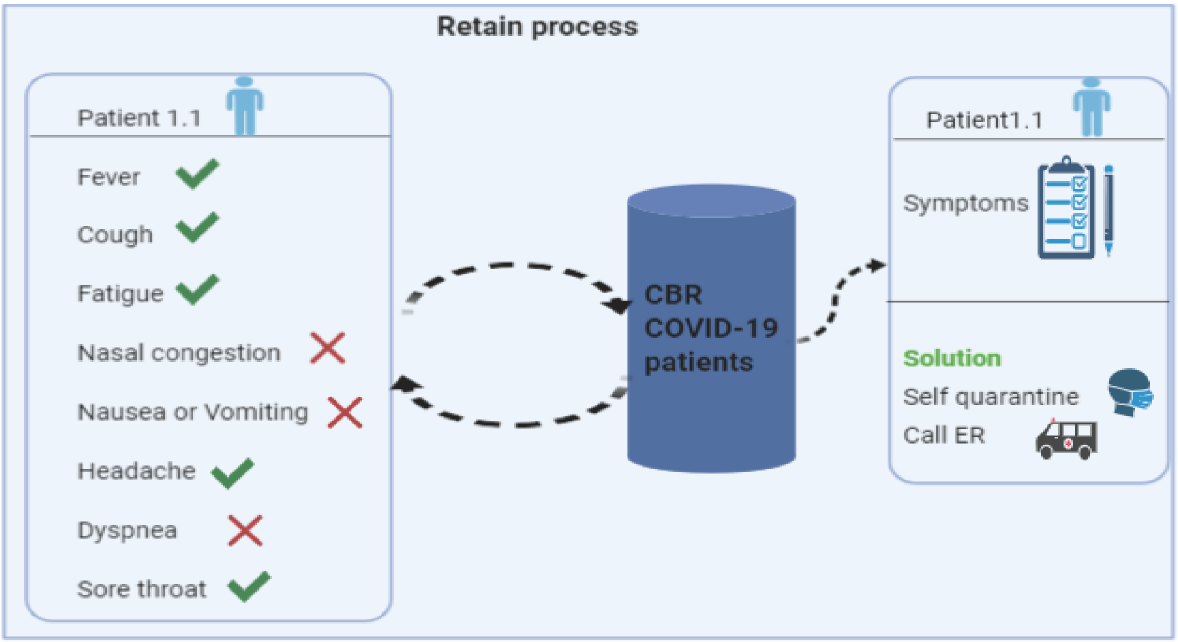

Fig. 2: CBR system for COVID-19 prediction

The maintenance process should be used to create a system that operates for a long time and interacts with significant amounts of data to assure the CBR efficiency, competence, and performance, to avoid mistakes and inaccuracies. The new suggested approach will be tested on a case-base COVID-19 dataset in the following and assessed using performance and competence standards.

4. Dynamic CBR maintenance proposed method.

Maintenance in CBR is a continuous process during the life cycle of the system and is achieved by adding, editing, and deleting cases to avoid redundancy or to reduce the scale of case bases. Case Base Maintenance (CBM) presents the process of refining the CBR system to attain enhancement of the accuracy, performance, and integrity of the system [3], and the case base maintenance is considered the most important issue when used for medical purposes because of the massive databases and the need for accurate results.

The main policies of CBM approaches are the CBM optimization, CBM Partitioning and Dynamic CBM. Each policy satisfies the reductionist approach preserving as much as possible the competence of the case base system but suffers also from some limitations especially with the existence of noisy cases, since it depends on the type of the stored cases and the complexity to run these policies on large incremental case bases which is our current problem. Our proposed method aims to maintain the CBR system on different levels, considering the limitations cited above, we propose a new efficient technique of machine learning handling both noise and the dynamic change of the medical case bases over time. As medical databases are known for their large size and to handle the incremental aspect of the cases in the CBR medical system, a dynamic method is required simultaneously it must detect noises without the need to declare the input parameters for the clusters, the clustering technique needed to address these characteristics is the DBSCAN-GM [4]. Furthermore, to ensure credibility as the medical area is a vital and real-world domain and to simulate the human reasoning, the clustering technique DBSCAN-GM is run with the fuzzy set theory Soft DBSCAN [5].

Combining these two methods results into a Dynamic Soft DBSCAN-GM technique, solving several problems at once, the large size of the medical case base will be reduced containing only the competent cases in a dynamic way having no issue in adding cases over time. Our approach consists of two steps, at first, we partition the case base incrementally generating a dynamic number of clusters that determines as well as possible noisy data, second step competence groups are assigned to the determined clusters to detect the different types of cases to finally alleviate the case base and preserve the competence and performance of the system, the different types of the detected clusters are detailed below:

4.1. Noise cases detection

The uncertainty management of the soft DBSCAN-GM technique detects and automatically generates the noise cases which will be deleted at first since their existence lowers the competence of the system.

4.2. Similar cases detection

After eliminating the noise cases, the remaining cases will be whether isolated or similar ones, For similar cases as they represent redundant cases will be deleted in order to reduce the case base size, only one case will be left for each cluster and will be calculated by a distance measurement as the nearest case to the center of each cluster, Then cases having the short distance to this one center case presenting the similar type will be deleted.

4.3. Isolated cases detection

Isolated cases define cases belonging to the clusters but very much distant from the center, their detection will be measured by a distance measurement contrarily to the similar type they have a large distant from the center, and they will not be deleted considering they represent and covers only themselves.

The novel DSCBM method is resumed in first partitioning the case base, then types of cases are detected, noise cases are deleted, second similar cases will also be deleted after keeping only one of each group to replace the whole set, isolated cases will not be touched.

5. Results and Discussion

TABLE I: The Arrangement of Channels

Dataset	Ref	Instances	Attributes
SARS-COV2	Cov-19	127	21

Experimentation and the analysis of the 'Dynamic Soft Case Base Maintenance' approach, are developed with Matlab R2020a (Version 9.8), and experiments are carried on COVID-19 data from **[6]** data repository for machine learning data sets, Below an example of 10 patients (See Figure 3) with details of the data-set in (Table I) and for the metric distance for this data set as it presents a mix of multiple values numeric and binary.

The Gower distance measure **[7]** is applied consisting on Manhattan distance for the numeric values, and dice distance for the binary values with the following formula: $d(i,j) = \frac{1}{p}\sum_{i=1}^{p} d_i j^{(f)}$

Where $d_{i,j}^{(f)}$ presents the type of the computed distance of different values.
Measurements for the effectiveness and performance of the new maintenance method are concluded through several evaluation criteria, and results are compared with the initial non-maintained CB, non-dynamic CBM approaches, and dynamic CBM approach.

5.1. Evaluation Criteria

- Storage size [S%]: after applying maintenance and obtaining a reduced case base size, the reduction size rate is calculated as follows:

  S = (Case base size after maintenance/ Initial training case base size) $*$ 100

- Retrieval time [t(s)]: the time criterion demonstrates the performance of the method and is calculated in seconds exerted in 1-Nearest Neighbor algorithm.

- Accuracy [PCC%]: referring to the percent of correct classification, determined as a percentage, applying the 1-Nearest neighbor algorithm along with the method of 10-Fold Cross Validation, this criterion is defined as:

PCC = (Well classified instances \ Total classified instances) * 100

- Competence [COMP %]: the competence measurement presents the number of problems that can be solved, this criterion will define the global competence of the case base by calculating the coverage set of all cases.

5.2. Experimental results

In order to evaluate the DSCBM approach, it has to be compared to well other known maintenance techniques: First, it will be compared to the initial CBR, then the optimization and partitioning approaches such as CNN [8], RNN [9], ENN [10], COID [11], WCOID [12] and finally with a dynamic approach which is DMCB [13], Comparison analysis will be based on the mentioned evaluation criteria.

5.2.1. Initial CBR and DSCBM:

To highlight our approach ability in handling the dynamic aspect of maintenance, measurements and experiments were applied on different random sizes of data and dynamicity is presented through evolving case bases by first partitioning the case base into two random subsets, secondly selecting a subset of cases, then the case base is incremented randomly with the rest of cases from the second subset of data.

Obtained results from (Table II) showing performance and competence rates in terms of storage size comparing to the initial CBR. We note a reduction size reaching 40%, which remarkably decreases the retrieval time. For the accuracy rate results our DSCBM method is achieving better accuracy reaching 99,01% which proves that DSCBM method is highly preserving and improving the performance of the system along with the competence which is more efficient than the initial CBR.

TABLE II: INITIAL CBR AND DSCBM

		I-CBR				DSCBM			
		Performance			Competence	Performance			Competence
Dataset	Learning Case	S%	T%	PCC%	Comp%	S%	T%	Pcc%	Comp%
Covid-19	75	100	0,0188	83,10	88,10	71,1	0,0098	96,30	98,30
	127	100	0,0196	84,89	91,16	62,8	0,0123	99,01	97,04

5.2.2. DSCBM and other methods comparison

Comparing to the other methods, our approach marks a reduction rate of only 62,78% which is totally explainable due to the nature of our tested dataset, and our method manages high level of uncertainty in data, therefore it deletes only 100% detected noisy data (See Table III). In terms of search time for the retrieval, our method strikes the best result which ascertains that our technique keeps only relevant data. Similar impressive observations are obtained for the PCC criterion as the PCC reaches 99,01% . After enhancing the performance of the CBR system, our approach also preserves its competence, the realized competence rate is up to 97%.

TABLE II: DSCBM and other methods

	CNN	RNN	ENN	COID	WCOID	DMCB	DSCBM
S%	42.11	57,82	64,49	37,10	33,89	24,76	62,78
T%	0,3420	0,4220	0,3730	0,1950	0,1660	0,0401	0,0123
PCC%	71,09	73,98	87,62	92,11	92,93	98,06	99,01
Comp%	78,50	78,26	83,96	87,22	87,39	96,66	97,04

6. Conclusion and Future Work

In this paper, we proposed a new maintenance policy named Dynamic Soft Case Base Maintenance applied to a medical CBR framework for novel epidemiology crisis (COVID-19), we focused on managing the dynamic aspect of such domain along with the uncertainty and ambiguity of medical unknowing diseases data, the obtained results confirms that our method generates the most relevant and competent data for such a delicate domain. As future work we aim to explore the remaining phases of the CBR system introducing the dynamic soft model.

7. Funding Acknowledgements

This work is supported by Ministry of higher education and scientific research of the republic of Tunisia.

8. References

[1] R. C. Schank, Dynamic memory: a theory of reminding and learning in computers and people. Cambridge Cambridgeshire] : New York: Cambridge University Press, 1982.

[2] A. Smiti and M. Nssibi, “Case based reasoning framework for COVID-19 diagnosis,” Ing´enierie des Syst`emes d Inf., vol. 25, no. 4, pp. 469–474, 2020. [Online]. Available: https://doi.org/10.18280/isi.250409

[3] D. C. Wilson and D. B. Leake, “Maintaining Case-Based Reasoners: Dimensions and Directions,” Computational Intelligence, vol. 17, pp. 196–213, May 2001. https://doi.org/10.1111/0824-7935.00140

[4] A. Smiti and Z. Elouedi, “Dynamic dbscan-gm clustering algorithm,” 2015 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI), pp. 311–316, 2015. https://doi.org/10.1109/CINTI.2015.7382941

[5] A. Smiti and Z. Eloudi, “Soft dbscan: Improving dbscan clustering method using fuzzy set theory,” in 2013 6th International Conference on Human System Interactions (HSI), 2013, pp. 380–385. https://doi.org/10.1109/HSI.2013.6577851

[6] kaggle data sets, https://www.kaggle.com/datasets.

[7] J. C. GOWER, “Some distance properties of latent root and vector methods used in multivariate analysis,” Biometrika, vol. 53, no. 3-4, pp. 325–338, 12 1966. [Online]. Available: https://doi.org/10.1093/biomet/53.3-4.325

[8] P. Hart, “The condensed nearest neighbor rule (corresp.),” IEEE Transactions on Information Theory, vol. 14, no. 3, pp. 515–516, 1968. https://doi.org/10.1109/TIT.1968.1054155

[9] G. Gates, “The reduced nearest neighbor rule (corresp.),” IEEE Transactions on Information Theory, vol. 18, no. 3, pp. 431–433, 1972. https://doi.org/10.1109/TIT.1972.1054809

[10] D. L. Wilson, “Asymptotic properties of nearest neighbor rules using edited data,” IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-2, no. 3, pp. 408–421, 1972. https://doi.org/10.1109/TSMC.1972.4309137

[11] A. Smiti and Z. Elouedi, “Coid: Maintaining case method based on clustering, outliers and internal detection,” 2010. https://doi.org/10.1007/978-3-642-13265-0_4

[12] A. Smiti and Z. Elouedi, “Wcoid: Maintaining case-based reasoning systems using weighting, clustering, outliers and internal cases detection,” 2011 11th International Conference on Intelligent Systems Design and Applications, pp. 356–361, 2011.

https://doi.org/10.1109/ISDA.2011.6121681

[13] A. Smiti and Z. Elouedi, “Dynamic maintenance case base using knowledge discovery techniques for case based reasoning systems,” Theoretical Computer Science, p. S0304397519304189, Jul. 2019

https://doi.org/10.1016/j.tcs.2019.06.026