

Case Based Reasoning Framework for COVID-19 Diagnosis

Abir Smiti^{1*}, Maha Nssibi²

¹ LARODEC, Université de Tunis, Institut Supérieur de Gestion de Tunis, 41 Avenue de la liberté, cité Bouchoucha, 2000 Le Bardo, Tunisia

² Ecole Supérieure d'Economie Numérique, Université de Manouba, Technopole de la Manouba, Manouba CP 2010, Tunisia

Corresponding Author Email: smiti.abir@gmail.com



<https://doi.org/10.18280/isi.250409>

ABSTRACT

Received: 9 June 2020

Accepted: 13 August 2020

Keywords:

machine learning, case based reasoning, clustering, classification, COVID-19 pandemic, diagnosis, prediction

The expanding area of Artificial Intelligence is playing a vital role in healthcare practices and research, and as medical field is rich in data can become difficult to interpret, the AI techniques present the preeminent solution to enhance the medical field achievements, thus as novel epidemiology and pathogens presents a critical and emerging issue for global health, the aim of the work presented in this paper is to structure a CBR framework that aid in the patients diagnosis of novel epidemiology presence, the novel pandemic Corona-virus disease (COVID19). The objective of this study is to highlight the Case Based Reasoning (CBR) AI method which is one of the most successful applied methods in the medical field, used for analysis, prediction, diagnosis, and recommendation treatment. This study proposes a CBR conceptual framework for COVID-19 disease prediction, able to aid in the diagnosis, to provide self-health assistant and to guide people in self testing and checking.

1. INTRODUCTION

Artificial intelligence applied to medical domain has recently made headlines, as health systems became overburdened and under resourced to handle the new challenges of daily life with the growing populations, a massive number of researches is ongoing and with machine learning tools models are developed to predict diseases, illness, aid in early diagnosis, drug discovery and manufacturing, and as novel epidemiology and pathogens presents a critical and emerging issue for global health, especially viral diseases that are instantaneously and easily transmissible, highly contagious and have uncharacteristic infectious periods, like the major current pathogen that is targeting the human respiratory system, the novel pandemic Corona-virus disease COVID-19, which resulted in major quarantines throughout the world to prevent further spread, caused by severe acute respiratory syndrome (SARS-COV2) [1], that represents the causative agent of a potentially lethal disease.

Precedent outbreaks of corona-viruses are the severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS) and they are characterized as agents of massive public health threat, the incubation period of this virus is set from 2 to 10 days [2], depending on the age and the health status of the patient within 6 to 40 days with a median of 14 it can cause death, the major symptoms of COVID-19 illness are fever, cough, shortness of breath, sore throat and fatigue among other [2], and since COVID-19 symptoms are very similar to other respiratory diseases and no definitive vaccine exists yet, the treatment of patients is based on past situations and similar or close viruses which is an identical process of an artificial intelligence method the Case Based Reasoning system (CBR) [3], that summarizes a common human daily behavior in solving problems to a reasoning method applied in the field of artificial intelligence by

developing a computer programs that generates solutions to new problems based on past cases, presented as a set of related facts and knowledge.

CBR models the human cognitive thinking and reasoning, it imitates the same mechanism of reusing past experiences and projecting them on new encountered problems. Previous solved problems, or cases, are reused in an intelligent form of knowledge, relying on the basis that the more two problems are similar or close, more similar their solutions will be.

CBR first roots were in the work of cognitive psychologist and artificial intelligence theorist, Roger Schank [3], studying the way humans solve problems he found out that most people when they encounter new problems derives solutions based on past experiences with the same or close conditions and situations.

Applications fields of CBR ranges from law to medical to machine learning and engineering to management and finance, etc.

In the medical field CBR is modeled for all the tasks of diagnosis, treatment, planning and aiding, achieved by gathering cases and implemented by identifying significant features that construct a case, the same way clinicians doctors and physicians relies on past experiences to determine a new patient status.

CBR system as learning method in the area of AI stores the cases referred as the training examples in order to access them later to solve new problems or to get a prediction by extracting those cases that are similar or close to the new ones, and this is achieved by accumulating data to build a model of understanding of a new problem by storing the conditions of the new problem, then establishing a collection of rules that dictate when a given situation, we have one or more actions to resolve it, the system is then focused on recognizing the similarity of new problems to existing ones, where a case is a record of a past situation or a problem that has been already

solved, information are recorded considering the domain the case will be associated with, detailed specifications of the situation, description of all the attributes and circumstances of the environment, along with the solution of that case with all the facts and the steps making it successful solution and a solved case, and it can be represented in different types and formats which makes CBR the most flexible method.

This cases are indexed in a form of a computational data structure that can speed up the search on the stored cases in memory and to pres-elect cases according to some criteria for example with important words such as terminologies (e.g. nausea, itching, vomiting...) and their synonyms, or adjectives (e.g. contagious, inflammation, abdominal pain or discomfort...) and verbs (e.g. occur, function, cause...) that make cases more precise. The input for the search process to find similarities can be mapped in different formats, and different methodologies are applied whether manual or automated such as prototypes, symbols, vocabularies, clusters, generalized cases, inverted indexes and networks.

CBR system consists on searching for past experiences, adapting them to the new one, testing the solution to see whether it works or not, storing the new one into memory if efficient and is performed with a 4R process [4] (Figure 1).

Retrieval phase which consists on finding a case that is close to the current one to retrieve the adequate solution.

Mechanism of finding a solved case that is similar to the present situation, this phase is achieved by relying on similarity algorithms or heuristic functions. A criterion and a mechanism of control are required to search and retrieve the most suitable and close case to the current one from the stored cases.

This search can be conducted in two different ways: whether a search for a whole case with all its features or by portions as often features of many cases are built from other cases, also the process of retrieving cases relies on the case base model and the normalizing procedures that has been used to store the cases and making it possible to apply distance metrics.

Reuse phase where the retrieved case is proposed as a valid solution to the current case, when it's a simple case, this phase

is applied without any changes but in most cases it may involve adapting it as needed in order to fit the features and conditions of the new case.

Adaptation process recognizes the differences between the new and retrieved cases by very different methods as substitution, transformation, and derivation.

Revision phase is basically mapping the solution of the previous retrieved case to the current one in real, it can be achieved through a series of simulations or tests to evaluate the proposed solution performance and to determine how well it fits the new case.

Retention phase after a successful execution, this final phase occurs to store and update the resulting case in memory of the case base by determining whether it's useful or not, some cases may be rejected, other cases can involve removing old ones from the case base to improve the CBR system and making the CBR a learning system.

Each of these phases interact with the collection of cases, by editing the case base or retrieving the solutions, which makes CBR an incremental learning method being constantly edited as the system searches for the closest case to the current one and constantly updates the case base.

This 4R process of the CBR system provides a problem-solving system that can generalize its previous experiences allowing an application for the solutions of solved problems made in similar and close circumstances to new situations.

In this paper, we propose a conceptual CBR framework applied to the medical domain for the diagnosis and prediction of the novel epidemiology the COVID-19 virus.

This framework compares COVID-19 patients of the CBR system case base and predicts patients who have symptoms related to COVID-19 disease.

The rest of the paper is organized as follows: Section 2 presents the medical CBR related work with the existing applications of CBR in medical field. Section 3 provides a summary of the new epidemiology. Section 4 describes the proposed framework of CBR for the prediction of COVID-19 disease, and Section 5 concludes this work and presents the future work.

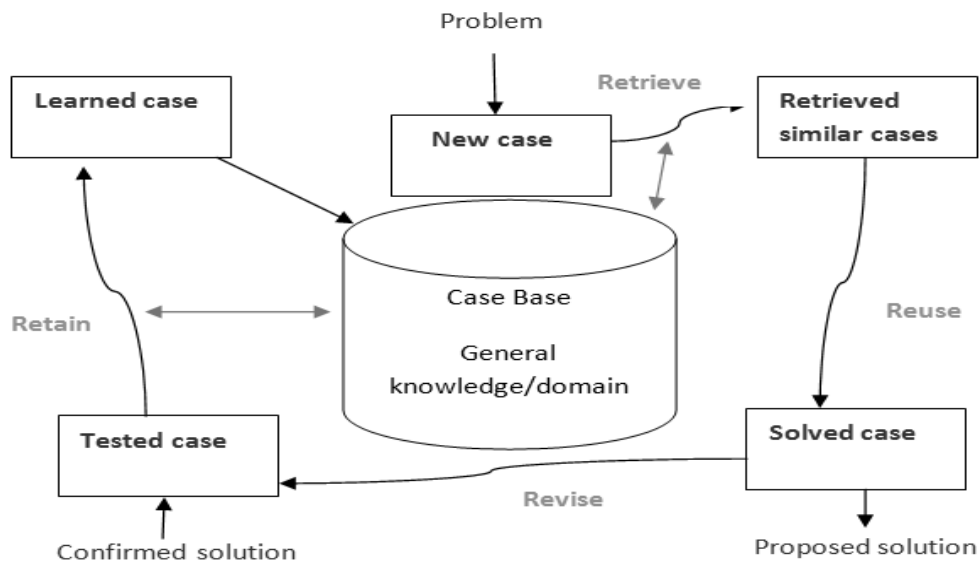


Figure 1. Case based reasoning cycle

Table 1. Medical CBR systems

CBR system	Description
CASSEY	Analyses a set of symptoms of patients with heart disease and delivers a diagnostic explanation [5]
BOLERO	Diagnosis pneumonia, the cause of it and learns from its success and failures [6]
SHRINK	Helps in psychiatric diagnosis [7]
MNAOMIA	Treating psychiatric eating disorder [8]
AUGUST	Helps prescribing neuroleptic drugs for Alzheimer patients [9]
SOMNUS	Assists with the respiratory therapy in the sleep disorders [10]
CARE PARTNER	Supports long term follow up care of stem cell transplant patients [11]
KASIMIR	Diagnosis of breast cancer [12]
SIDSTOU	Aid in tutoring of tourette syndrome [13]
MARLING	Diabetes diagnosis and decision [14]

2. CASE BASED REASONING AND THE MEDICAL FIELD

Medical field is rich in data and previous experiences and it's known that the diagnosis systems use previous records as a source of base experiences to interpret and determine new patient problems.

Clinicians and doctors generally start to practice with initial experiences, then they use these previous experiences to derive a solution for the new problems or situations, they use the diagnosis of the previous cases and develops a hypothesis about the diagnosis of the new one.

As obvious remembering past experiences or cases has the main advantage of avoiding and repeating past errors and faults, that way CBR focuses only on the significant features of the new problem, no knowledge or rules are required to solve problems, the system learns through the reuse.

The knowledge task is reduced as it consists on only collecting relevant past cases representing them and storing them.

In the approach of medical CBR application, the knowledge task is just resembling the physicians reasoning about patients and copying how they use the expertise in cases.

No explicit knowledge of the domain is needed, it can be applied even if the existing rules are least possible.

No need for full knowledge of all clinic experts to build the system, CBR is automated and based on existing cases.

The major problem of incomplete data that can decrease the efficiency of other methods and systems is not an obstacle, CBR relies just on similarity between cases to find the solution.

It can be adjusted to simply fulfill the requirements of a certain surgeon or a clinician and can be abstracted by generalizing the cases for adequate hypothesis.

An intuitive and dynamic learner as cases are added over time to the case base, CBR reasons with a large variety of situations, provides a much higher success and presents solutions to the diagnosis of diseases or medical conditions, classification of patients their status and history, planning a procedure for therapy or treatment, creating promising opportunities to handle increasingly large, complex, incomplete, and uncertain data in clinical environments.

The process is transparent and easily understandable by users which justifies decisions and increases approval. Below a summary of the most famous CBR medical applications is presented (Table 1).

Researchers are working on medical CBR with diverse applications, ranging from diagnosis and decision support systems, classification systems, planning systems to tutoring systems in order to deliver a preminent method to aid and

enhance the efficiency of the healthcare domain, and it has been proven that CBR is very useful for problem solving in health science.

3. THE EPIDEMIOLOGY AND PATHOGENESIS OF CORONA VIRUS DISEASE (COVID19)

Corona virus is one of the crucial pathogens and an infectious disease that deteriorates the human respiratory system, the disease name is corona virus disease (COVID-19) and the virus name is severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) [1].

Precedent outbreaks of corona viruses are the severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS) and they are characterized as agents of massive public health threat.

In late December 2019, a group of patients with an initial diagnosis of pneumonia of a strange etiology was admitted to hospitals, then these patients were epidemiologically associated to a seafood and wet animal market in Wuhan, Hubei Province, China [1].

Early reports predicted the beginning of a potential Corona virus given the estimation of the reproduction number for the 2019 Novel Corona virus [10], named after by World Health Organization on Feb-2020 COVID-19.

3.1 Chronology of COVID-19

The first cases were reported in December 2019 [11]. From December 18, 2019 to December 29, 2019, five patients were admitted to hospital with severe respiratory distress syndrome and one of them died [1].

January 2, 2020, 44 patients had been identified as having laboratory confirmed COVID-19 infection. January 22, 2020, a total of 571 cases of COVID-19 were reported in 25 districts and cities in China, and the China National Health Commission reported the details of the 17 deaths. January 25, 2020, a total of 1975 cases were confirmed with the COVID-19 infection in mainland China with a total of 56 deaths. January 30, 2020, 7734 cases have been confirmed in China and 90 other cases have also been reported from a huge number of countries including Taiwan, Thailand, Vietnam, Malaysia, Nepal, Sri Lanka, Cambodia, Japan, Singapore, Republic of Korea, United Arab Emirates, United States, The Philippines, India, Australia, Canada, Finland, France, and Germany [12]. And then the situation escalated to a pandemic declared officially by World Health Organization on March 11, 2020 leading to a global lock-down of millions of people.

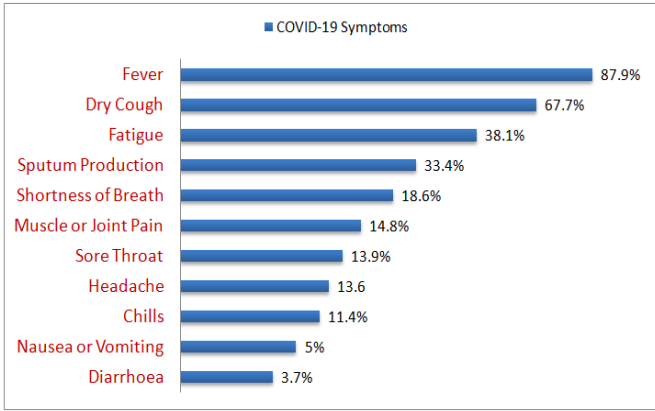


Figure 2. COVID-19 symptoms rate

3.2 Symptoms and incubation period of COVID-19

The symptoms of COVID-19 virus manifest after an incubation period of 2 to 10 days [13].

The period from the beginning of first symptoms to death ranges from 6 to 41 days with a median of 14, this period depends also on the age and the health status of the patient. The most common symptoms of COVID-19 illness are fever, dry cough, and fatigue among other, with the following rates [14] (See Figure 2).

4. CASE BASED REASONING FRAMEWORK FOR COVID-19 DIAGNOSIS

As mentioned, AI has a vital role in fighting and helping against the new epidemiology crisis, with different methods and approaches, in this section, a presentation of a CBR framework will be detailed as Case Based Reasoning is stated one of the most successful applied AI methods in the medical field.

As explained above the process of CBR reasons with previous experiments and in this framework cases are used to find a match for a particular set of symptoms in order to conclude a diagnosis or a detection of COVID-19 virus infection, below an explanation on how CBR works based on the 4R process [4], retrieve phase, reuse, revisit and retain (See Figure 3).

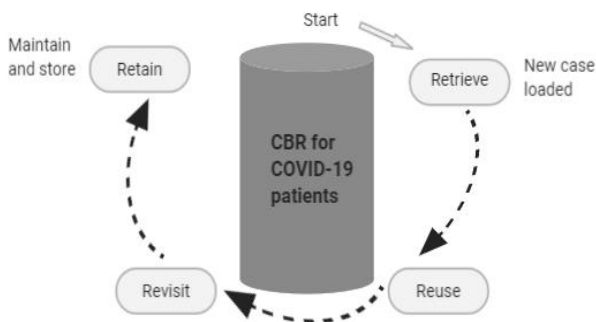


Figure 3. CBR process

This framework has two main processes:

- Classification process in the retrieval phase with K-Nearest Neighbor algorithm [4], that classifies the patient records, based on age, gender, and symptoms.
- Prediction process in final phase, to predict and to determine whether the patient has the COVID-19 or not.

The overview of the method is in (Figure 4) and detailed below.

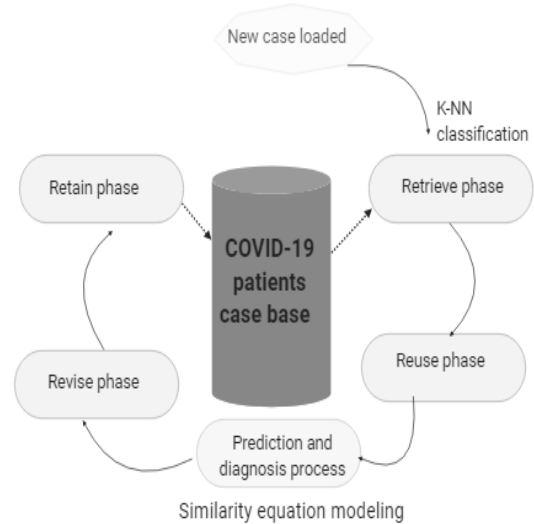


Figure 4. CBR framework for COVID-19 diagnosis

- Case Base: the cases are structured as a pair of case/solution:

	Patient 1.0	Patient 1.1
	Problem (Symptoms)	Problem (Symptoms)
	Fever ✗	Fever ✓
	Cough ✗	Cough ✓
	Fatigue ✗	Fatigue ✗
	Nasal congestion ✗	Nasal congestion ✗
	Nausea or vomiting ✓	Nausea or vomiting ✗
	Headache ✗	Headache ✗
	Dyspnea ✓	Dyspnea ✗
	Sore throat ✗	Sore throat ✓

Figure 5. Case base cases

The case $C_i=(a_i, v_i)$ is organized as attributes vector $a_i=(a_{i1}, a_{i2}, \dots, a_{in})$ and vector values $v_i=(v_{i1}, v_{i2}, \dots, v_{in})$ defining the patient detected symptoms (Figure 5).

The solution S_i is represented as vectors $S_i=(s_{i1}, \dots, s_{in})$ defining the related disease to the given set of symptoms (Figure 6).

- Retrieval phase: to calculate the distance and define the similarities between the new patient and the stored cases (Figure 7), two measurements are conducted:

1. Local similarity measurement to define the similarity metric for the attributes of the case: $sim(C_{newi}, C_{ji})$.
2. Global similarity measurement to define global similarity metric to compare cases: $sim(C_{newi}, C_j) = \sum_{i=1}^n w_i sim_i(C_{newi}, C_{ji})$ with $\sum_{i=1}^n w_i = 1$ where $0 < w_i < 1$ is the weight of i-th attribute, and as a metric measurement the weighted Euclidean distance is applied.

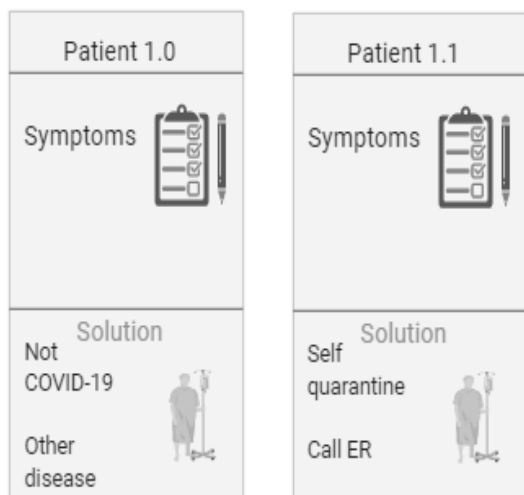


Figure 6. Case base solutions

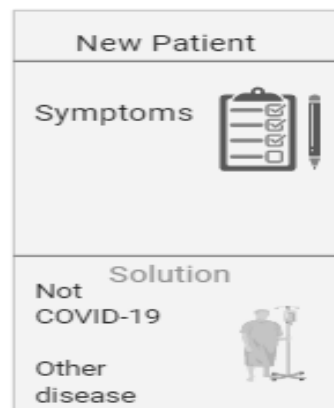


Figure 8. Retain phase

The CBR system for COVID-19 diagnosis and prediction can help in decision making for the users whether simple individuals to check their status when they fell like they have symptoms related to the virus, or for clinicians to help ease the diagnosis as the symptoms of this virus and the respiratory diseases like flu or bronchitis are very close.

The case based reasoning (CBR) system relies on the same cognitive thinking and reasoning of doctors, biologists, physicians and all human beings, thusly when first the novel virus appeared, first insight was to search for previous viruses that manifest alike, therefore CBR is one of the most suited reasoning methodology in medical field, particularly when it's about unknown diseases, or when basic rules do not apply.

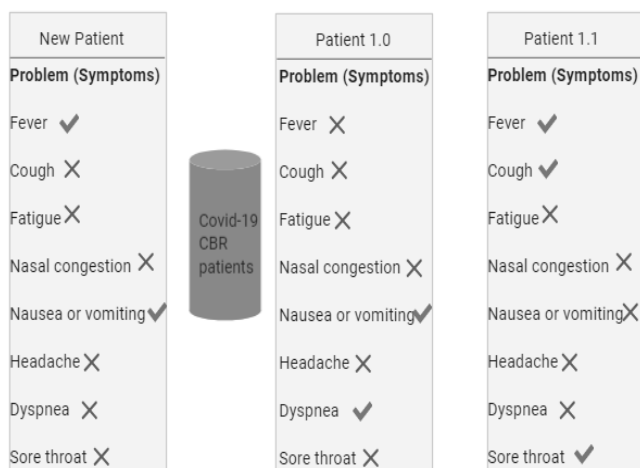


Figure 7. Retrieval phase

- Reuse phase: reuse of the chosen solution based on the K-NN retrieval with similarity measurements, the solution of the new situation s_{new} is the solution of the chosen case s_{K-NN} (Figure 8).
- Revision phase: decision and verification are based on simulation as cases are collected from real medical data of hospitals.
- Retain phase: after processing cases one by one, the output will be sent to the users showing percentage results in the terms of an estimation and proposes actions such as should the user be self-quarantined, calls ER or that their symptoms are not related to COVID-19 virus, and then the result of the new case will be stored in the CBR case base for future reuse (Figure 8).

5. CONCLUSION AND FUTURE WORK

CBR is the most suitable method for the medical field as it presents the cognitive reasoning, explicit experiences, automatic decision making with detailed explanation and can aid in building intelligent diagnosis systems and as developed in this paper it can help in fighting against the new epidemiology.

In this paper, we have proposed a conceptual CBR framework for COVID-19 disease diagnosis, the result of this framework is quite significant especially that this disease manifestation is very similar to respiratory diseases at first and can mislead an interpretation of symptoms by that this work helps clinicians in decision making for patient's status, warn and comfort users depending on the situation.

To ensure the CBR efficiency, competence, and performance, to avoid errors and inaccuracies, maintenance process should be applied.

Therefore, as a future work, we aim to build a system that works for a long period of time and deals with large amount of data by introducing a maintenance phase to the CBR system, and we aim to compare our framework with other prediction techniques with different data bases of various diseases.

REFERENCES

- [1] Bogoch, I.I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M.U.G., Khan, K. (2020). Pneumonia of unknown aetiology in Wuhan, China: Potential for international spread via commercial air travel. *Journal of Travel Medicine*, 27(2). <https://doi.org/10.1093/jtm/taaa008>

- [2] Toit, A.D. (2020). Outbreak of a novel coronavirus. *Nature Reviews Microbiology*, 18: 123-123. <https://doi.org/10.1038/s41579-020-0332-0>
- [3] Shank, R.C. (1982). *Dynamic memory: A theory of reminding and learning in computers and people*. Cambridge University Press, Cambridge, New York.
- [4] De Mantaras, R.L., Mcsherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Faltings, B., Maher, M.L., Cox, M.T., Forbus, K., Keane, M., Aamodt, A., Watson, I. (2005). Retrieval, reuse, revision and retention in case-based reasoning. *The Knowledge Engineering Review*, 20(3): 215-240. <https://doi.org/10.1017/S0269888906000646>
- [5] Smiti, A., Elouedi, Z. (2020). Dynamic maintenance case base using knowledge discovery techniques for case based reasoning systems. *Theoretical Computer Science*, 817: 24-32. <https://doi.org/10.1016/j.tcs.2019.06.026>
- [6] Smiti, A. (2020). When machine learning meets medical world: Current status and future challenges. *Computer Science Review*, 37: 100280. <https://doi.org/10.1016/j.cosrev.2020.100280>
- [7] Kolodner, J.L., Kolodner, R.M. (1987). Using experience in clinical problem solving: Introduction and framework. *IEEE Transactions on Systems, Man, and Cybernetics*, 17: 420-431. <https://doi.org/10.1109/TSMC.1987.4309058>
- [8] Bichindaritz, I. (1996). Mnaomia: Reasoning and learning from cases of eating disorders in psychiatry. *Proc AMIA Annu Fall Symp.*, 1996: 965.
- [9] Marling, C., Whitehouse, P. (2001). Case-based reasoning in the care of alzheimer's disease patients. *ICCBR 2001: Case-Based Reasoning Research and Development*, pp. 702-715. https://doi.org/10.1007/3-540-44593-5_50
- [10] Kwiatkowska, M., Atkins, M. (2020). Case representation and retrieval in the diagnosis and treatment of obstructive sleep apnea: A semio-fuzzy approach. *IEEE Annual Meeting of the Fuzzy Information, 2004. Processing NAFIPS '04*, Banff, Alberta, Canada. <https://doi.org/10.1109/NAFIPS.2004.1337383>
- [11] Bichindaritz, I., Kansu, E., Sullivan, K.M. (1998). Case-based reasoning in care-partner: Gathering evidence for evidence-based medical practice. *EWCBR 1998: Advances in Case-Based Reasoning*, pp. 334-345. <https://doi.org/10.1007/BFb0056345>
- [12] Lieber, J., Bresson, B. (2000). Case-Based Reasoning for Breast Cancer Treatment Decision Helping. In: Blanzieri E., Portinale L. (eds) *Advances in Case-Based Reasoning. EWCBR 2000. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, vol. 1898. https://doi.org/10.1007/3-540-44527-7_16
- [13] Sharaf-El-Deen, D.A., Moawad, I.F., Khalifa, M.E. (2014). A new hybrid case-based reasoning approach for medical diagnosis systems. *Journal of Medical Systems*, 38: 9. <https://doi.org/10.1007/s10916-014-0009-1>
- [14] Marling, C., Shubrook, J., Schwartz, F. (2008). Case-based decision support for patients with type 1 diabetes on insulin pump therapy. *Proc.9th Eur. Conf.Case-Based Reason: (ECCBR)*. Berlin, Germany: Springer Verlag, pp. 325-339. https://doi.org/10.1007/978-3-540-85502-6_22