Becalm: intelligent monitoring of respiratory patients

Juan A. Recio-Garcia, Belen Diaz-Agudo, and Arturo Acuaviva-Huertos

***Abstract*— The Becalm project is an open and low-cost solution for the remote monitoring of respiratory support therapies like the ones used in COVID-19 patients. It is a combined architecture based on a Case-Based Reasoning (CBR) decision-making system for the remote monitoring, detection, and explanation of risk situations for respiratory patients using a low-cost non-invasive mask.**

This paper describes the mask and the sensors that allow remote monitoring. Then, it describes the intelligent decision-making system that detects anomalies and raises early alerts that are visualized and explained to healthcare professionals. This detection is based on the comparison of cases that represent patients using a set of static variables, plus the dynamic vector of the patient time series from sensors.

The experiments reported in this paper are based on a synthetic data generator that simulates realistic patients using a synthesis process developed from the analysis of the available clinical literature. This process has been verified with real data and allows the validation of the reasoning system with noisy and incomplete data, threshold values, and life/death situations. Besides, we have evaluated three different distance metrics for the reasoning system in either optimal situations or cold-start and noisy situations. Our results demonstrate promising results and good accuracy for the proposed low-cost method to supervise COVID-19 patients for medical staff.

***Index Terms*— Case-based Reasoning, eXplainable Artificial Intelligence, Artificial Intelligence of Things, non-invasive respiratory device**

I. Introduction

In recent years there has been a substantial increase in proposals to apply Artificial Intelligence (AI) to the healthcare industry [1], [2]. This trend has been intensified by the recent events arising from the outbreak of the COVID-19 crisis [3]–[5]. COVID-19 patients in the hardest pandemic time could not be cared for in hospitals with adequate technical means, and healthcare professionals had not been able to do individualized supervision [6], [7].

The need, the high cost, and the difficulties for this individualized supervision are motivation aspects behind the Becalm project: an open project promoted by a non-profit organization after the COVID-19 pandemic[1]. BeCalm aims to provide an open and low-cost AI solution for the monitoring of oxygen therapies for respiratory patients when human healthcare resources are not fully available, for example, in developing and underdeveloped countries, rural areas, or overwhelmed healthcare infrastructures.

Fig. 1: The Becalm mask is a low-cost solution for monitoring COVID-19 patients

Becalm is based on open-source software contributions and widely available components. The Becalm mask (Figure 1) is an Internet of Things (IoT) device consisting of a non-intrusive ventilator mask, a pulse oximeter, a CO_2 concentration meter, and a pressure sensor. This mask can use both local or wide area network protocols to transmit the information provided by the sensors to a central point of control that allows the remote monitoring by healthcare staff of hundreds of patients at the same time without the need to move and be physically close to them. This individualized medical supervision is essential for fast assistance in risky situations. The centralized remote monitoring system is based on an intelligent decision-making system able to raise early alerts to healthcare professionals.

This paper describes the implementation of the decision-making system for the remote monitoring and detection of respiratory problems for patients using the Becalm mask through Case-Based Reasoning (CBR). CBR is an experience-

All authors belong to the Universidad Complutense de Madrid (UCM). Prof. Diaz-Agudo and Recio-Garcia are associated to the Institute of Knowledge Technology of the UCM (e-mails: jareciog@ucm.es, belend@ucm.es, arturacu@ucm.es)

[1]IDATIS. https://www.idatis.org/proyecto-becalm/

based approach to solving new problems by adapting previously successful solutions to similar problems [8]. There are previous applications of CBR for the detection and diagnosis of COVID-19, as well as other infectious diseases [9]–[11]. These existing models are based on cases that are encoded numerically or categorically using vector attributes representing disease symptoms. This way, the Becalm decision system combines static descriptive variables and dynamic values from sensors and manages the information in the form of time series, such as oscillations in peripheral oxygen saturation or pulsations in a certain window of time.

One advantage of our CBR proposal is that it provides a general scheme that can be integrated with different sets of clinical IoT sensors, not only for COVID-19 patients but also for other contexts and applications. Although in this paper, we will focus on the monitoring of respiratory anomalies caused by COVID-19, the CBR technique is easily applicable to other situations in which static clinical data assembles with time series from sensors.

The use of CBR has the additional benefit of providing explanations about the causes of the alert that other black-box decision-making models, such as neural networks, are not able to provide. There is a huge body of recent works on CBR demonstrating its ability to provide an explanation, transparency, and trust of the reasoning results based on the underlying examples, particularly in the healthcare domain [12], [13].

Another contribution of this paper arises from an insurmountable situation: the lack of real patient data to experiment and validate the research. It was impossible to get a proper real data set from patients, mainly because the mask device was improvised during the worst days of the pandemic due to the current emergency medical situation and the limited resources. The experiments reported in this paper are based on a synthetic data generator (described and evaluated in Section III) that simulates realistic patients using a synthesis process developed from the analysis of the available clinical literature. This process is verified through the comparison to a real data set and has the clear advantage of allowing us to extensively validate and test the alert and explanation systems with limit values and life/death situations.

Section II introduces the Becalm mask, its IoT components, and the type of data and parameters collected. Section III describes the synthetic data generator that we have developed to be able to simulate the cases collected using the Becalm mask. Section IV describes the proposed CBR approach to monitor and detect situations of patients that are potentially at risk of suffering some deterioration of the respiratory disease. Section V presents an empirical evaluation of the CBR system. Section VI focuses on the explainability of the monitoring process to healthcare professionals, and Section VII concludes the paper and outlines some lines of future work.

II. Becalm mask

The Becalm mask is an IoT solution consisting of a non-intrusive ventilator (NIV) device using a low-cost commercial scuba diving mask, a pulse-oximeter, a CO_2 concentration meter, and a pressure sensor. The project also integrates a simple Raspberry Pi computer board to manage hardware resources and transmit the data to a central server. Figure 2 shows all the components required to build the mask. All these components are widely available or can be 3D printed. This paper focuses on the software CBR solution for the intelligent monitoring of patients. In this section, we provide a brief functional description of its IoT components, whose technical details can be found at the project's open repository [2].

Fig. 2: Components of the Becalm mask

- **Non-intrusive ventilation**. The fan mask has been adapted from a scuba broadly available diving mask. This mask has been used in other proposals to solve the problem of shortage of respirators during the crisis caused by the COVID-19 pandemic [14], [15].
- **Pressure sensor**. The most vital measurement during monitoring is the mask internal pressure (MIP) which provides details on the correct operation of the mask, as well as possible problems that the patient may experience. The pressure monitoring in the Becalm mask is provided by the BMP280 sensor developed by BOSCH[3]. This sensor is capable of measuring atmospheric pressure, as well as temperature. Its small size and low consumption make it ideal for application in non-intrusive respirator masks.
- **Pulse Oximeter and Heart-Rate Sensor**. The pulse oximeter is a medical device that allows peripheral oxygen concentration (SpO_2) to be monitored non-intrusively. Concretely, the Becalm mask uses the MAX30100 Oximeter and Heart-Rate Sensor[4]. It is also capable of indicating the heart rate and pulse of the patient to be monitored. The value of peripheral oxygen

[2]https://github.com/idatis-org/becalm-station

[3]https://www.bosch-sensortec.com/products/environmental-sensors/pressure-sensors/bmp280/

[4]https://www.maximintegrated.com/en/products/sensors/MAX30100.html

saturation can alert clinicians to improperly assisted respiration, while heart rate can indicate a patient's stability.

- **Air quality sensor**. During monitoring, it is vitally important to ensure that not only the vital signs seem adequate for a patient, but also to be able to guarantee the quality of the air that they are consuming. Thus, we use carbon monoxide (CO_2) concentration measurements capable of estimating air quality. More specifically we use the Adafruit CCS811 gas sensor[5], that is capable of monitoring air quality in fast cycle times and with low energy consumption.

III. Synthetic patient data generator

Becalm experimentation requires evaluating the system with limit values and life/death situations. A collateral contribution of this paper arises from the requirements of evaluating and experimenting with different computational and representation techniques for patient data. Although some testing data have been collected using the Becalm respiration mask, our goal is to be able to obtain a sufficiently complete and extensive database on which to perform experimental tests. In this section, we describe a procedure to realistically generate an artificial database that simulates patients and critical situations to be able to evaluate the system performance within different techniques and algorithms emulating heterogeneous situations.

The goal of this generation process is to be able to characterize how monitored parameters can vary for the subjects in real-time (e.g. peripheral oxygen saturation or pressure in the mask), simulating scenarios of rest, suffocation, hyperventilation, or other conditions.

The following subsections first summarize the patient variables and values obtained from the sensors of the Becalm monitoring system (described in Section II), introducing a mathematical model that formalizes it. Subsequently, we introduce the strategy used to generate a time series simulating the hypothetical evolution of patients, and its validation through the comparison to a real dataset.

A. Variables

Becalm mask allows the monitoring of different parameters related to dynamic spirometric factors of the patient such as mask pressure, CO_2 emission, or peripheral oxygen saturation (SpO_2). Additionally, we consider other descriptive static variables of the patient: age, height, sex, and tobacco consumption. These parameters, described below, will be grouped into a *feature vector* that represents the state of a patient.

1) Static descriptive variables:

- **Height**. According to existing literature [16], we can model it as a normal distribution ($Male \sim N[171.0, 5]$, $Female \sim N[159.5, 5]$ cm).
- **Age**. We consider that it follows a discrete uniform distribution within range, $Age \sim U(16, 65)$ [16].
- **Sex**. An evenly distributed binary variable.
- **Tobacco consumption**. Although this variable may be correlated to the previous ones, we simplify its representation as a binary value distributed according to a 20:100 ratio (smokers/non-smokers) based on averaging existing studies [17].

2) Time series features: While using the mask the sensors capture data that are recorded in time series using different time windows. The CBR system reasons with knowledge about the evolution of these dynamic variables such as the heart rate or the pressure inside the mask. Each case represents one patient and the different time series characterize the evolution of this patient.

- **Resting heart rate**. A normal resting heart rate for adults lies somewhere between 60 and 100 beats per minute (bpm), and varies based on age group and gender [18]–[20]. Therefore, we will consider a normal distribution $RHR \sim N(75, 4)$ bpm.
- **Carbon dioxide inside the mask**. Average values of indoor CO_2 are between 300-400 ppm (parts per million), being 1000 ppm its limit to avoid physical damages [21]. Additionally, there is some consensus in the literature about hypoventilation and breathing resistance in ventilators, the latter being able to contribute to the rebreathing of CO_2 and thus increasing the concentration of the gas in the body to symptomatic levels [22]. Thus, considering the possible impact of the mask on respiratory function, we will set the possible resting carbon dioxide values of a subject following a normal distribution $CO_2 \sim N(450, 10)$ ppm.
- **Peripheral oxygen saturation**. SpO_2 is defined within range $(0, 100]$ and represents the percentage of oxygen-carrying hemoglobin in the blood relative to the total amount of hemoglobin. This variable allows estimating if the respiratory behavior is being adequate. According to the available literature, we can define the following reference distributions: $Ref_{SpO_2}(male) \sim N(97.1, 0.5)$ and $Ref_{SpO_2}(female) \sim N(96.6, 0.5)$ [23]–[32]. However, these studies also point out that the peripheral oxygen saturation depends on a multitude of variables, such as age, race, physical condition or smoking, being age and tobacco consumption the most relevant. Age is correlated to the decrease of oxygen saturation, whereas tobacco consumption may represent a significant variation. Therefore, we modify the previous normal distribution as it follows:

$$SpO_2 = Ref_{SpO_2}(sex) - \frac{age}{maxAge} - itc_{SpO_2}(sex) \quad (1)$$

  where the impact of tobaco consumption depends on the sex: $itc_{SpO_2}(male) \approx -0.7\%$ and $itc_{SpO_2}(female) \approx -0.4\%$ for smoker patients [31], [32] and $itc_{SpO_2} = 0$ otherwise. Additionally, $\frac{age}{maxAge}$ represents the impact of age in the oxygen saturation.
- **Mask Internal Pressure**. This variable ranges from 100500 y 101500 Pascals according to experimentation using the Becalm device and unfortunately lacks standardized reference values. However, it is possible to correlate it with the forced vital capacity (FVC) which is the amount of air that can be forcibly exhaled. Again, this variable is influenced by other parameters such as age, height, tobacco consumption or sex [33], [34]. From

[5]https://learn.adafruit.com/adafruit-ccs811-air-quality-sensor

Variable	Interval	Variability	Source
RHR	$[60, 100]$	$N(0, 2)$	[36]
SpO_2	$[93, 100]$	$N(0, .75)$	[37]
CO_2	$[300, 900]$	$N(0, 7)$	[21]
MIP	$[100900, 101100]$	$N(0, 10)$	[33], [34]

TABLE I: Configuration parameters for the reference series.

Variable	Decrement Interval	Increment Interval	Variability
RHR	$[30, 80]$	$[80, 130]$	$\Gamma(2, 2)$
SpO_2	$[85, 90]$	-	$\Gamma(2, .3)$
CO_2	$[10, 400]$	$[600, 1000]$	$\Gamma(2, 7)$
MIP	$[100800, 101000]$	$[101000, 101200]$	$\Gamma(5, 10)$

TABLE II: Configuration parameters for the incremental and decremental series.

the analysis of existing reports about its influence we can estimate FVC as:

$$FVC = li(age, sex) - itc_{FVC} \tag{2}$$

where the $li()$ function is the linear interpolation of the tabular data reported in [35] and itc_{FVC} represents it the impact of tobacco consumption (in our setup $itc_{FVC} = 0.1$). Once the FVC is estimated we can correlate the mask internal pressure using a transformation function that re-scales the FVC value from its original domain to the MIP range:

$$MIP = transform(FVC) \tag{3}$$
$$transform : [2.8, 5.2] \rightarrow [100500, 101500]$$

B. Temporal series generation

As we have described, each case records variables that evolve over time during the monitoring of each patient. Therefore, we will assume that these values should be stable when there is no medical emergency, but define a marked variation in the other case. We generate synthetic time series to simulate the use of the mask, focusing on generating appreciable variations in the values of the variables monitored by the Becalm mask sensors.

1) Reference series: are those temporal series generated to simulate a non-danger state for the patient. The generation process is defined as follows:

1) Define the reference $[min, max]$ interval of the variable. For example, the heart rate variable ranges from 60 to 100 bpm as defined in the previous section.
2) Define the initial value p_0 and the length of the series n.
3) Generate the following element with:

$$p_i = p_{i-1} + S + V + J$$

where (S)inusoidal reproduces a smooth repetitive oscillation $S \sim sin(i)$, and (V)ariability follows a normal distribution $V \sim N(\mu, \sigma)$. Finally, (J)ittering adds random noise. In case $p_i \notin [a, b]$ that value is discarded and a new point is generated.

Table I presents the configuration parameters used to generate the reference series for each variable.

2) Anomalous series: simulate the scenario of a possible failure of the organism. We simulate a marked variation in some of the variables. The time series reporting such anomalies will include values outside the resting range defined in the previous subsection. These anomalies can be identified as values above the upper limit (series that we will call *incremental*) or below the lower limit (*decremental* series). To reproduce such anomalous series with substantial incremental or decremental variations, we can use a gamma distribution as follows:

1) Define the reference $[min', max']$ anomalous interval of the variable. For example, an anomalous heart rate above the reference range can be defined as $[100, 130]$.
2) Define the initial value p_0 an length of the series n.
3) Generate the following element with

$$p_i = p_{i-1} + S + t \cdot V + J$$

where $t = 1$ for incremental series and $t = -1$ otherwise. (S)inusoidal reproduces a smooth repetitive oscillation $S \sim sin(i)$. (V)ariability follows a gamma distribution $V \sim \Gamma(\alpha, \lambda)$. And (J)ittering is a random noise function. Once the data points are below the min' or above the max' anomaly limits we can apply the same procedure than regular series within range $[min', max']$.

Table II presents the configuration parameters used to generate the incremental or decremental series for each variable.

Parameters reported in Tables I and II have been inferred from the analysis of existing literature. Nevertheless, they can be re-configured according to the criteria of the medical expert to particularize the simulation to a custom segment of the population.

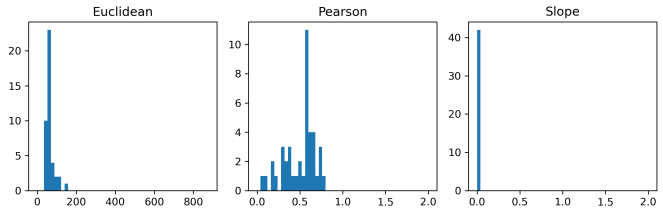
C. Dataset validation

In order to validate our synthetic time series to simulate the use of the mask, we have compared the resulting series to a real dataset. Although there are no validated datasets available for the $C0_2$ and MIP values, as they are completely dependent on the type of mask being used, we can find reference datasets for the heart rate and saturation variables. Concretely, we have chosen the BIDMC PPG and Respiration Dataset [38], which has been used to evaluate the performance of different algorithms for estimating respiratory rate. This dataset contains signals extracted from the much larger MIMIC II matched waveform Database [39], acquired from critically-ill patients during hospital care at the Beth Israel Deaconess Medical Centre (Boston, MA, USA).

The validation is based on the hypothesis that our synthetic method generates series that are very close to the real ones found in the BIDMC dataset. Therefore, we have compared the 53 recordings in BIDMC for the RHR and SPO_2 variables to a synthetic dataset generated through the method proposed in Section III.

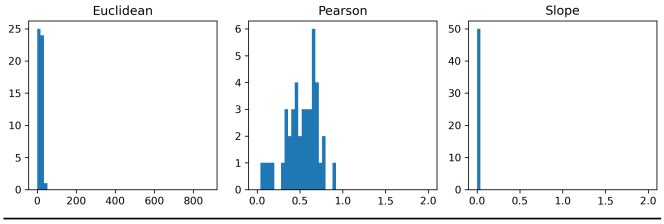
Results are shown in Figure 3 which shows the distribution of the similarity values between each record of the

(a) Resting Heart Rate



Sim. Function	Mean	Std	Min	25%	50%	75%	Max
Euclidean	65.32	21.14	43.66	52.79	59.08	67.01	152.13
Pearson	0.50	0.18	0.05	0.37	0.57	0.62	0.76
Slope	6.79e-5	7.91e-5	2.15e-7	1.38e-5	4.02e-5	9.63e-5	3.08e-4

(b) Peripheral oxygen saturation



Sim. Function	Mean	Std	Min	25%	50%	75%	Max
Euclidean	18.74	9.31	1.00	12.25	17.66	27.23	39.59
Pearson	0.52	0.19	0.04	0.42	0.53	0.66	0.89
Slope	1.04e-5	2.27e-5	5.82e-5	1.80e-6	4.25e-6	7.82e-6	1.47e-4

Fig. 3: Histogram distribution and descriptive statistics for the similarity values between each record of the BIDMC dataset and its most similar time series in the synthetic dataset (n=500) for the RHR (a) and SPO_2 (b) variables.

BIDMC dataset and its most similar time series from the synthetic one. For each variable, we display the minimum similarity according to the three functions used by the case-based reasoning decision-making system presented in the next section: Euclidean, Pearson, and Slope-comparison. As we can observe, the Euclidean distance -that will be later confirmed as the best similarity strategy in Section V- achieves very good results. Pearson similarity is not so good, probably due to the jittering effect of the time series, whereas slope (the baseline metric) does not provide significant results.

IV. Case Based Reasoning for Patient Monitoring

Case-based reasoning (CBR) is a well-known AI reasoning and learning paradigm that is inspired by memory-based human problem-solving in which instances of earlier problem-solving are remembered and applied to solve new problems. [40], [41] The fundamental assumption of CBR is that similar problems have similar solutions: a patient with similar symptoms will have the same diagnosis, or the price of a house with similar accommodation and location will be similar, and similar trials provide reusable decisions based on legal precedents [42]. In this way, specific experiences are memorized and later remembered and reused when similar situations arise. This approach contrasts with rule-based or theory-based problem-solving in which knowledge of how to solve a problem is applied. A doctor diagnosing a patient's symptoms may apply knowledge about how diseases manifest themselves, or she may remember a previous patient who demonstrated similar symptoms and thus apply a case-based

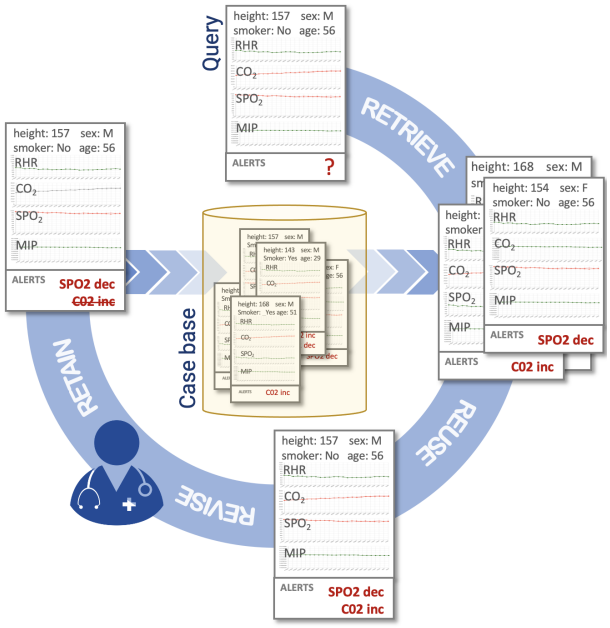


Fig. 4: General schema of the CBR process.

approach. One additional advantage of the CBR approach is the interpretability of results, as one or more previous patients are provided as sustainable evidence [12]. In the domain of this paper, the CBR assumption assumes that *similar* patient situations will have *similar* evolution. The similarity is a key aspect in CBR as a new unknown situation is compared with previous cases to predict their evolution and thus detect, alert and explain different anomalous situations. Cases in the case base encode the information referring to the clinical situation of the patient (as it was described in Section III-A), subsequently allowing us to understand the reason why the health status of a certain subject deteriorates or improves.

According to the available medical literature, it is common to find correlations when comparing clinical conditions between patients, and mostly intra-patients [43]–[45]. For this reason, when sufficient data is available, we will consider comparing a new case to be analyzed with past cases from other patients and previous states of the same.

The next subsection describes the representation of a case using static and dynamic variables, plus the vector of the time series. The rest of the sections cover the CBR process of the Becalm anomaly detection system. As Figure 4 illustrates, our CBR process follows the typical 4Rs stages [41] consisting of (1) an initial retrieval step that compares the query to the previous cases stored in the case base using a similarity function, (2) a reuse step where solutions from the most similar cases are combined to provide a solution to the given query, (3) an optional external revision of the proposed solution by a domain expert, and (4) the last step consisting on storing the confirmed new case (query+solution) in the case base.

A. Case representation

We define a case C_i for a patient i as a description D_i that groups numerical elements and time series referred to the

patient together with the *anomalies* (A) to be reported to the healthcare professional. It can be expressed as:

$$C_i = \langle D_i, A \rangle$$

The description of the patient encodes the variables defined in Section III-A:

$$D_i = \langle \overline{sdv}, \{RHR\}_{t=0}^{N}, \{CO_2\}_{t=0}^{N}, \{SpO_2\}_{t=0}^{N}, \{MIP\}_{t=0}^{N} \rangle$$

where $\overline{sdv}$ contains the static descriptive variables: height (h), age (a), sex (s), and smoker (m). $\{RHR\}_{t=0}^{N}$ is a series of N elements referring to the variation of the heart rate of the patient i -th in the last interval of specified time, $\{CO_2\}_{t=0}^{N}$ is a time series of elements that define the variation in the amount of carbon dioxide emitted by the patient in the last period of time, $\{SpO_2\}_{t=0}^{N}$ is the time series referred to the variation of the peripheral oxygen saturation and $\{MIP\}_{t=0}^{N}$ that of the variation in pressure within the mask. Note that in the previous definition, the concept of time period to which the time series refer to has not been specified on purpose. In general, you can define a time mesh based on seconds, minutes or hours, setting the number of intervals created as N.

The solution of the case encodes the class corresponding to the anomaly detected within the case. It can be any of the following values: ok (no anomaly detected), RHR_{inc} , RHR_{dec} , SpO_{2dec} , CO_{2inc} , CO_{2dec} , MIP_{inc} , MIP_{dec}.

B. Similarity based retrieval

Case retrieval computes the distance between a clinical query and the cases representing medical conditions for other patients. Next, we describe similarity metrics for each variable and how they are combined.

1) Case similarity and comparison: There are two numeric variables that refer to the value of height and age. Both values can be represented as natural numbers greater than zero, allowing a simple comparison between attributes using the usual Euclidean distance or any other analogous metric. For Boolean variables, we use equal similarity. More interesting is the comparison of temporal series. In the scientific literature referring to the study of data related to time series there are many different techniques [46]–[50]. However, among all of them, we will highlight the comparison using metrics derived from the L_p or Minkowski norm for its simplicity and its direct applicability to our case study. Note that even though the L_p is coordinate independent from the vector, the order in the series $\{x\}_i^n$ and $\{y\}_i^n$ determines the values of the distance. We will restrict ourselves to measurements of patients in the same time horizon so that it is not necessary to use techniques such as Dynamic Temporal Alignment or correction factors to adjust the series, and these can be compared directly. In turn, it is not necessary to apply any correction to the segments to align the series, since we will assume equispaced intervals that correspond to the same time horizon obtained for all patients.

To compute the similarity between cases we compute a similarity function between the different attributes of cases C_i and C_k:

$$\mathcal{L}(C_i, C_k) = \frac{\sum_{j \in V_i \cap V_k} sim(D_i^j, D_k^j)}{card(V)} \quad (4)$$

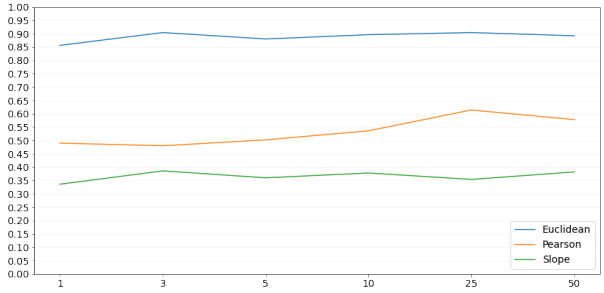


Fig. 5: Impact of k parameter. The three different similarity metrics are portrayed: euclidean distance (blue), Pearson correlation (orange) and slope comparison (green).

where V_i and V_k represents respectively the set of variables describing each case i and k, D_i^j and D_k^j are the values of the j-th attribute for each i-th and k-th description of the case, and $sim(\cdot, \cdot)$ is the similarity function that computes the distance for numerical variables and series. This approach is designed to effectively minimize learning over-generalization, and it is based on information retrieval techniques well-known in the field [51]–[53].

In particular, for our case study, we will compare three different distance metrics: Euclidean distance, Pearson correlation, and slope correlation. Being the latest, a basic metric based on the comparison of the slopes corresponding to the linear regression of the data series.

2) Case reuse: The following step in the CBR cycle is solution reuse. Here, we have defined a simple voting strategy that aggregates the anomalies $[A_1, \ldots, A_k]$ of the k-nearest neighbors of the query:

$$sv(A_1, \ldots, A_k) = \arg\max_x |\{x \in \cup_{i \in \{1,\ldots,k\}} A_i\}|. \quad (5)$$

The CBR process designed for the BeCalm mask and described in Figure 4 also includes a revision stage where a health professional confirms the outcome of the decision-making engine. This stage will be implemented once the platform is deployed in real scenarios. Finally, the last stage of the CBR cycle is the retention of the new solved case in the case base.

Once, we have defined our CBR process, the following section presents its experimental evaluation

V. Experimentation

In order to test the different strategies we run an experiment against a synthetically spawned sample of 5000 random patients. These simulations were provided by the generator described in Section III and tagged with four different classes: non-danger state, anomalous state generated by CO_2 incremental series, and/or by pressure incremental series. Note that as part of the data preprocessing, we also performed feature scaling to ensure that variables with a broad range of values do not disproportionately contribute to the explanation. In particular, we applied min-max normalization over the different values.

	precision	*recall*	*f1*
CO_{2dec}	0.92	0.92	0.92
CO_{2inc}	0.90	0.97	0.93
RHR_{dec}	0.97	0.95	0.96
RHR_{inc}	0.95	0.88	0.91
MIP_{dec}	0.98	0.93	0.95
MIP_{inc}	0.94	0.93	0.94
ok	0.60	0.73	0.66
SpO_{2dec}	1.00	0.93	0.96
accuracy			0.91
average	0.92	0.91	0.91

TABLE III: Global results using 10% hold-out cross validation and the Euclidean similarity metric (k=3)

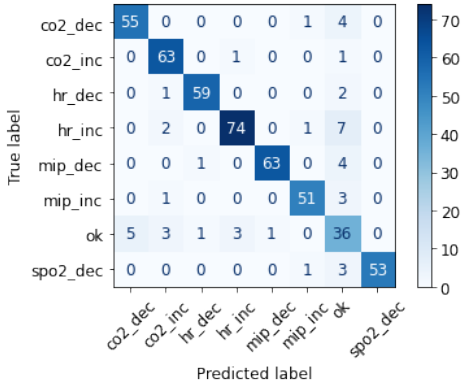


Fig. 6: Confusion matrix corresponding to Table III: 10% hold-out cross validation and Euclidean similarity metric (k=3).

Experimentation has been structured into three phases corresponding to the analysis of (1) optimal configuration, (1) cold-start behavior, and (2) noise tolerance. All these experiments evaluated performance using 10% Hold-out cross-validation.

A. Optimal configuration

The first analysis corresponds to the performance of the three similarity metrics according to the k parameter. This evaluation aims to find the optimal configuration of our system. Accuracy results are displayed in Figure 5, where we can observe that Euclidean distance is clearly the most efficient similarity strategy, being $k = 3$ the optimal configuration for the number of retrieved neighbors, as higher values do not report significant improvements. Pearson achieves an accuracy close to 60% and slope-based similarity does not obtain relevant results.

Following the previous results, we performed an in-deep analysis of the performance with the optimal configuration ($k = 3$ and Euclidean distance). Table III reports the performance of the CBR system according to standard evaluation metrics (precision, recall and f1-score). As we can observe, the average accuracy is .91, a very good figure considering a theoretical baseline of 12.5% corresponding to a random 8-class classifier. As expected, most of the misclassifications correspond to the 'ok' class, leading to false negatives that must be overcome through human supervision.

B. Cold-start behaviour

The next evaluation simulates the learning process of our CBR system, analyzing its behavior as the case base grows by adding new randomly generated cases. This evaluation is aimed to evaluate the cold-start scenario when few cases are available.

This way we begin with 10 patients belonging to each class, ending up with a case base of 500 patients. Figure 7 shows the evolution of accuracy against the addition of new cases for the three proposed similarity approaches. Results show that the optimal configuration using Euclidean distance is able to achieve acceptable performance in the cold start situation when there are few similar cases available in the case base. After the cold start situation, the experiments show how the model performance quickly stabilized.

C. Noise tolerance

Following evaluation analyses tolerance to missing or incorrect sensor readings. This analysis is quite significant as the BeCalm mask has been designed to be used remotely, by sending sensor readings to the centralized monitory repository that uses the case-based anomaly detection system presented in this paper. Therefore, this distributed architecture may have problems of incorrect or missing sensor readings, together with problems derived from the transmission of these readings using either local or wide area network protocols. This way, the CBR system must be tolerant to this kind of errors.

To evaluate the impact of noisy readings in the performance we have simulated two different types of errors:

- Missing readings. This error simulates that readings are not received by the centralized monitoring system due to either sensor or transmission failures. In this case the system uses the latest received value: $p_i = p_{i-1}$.
- Erroneous readings. In this scenario sensor readings are incorrect and do not reflect the actual value of the variable. This type of error is usually rooted in an incorrect sensor reading, as transmission protocols include redundancy checks to avoid accidental changes to raw data. To simulate this error we use a random value within the data range of the time series: $p_i = random()$

Figure 8 reports the accuracy of our anomaly detection system with levels of noise starting from the 75% of the raw data. This figure shows that our case-based approach is very tolerant of missing values. This is a coherent result as our methods are designed to identify the tendencies of time series, and in this case, the previous value is used when no reading has been received. Here missing values do not change the tendency of the series because it keeps stable until a new value is received. On the other hand, erroneous readings do change the tendencies and therefore the evaluation reports a very poor performance.

VI. Interpretability of the Case-based Anomaly Prediction System

In the domain of this paper, it is extremely important to explain the system's predictions to be able to trust them.

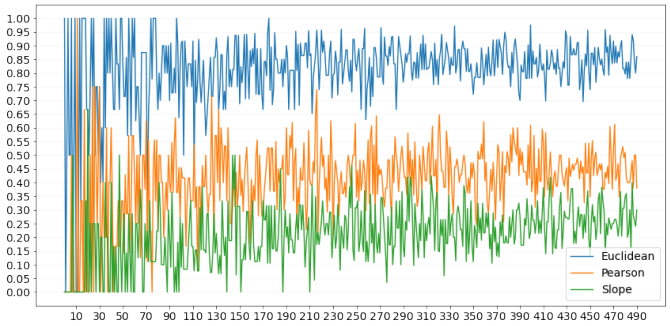


Fig. 7: Visual plot analyzing the learning process of the CBR system. The three different approaches for the similarity computation are portrayed: euclidean distance (blue), Pearson correlation (orange) and slope comparison (green). Y-axis represents accuracy.

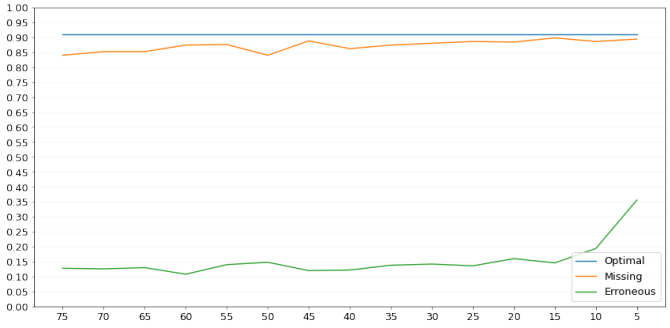


Fig. 8: Noise evaluation. X-axis represents the level of noise (percentage of raw data) that was introduced in the time series either simulating missing or erroneous readings. The optimal (noise-free) accuracy is also portrayed. Y-axis represents accuracy.

Especially because healthcare professionals make a decision based on the predictions and the results of this decision concern human safety. A machine learning model is better *interpretable* than another model if its decisions are easier for a human to comprehend than decisions from the other model. *Interpretability* can be defined as the degree to which a human can consistently predict the model's result [54]. The higher the interpretability of a machine learning model, the easier it is for someone to comprehend why certain decisions, or predictions, have been made.

At a high level, the literature distinguishes between two main approaches to interpretability:

- *model-specific* (also called transparent or white box) models
- *model-agnostic* (post-hoc) surrogate models to explain black box models [55].

CBR is typically considered as an interpretable model [56] as similar cases can be used as supporting evidence and, then, as explanations themselves. Using cases to explain would be *model-specific* explanation method as cases are used as explanations. Some other approaches, separate the explanations from the machine learning model (= model-agnostic methods) and use CBR as a post-hoc explanation model. For example, in [57] authors make a review of the twin-systems post-hoc explanation-by-example method. In twin systems, a black-box model is explained by "twinning" it with a more interpretable CBR system, by mapping the feature weights from the former to the latter. When reviewing the eXplainable Artificial Intelligence (XAI) literature we also find out that many approaches recognize that the context of an explanation situation and the goal of the user in that situation, influence what is and what is not a good explanation [56]. Our approach to explanation aims to generate personalized visual reports to explain what the data from the sensors tell us about a patient and their context, patterns in data, how a patient evolves in time, how they relate with its context, and why this patient is similar or different regarding others. We propose to personalize the explanation, using visual plots that are better than others for a particular goal (temporal evolution, current state, comparison with other patients), for the type of user – patient, doctor, or data scientist– and for the type of data that we want to explain.

Explanation reports can be personalized using different user profiles. The doctor's explanation report contains the following information:

- The query (patient) and the retrieved similar case data. The report relates the patient with other patients already diagnosed who present a similar clinical state.
- A visual representation of the comparison of the most significant variables used to conclude that the patients were similar. This can range from bar charts to dot graphs, depending on the attribute type compared.
- A textual explanation for the anomalies detected, generated from adjusting the case description and the types of alarm from the already classified past case in the knowledge base.

To illustrate this process we have created a UI prototype that exemplifies how detected anomalies are explained to health professionals. Figure 9 contains a screenshot of this application. It has been designed as a responsive web-app to be executed remotely using different types of devices. The explanation report shows an example where patient 10 (63 years old non-smoker and 177 cm tall) is motorized and our case-based anomaly detection system identifies two patients with a similar evolution. Therefore, the anomalous series are highlighted (red color) together with the previous evolution of these similar patients (light red). Additionally, the system generates a textual description of the anomalies being detected and proposes a response action. This suggested action is computed using the type of anomaly assigned to the retrieved past case when the case was retained in the CBR learning cycle.

VII. Conclusions and Future Work

In this paper, we have introduced the Becalm project as a low-cost solution for monitoring COVID-19 patients. We have described the mask and the sensors that allows remote monitoring. An intelligent decision-making system based on Case Based Reasoning (CBR) is able to raise early alerts to the healthcare professionals and generate personalized explanations. One advantage of our CBR proposal is that it provides

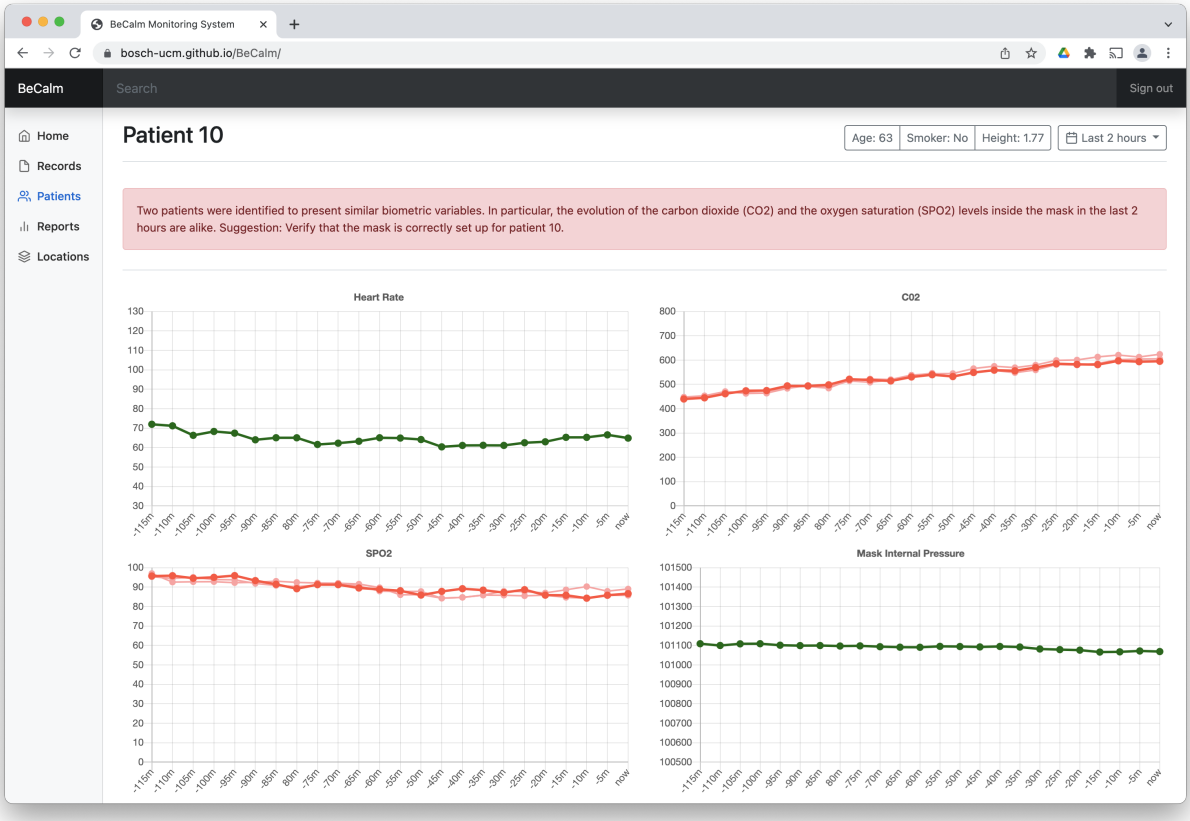


Fig. 9: User interface generated to assist the physician when diagnosing a new patient. The evolution of the patient being diagnosed is depicted using line charts over a time period. In case anomalies are detected, the anomalous time series corresponding to the retrieved past cases which suggest what issue might be facing the patient is also shown. A textual post-hoc explanation is also automatically generated.

a general scheme and allows an automatic and autonomous process of monitoring patients in real-time. In this paper, we have specifically described and evaluated an application for monitoring of respiratory anomalies caused by COVID-19. However, the CBR technique is novel, reusable, and easily applicable to other situations in which static clinical data assembles with time series from sensors. Cases in the CBR system combine static descriptive variables, and dynamic values from sensors, and manage the information in the form of time series, such as oscillations in peripheral oxygen saturation or pulsations in a certain window of time.

One of the greatest challenges in CBR systems is similarity computation. We have proposed three different metrics and evaluated their performance to obtain an optimal configuration of the system. We also have evaluated the behavior of the system in cold-start scenarios, when few cases are available, showing that the proposed method is suitable. Finally, as BeCalm has been designed to work remotely we have tested the noise tolerance of our anomaly detection system. Results demonstrate that this approach is able to manage noisy signals without losing significant performance.

As future work, we will apply user groups or profiles to make specific comparisons inter and intra-groups. For example, the system could store data referring to the elderly, marking this profile and allowing an instance to be compared only with the data of a certain group.

The use of CBR has the additional benefit of providing explanations about the causes of the alert. We have described the explanation system and illustrated it with examples, but as future work, we have planned a complete evaluation of the quality of the generated explanations with healthcare professionals. Additionally, we will compare its performance with other Machine Learning techniques such as Artificial Neural Networks. This comparison will let us determine the benefits of the performance-explainability trade-off achieved by the CBR approach. Also as future work, BeCalm could be validated with respect to reference medical devices, following a well-defined test protocol and deriving the metrological performance of the whole system, for each of the measured quantities. Such a consideration could be added here.

VIII. Appendix

The complete source code, datasets, and evaluation code of the case-based anomaly detection system for the BeCalm mask presented in this paper, together with the prototype of the explanation UI, are available at: `https://github.com/BOSCH-UCM/BeCalm`

Acknowledgments

This research is a result of the Horizon 2020 Future and Emerging Technologies (FET) programme of the European Union through the iSee project (CHIST-ERA-19-XAI-008, PCI2020-120720-2) funded by MCIN/AEI/10.13039/501100011033 and European Union "NextGenerationEU"/PRTR". This work has been also partially supported by the PERXAI project PID2020-114596RB-C21 funded by MCIN/AEI/10.13039/501100011033 and the BOSCH-UCM Honorary Chair on Artificial Intelligence applied to Internet of Things (https://www.ucm.es/catedrabosch).

The Becalm Mask project has been promoted and developed by Enrique Melero and Felipe Santi in collaboration with IDATIS.

References

[1] J. He, S. L. Baxter, J. Xu, J. Xu, X. Zhou, and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine," *Nature Medicine*, vol. 25, no. 1, pp. 30–36, Jan. 2019. [Online]. Available: https://doi.org/10.1038/s41591-018-0307-0

[2] N. Zahid, A. H. Sodhro, U. R. Kamboh, A. Alkhayyat, and L. Wang, "Ai-driven adaptive reliable and sustainable approach for internet of things enabled healthcare system," *Mathematical Biosciences and Engineering*, vol. 19, no. 4, pp. 3953–3971, 2022. [Online]. Available: https://www.aimspress.com/article/doi/10.3934/mbe.2022182

[3] M. M. Islam, T. N. Poly, B. Alsinglawi, L.-F. Lin, S.-C. Chien, J.-C. Liu, and W.-S. Jian, "Application of artificial intelligence in COVID-19 pandemic: Bibliometric analysis," *Healthcare*, vol. 9, no. 4, p. 441, Apr. 2021. [Online]. Available: https://doi.org/10.3390/healthcare9040441

[4] M. M. Rahman, F. Khatun, A. Uzzaman, S. I. Sami, M. A.-A. Bhuiyan, and T. S. Kiong, "A comprehensive study of artificial intelligence and machine learning approaches in confronting the coronavirus (COVID-19) pandemic," *International Journal of Health Services*, May 2021. [Online]. Available: https://doi.org/10.1177/00207314211017469

[5] M. S. Al-Rakhami, M. M. Islam, M. Z. Islam, A. Asraf, A. H. Sodhro, and W. Ding, "Diagnosis of covid-19 from x-rays using combined cnn-rnn architecture with transfer learning," *medRxiv*, 2021. [Online]. Available: https://www.medrxiv.org/content/early/2021/08/09/2020.08.24.20181339

[6] U.S. Department of Health and Human Services, "Hospital experiences responding to the covid-19 pandemic," April 2020. [Online]. Available: https://oig.hhs.gov/oei/reports/oei-06-20-00300.pdf

[7] Z. Zhuang, P. Cao, S. Zhao, L. Han, D. He, and L. Yang, "The shortage of hospital beds for COVID-19 and non-COVID-19 patients during the lockdown of wuhan, china," *Annals of Translational Medicine*, vol. 9, no. 3, pp. 200–200, Feb. 2021. [Online]. Available: https://doi.org/10.21037/atm-20-5248

[8] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI Communications*, vol. 7, pp. 39–59, 08 2001.

[9] O. N. Oyelade and A. E. Ezugwu, "A case-based reasoning framework for early detection and diagnosis of novel coronavirus," *Informatics in Medicine Unlocked*, vol. 20, p. 100395, 2020. [Online]. Available: https://doi.org/10.1016/j.imu.2020.100395

[10] A. Smiti and M. Nssibi, "Case based reasoning framework for COVID-19 diagnosis," *Ingénierie des systèmes d information*, vol. 25, no. 4, pp. 469–474, Sep. 2020. [Online]. Available: https://doi.org/10.18280/isi.250409

[11] J. Duan and F. Jiao, "Novel case-based reasoning system for public health emergencies," *Risk Management and Healthcare Policy*, vol. Volume 14, pp. 541–553, feb 2021. [Online]. Available: https://doi.org/10.2147/rmhp.s291441

[12] J. A. Recio-García, B. Díaz-Agudo, A. Kazemi, and J. L. Jorro-Aragoneses, "A data-driven predictive system using case-based reasoning for the configuration of device-assisted back pain therapy," *J. Exp. Theor. Artif. Intell.*, vol. 33, no. 4, pp. 617–635, 2021. [Online]. Available: https://doi.org/10.1080/0952813X.2019.1704441

[13] A. Wijekoon, N. Wiratunga, K. Cooper, and K. Bach, "Learning to recognise exercises in the self-management of low back pain," in *Thirty-Third International Florida Artificial Intelligence Research Society Conference, USA*. AAAI Press, 2020, pp. 347–352. [Online]. Available: https://aaai.org/ocs/index.php/FLAIRS/FLAIRS20/paper/view/18460

[14] A. Noto, C. Crimi, A. Cortegiani, M. Giardina, F. Benedetto, P. Princi, A. Carlucci, L. Appendini, and C. Gregoretti, "Performance of easybreath decathlon snorkeling mask for delivering continuous positive airway pressure," *Scientific Reports*, vol. 11, no. 1, p. 5559, Mar 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-85093-w

[15] R. Fuentes-Alvarez, M. Alfaro-Ponce, F. Alvarado, J. A. Mora-Galvan, R. Q. Fuentes-Aguilar, and I. Matehuala Moran, "Full-face mask adapter as covid-19 personal protective equipment: design, tests and validation," *IEEE Latin America Transactions*, vol. 19, no. 6, p. 986–993, Jun. 2021. [Online]. Available: https://latamt.ieeer9.org/index.php/transactions/article/view/4373

[16] C. A. Max Roser and H. Ritchie, "Human height," *Our World in Data*, 2013, https://ourworldindata.org/human-height.

[17] [Online]. Available: https://worldpopulationreview.com/country-rankings/smoking-rates-by-country

[18] J. W. Mason, D. J. Ramseth, D. O. Chanter, T. E. Moon, D. B. Goodman, and B. Mendzelevski, "Electrocardiographic reference ranges derived from 79,743 ambulatory subjects," *Journal of electrocardiology*, vol. 40, no. 3, pp. 228–234, 2007.

[19] W. B. Kannel, C. Kannel, R. S. Paffenbarger Jr, and L. A. Cupples, "Heart rate and cardiovascular mortality: the framingham study," *American heart journal*, vol. 113, no. 6, pp. 1489–1494, 1987.

[20] R. Avram, G. H. Tison, K. Aschbacher, P. Kuhar, E. Vittinghoff, M. Butzner, R. Runge, N. Wu, M. J. Pletcher, G. M. Marcus, and J. Olgin, "Real-world heart rate norms in the health eHeart study," *npj Digital Medicine*, vol. 2, no. 1, Jun. 2019. [Online]. Available: https://doi.org/10.1038/s41746-019-0134-9

[21] M. S. M. Rhee, C. D. Lindquist, M. T. Silvestrini, A. C. Chan, J. J. Y. Ong, and V. K. Sharma, "Carbon dioxide increases with face masks but remains below short-term niosh limits," *BMC Infectious Diseases*, vol. 21, no. 1, p. 354, Apr 2021. [Online]. Available: https://doi.org/10.1186/s12879-021-06056-0

[22] E. J. Sinkule, J. B. Powell, and F. L. Goss, "Evaluation of N95 Respirator Use with a Surgical Mask Cover: Effects on Breathing Resistance and Inhaled Carbon Dioxide," *The Annals of Occupational Hygiene*, vol. 57, no. 3, pp. 384–398, 10 2012. [Online]. Available: https://doi.org/10.1093/annhyg/mes068

[23] S. Levental, E. Picard, F. Mimouni, L. Joseph, T. Samuel, R. Bromiker, D. Mandel, N. Arish, and S. Goldberg, "Sex-linked difference in blood oxygen saturation," *The Clinical Respiratory Journal*, vol. 12, 12 2017.

[24] M. Özdal, Z. Pancar, V. Çınar, and M. Bilgiç, "Effect of smoking on oxygen saturation in healthy sedentary men and women," 2017.

[25] S. D. Rahmanian, K. L. Wood, S. Lin, M. A. King, A. Horne, S. Yang, H. M. Wu, and P. T. Diaz, "Gender differences in pulmonary function, respiratory symptoms, and macrophage proteomics among HIV-infected smokers," *Scientifica*, vol. 2014, pp. 1–9, 2014. [Online]. Available: https://doi.org/10.1155/2014/613689

[26] S. Shrestha, L. Shrestha, and N. Bhandary, "Oxygen saturation of hemoglobin in healthy children of 2- 14 years at high altitude in nepal," *Kathmandu University medical journal (KUMJ)*, vol. 10, pp. 40–3, 10 2012.

[27] E. Pomeroy, J. T. Stock, S. Stanojevic, J. J. Miranda, T. J. Cole, and J. C. Wells, "Associations between arterial oxygen saturation, body size and limb measurements among high-altitude andean children," *American Journal of Human Biology*, vol. 25, no. 5, pp. 629–636, Aug. 2013. [Online]. Available: https://doi.org/10.1002/ajhb.22422

[28] I. B. MASTERS, A. M. GOES, L. HEALY, M. O'NEIL, D. STEPHENS, and M. A. HARRIS, "Age-related changes in oxygen saturation over the first year of life: A longitudinal study," *Journal of Paediatrics and Child Health*, vol. 30, no. 5, pp. 423–428, Oct. 1994. [Online]. Available: https://doi.org/10.1111/j.1440-1754.1994.tb00693.x

[29] K. Lee, K. P. Hui, W. Tan, and T. Lim, "Factors influencing pulse oximetry as compared to functional arterial saturation in multi-ethnic singapore." *Singapore medical journal*, vol. 34 5, pp. 385–7, 1993.

[30] M. W. Sjoding, R. P. Dickson, T. J. Iwashyna, S. E. Gay, and T. S. Valley, "Racial bias in pulse oximetry measurement," *New England Journal of Medicine*, vol. 383, no. 25, pp. 2477–2478, Dec. 2020. [Online]. Available: https://doi.org/10.1056/nejmc2029240

[31] P. Tsai, A. Harris, L. Kagemann, S. Kresovsky, R. Dinn, B. Siesky, L. McCranor, and Y. Catoira, "Arterial oxygen saturation decreases with age," *Investigative Ophthalmology & Visual Science*, vol. 45, no. 13, pp. 2592–2592, May 2004.

[32] A. S. Bhogal and A. R. Mani, "Pattern analysis of oxygen saturation variability in healthy individuals: Entropy of pulse oximetry signals carries information about mean oxygen saturation," *Frontiers in Physiology*, vol. 8, Aug. 2017. [Online]. Available: https://doi.org/10.3389/fphys.2017.00555

[33] A. T. Barroso, E. M. Martín, L. M. R. Romero, and F. O. Ruiz, "Factors affecting lung function: A review of the literature," *Archivos de Bronconeumología (English Edition)*, vol. 54, no. 6, pp. 327–332, Jun. 2018. [Online]. Available: https://doi.org/10.1016/j.arbr.2018.04.003

[34] E. Dugral and D. Balkanci, "Effects of smoking and physical exercise on respiratory function test results in students of university: A cross-sectional study," *Medicine*, vol. 98, no. 32, pp. e16 596–e16 596, Aug 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31393359

[35] "Lung function results 2007 to 2009," Nov 2015. [Online]. Available: https://www150.statcan.gc.ca/n1/pub/82-625-x/2010001/article/11088-eng.htm

[36] J. Hart, "Normal resting pulse rate ranges," *Journal of Nursing Education and Practice*, vol. 5, no. 8, Jun. 2015. [Online]. Available: https://doi.org/10.5430/jnep.v5n8p95

[37] I. M. Alhyari, M. A. Alabadi, G. J. Hijazin, and F. T. Alasasfeh, "SPO2 vital sign: Definition, ranges, and measurements," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 8, no. 7, Jul. 2018. [Online]. Available: https://doi.org/10.29322/ijsrp.8.7.2018.p7945

[38] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2017.

[39] J. Lee, D. J. Scott, M. Villarroel, G. D. Clifford, M. Saeed, and R. G. Mark, "Open-access mimic-ii database for intensive care research," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 8315–8318.

[40] S. Craw, *Case-Based Reasoning*. Boston, MA: Springer US, 2010, pp. 147–154. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_97

[41] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI Commun.*, vol. 7, no. 1, pp. 39–59, 1994. [Online]. Available: https://doi.org/10.3233/AIC-1994-7104

[42] S. Craw, *Case-Based Reasoning*. Boston, MA: Springer US, 2017, pp. 180–188. [Online]. Available: https://doi.org/10.1007/978-1-4899-7687-1_34

[43] D. Cattaneo, C. Gervasoni, P. Meraviglia, S. Landonio, S. Fucile, V. Cozzi, S. Baldelli, M. Pellegrini, M. Galli, and E. Clementi, "Inter- and intra-patient variability of raltegravir pharmacokinetics in HIV-1-infected subjects," *Journal of Antimicrobial Chemotherapy*, vol. 67, no. 2, pp. 460–464, 11 2011. [Online]. Available: https://doi.org/10.1093/jac/dkr498

[44] A. W. Burton, M. Filbet, R. Tayi, M. S. Perelman, and A. D. Knight, "Intra-and inter-patient variability of baseline pain intensity scores during breakthrough pain in cancer." *Journal of Clinical Oncology*, vol. 30, no. 15_suppl, pp. e19 578–e19 578, 2012. [Online]. Available: https://doi.org/10.1200/jco.2012.30.15_suppl.e19578

[45] V. M. Lugo, M. Torres, O. Garmendia, M. Suarez-Giron, C. Ruiz, C. Carmona, E. Chiner, N. Tarraubella, M. Dalmases, A. M. Pedro, C. J. Egea, M. Abellana, M. Mayos, C. Monasterio, J. F. Masa, R. Farré, and J. M. Montserrat, "Intra- and inter-physician agreement in therapeutic decision for sleep apnea syndrome," *Archivos de Bronconeumología*, vol. 56, no. 1, pp. 18–22, Jan. 2020. [Online]. Available: https://doi.org/10.1016/j.arbres.2019.02.014

[46] S. Hirano and S. Tsumoto, "Multiscale analysis of long time-series medical databases," *AMIA Annu Symp Proc*, pp. 289–293, 2003.

[47] A. Kianimajd, M. Ruano, P. Carvalho, J. Henriques, T. Rocha, S. Paredes, and A. Ruano, "Comparison of different methods of measuring similarity in physiologic time series," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 11 005–11 010, 2017, 20th IFAC World Congress. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896317333967

[48] J. Alexander, R. A. Edwards, M. Brodsky, L. Manca, R. Grugni, A. Savoldelli, G. Bonfanti, B. Emir, E. Whalen, S. Watt, and B. Parsons, "Using time series analysis approaches for improved prediction of pain outcomes in subgroups of patients with painful diabetic peripheral neuropathy," *PLOS ONE*, vol. 13, no. 12, p. e0207120, Dec. 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0207120

[49] S. Hirano and S. Tsumoto, "Cluster analysis of time-series medical data based on the trajectory representation and multiscale comparison techniques," in *Sixth International Conference on Data Mining (ICDM'06)*. IEEE, Dec. 2006. [Online]. Available: https://doi.org/10.1109/icdm.2006.33

[50] C. E. Kennedy and J. P. Turley, "Time series analysis as input for clinical predictive modeling: Modeling cardiac arrest in a pediatric ICU," *Theoretical Biology and Medical Modelling*, vol. 8, no. 1, Oct. 2011. [Online]. Available: https://doi.org/10.1186/1742-4682-8-40

[51] R. Duda, *Pattern classification and scene analysis*. New York: Wiley, 1973.

[52] A. Montazemi and K. Gupta, "A framework for retrieval in case-based reasoning systems," *Annals of Operations Research*, vol. 72, pp. 51–73, 1997.

[53] T. Eskridge, "Principles of continuous analogical reasoning," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 1, no. 3, pp. 179–194, 1989.

[54] B. Kim, O. Koyejo, and R. Khanna, "Examples are not enough, learn to criticize! criticism for interpretability," in *Annual Conference on Neural Information Processing Systems 2016, 2016, Barcelona, Spain*, D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, Eds., 2016, pp. 2280–2288.

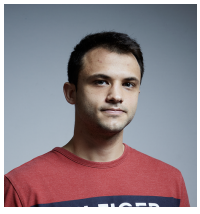
[55] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G. Yang, "XAI - explainable artificial intelligence," *Sci. Robotics*, vol. 4, no. 37, 2019. [Online]. Available: https://doi.org/10.1126/scirobotics.aay7120

[56] F. Sørmo, J. Cassens, and A. Aamodt, "Explanation in case-based reasoning-perspectives and goals," *Artif. Intell. Rev.*, vol. 24, no. 2, pp. 109–143, 2005. [Online]. Available: https://doi.org/10.1007/s10462-005-4607-7

[57] E. M. Kenny, C. Ford, M. S. Quinn, and M. T. Keane, "Explaining black-box classifiers using *post-hoc* explanations-by-example: The effect of explanations and error-rates in XAI user studies," *Artif. Intell.*, vol. 294, p. 103459, 2021. [Online]. Available: https://doi.org/10.1016/j.artint.2021.103459

Juan A. Recio-Garcia is Associate Professor at the Department of Software Engineering and Artificial Intelligence at the Complutense University of Madrid. Currently he holds the BOSCH-UCM Honorary Chair on Artificial Intelligence applied to Internet of Things. His research is focused on Case-Based Reasoning, Recommender Systems and eXplainable AI (XAI). He is lead investigator of several national-founded projects and has conducted several contracts with companies in the area of AI.

Belen Diaz-Agudo (Researcher ID: K-6046-2014. Orcid code: 0000-0003-2818-027X) is a professor at the Complutense University where she co-leads the GAIA (Group of Artificial Intelligence Applications) research group. Her research career is related to Case Based Reasoning, ontologies, Social recommenders and Explainable and Accountable AI systems. She is the leader of the CHIST-ERA project iSee - Intelligent Sharing of Explanation Experience by Users for Users https://isee4xai.com/.



Arturo Acuaviva-Huertos is a postgraduate student at Complutense University of Madrid (UCM). He holds a university degree in Mathematics and a degree in Computer Science. He has been collaborating with the Department of Software Engineering and Artificial Intelligence from the UCM as a Research Staff Support. His latest research has been focused on studying how to build AIoT systems.