Predicting ARDS Using the MIMIC II Physiological Database

TAOUM Aline[1,2], MOURAD-CHEHADE Farah[2], AMOUD Hassan[1], FAWAL Ziad[1]
[1] Azm platform for Research in Biotechnology and its Applications, EDST, Lebanese University, Tripoli, Lebanon
[2] Institut Charles Delaunay, ROSAS, LM2S, Université de Technologie de Troyes, UMR 6281, CNRS, Troyes, France

***Abstract*— Acute Respiratory Distress Syndrome (ARDS) is a critical lung condition occurring in ill patients. Like many other cardiac disorders, ARDS can be assessed by physiological measurements. This study aims to predict ARDS in hospitalized patients using only physiological signals as heart rate and breathing rate. An approach based on hypothesis testing is developed to detect whether subjects' signals deviate from their initial states. The approach is applied on mechanically ventilated subjects in the MIMIC II database. As results, a sensitivity going up to 85% is achieved, with a prediction remaining possible before 24 hours of ARDS occurrence.**

Keywords—Acute respiratory distress syndrome; conjunctive & disjunctive rules; hypothesis tests ; MIMIC-II database; prediction

I. INTRODUCTION

Patients in intensive care units (ICU) are vulnerable to many types of pathologies that degrade their health and may cause mortality. One of these pathologies is the Acute Respiratory Distress Syndrome (ARDS) [1], [2]. ARDS is caused by pulmonary edema that is the accumulation of fluid in the lungs preventing the exchange of oxygen and carbon dioxide. ARDS and Acute Lung Injury (ALI), a less severe form of ARDS, were defined by the American-European Consensus Committee in 1994 as they are acute conditions characterized by bilateral pulmonary infiltrates and severe hypoxemia. ARDS is diagnosed when the ratio of the partial pressure of oxygen in the patient's arterial blood (PaO_2) to the fraction of oxygen in the inspired air (FiO_2) is less than 200, whereas ALI is identified when the ratio is less than 300 [3].

ARDS is developed after a direct injury to the lungs as aspiration, trauma or pneumonia or an indirect injury to other parts of the body as sepsis or pancreatitis. Additionally, the onset of ARDS may be affected by initial ventilator settings [4]–[6]. Several studies have investigated the risk factors associated with ARDS. High tidal volume and high airway pressure are characterized as the most important risk factors associated with the onset of ARDS [7]. Boverman et al. have proved that severe sepsis is a principle stimulant of ARDS patients' mortality [8]. There are very few studies conducted on predicting ARDS. One of them is the work done by Ennett et al. They have developed a fusion algorithm to predict the development of ARDS in ICU patients [9]. Four rule sets were developed to obtain a decision system using clinical data such as plateau pressure (PP), pulse oximetry (SpO2) and heart rate (HR) from the MIMIC-II clinical database [10]. They obtained a sensitivity of 60% and a specificity of 82% on a single rule basis, and a sensitivity of 50% and a specificity of 90% for combined rules. These rules are obtained by a comparison to numerical adjusted thresholds. Another study has presented a method for detection of respiratory complications followed by an unexpected hospital death [11]. A Markov Model based algorithm was developed using physiological patterns extracted from respiration rate (RR) and pulse oximetry (SpO_2) signals. The proposed algorithm detects various levels of relevant patterns from signals before the Markov Model is executed. This approach has achieved a true positive rate of 92%, however it does not consider any specific pathology.

Our study aims to predict the development of ARDS prior to the critical event by studying the variability of heart rate and breathing rate measurements. These signals were used for the reason that they are non-invasive measurements which can be collected outside hospital units. The pathophysiological disorders of ARDS are much more complicated to be detected by pre-determined thresholds. The proposed method, based on the hypothesis testing principle, consists of detecting whether the signals diverge from their starting states or not. Disjunctive and conjunctive rules are then applied to combine information of both types of signals. The approach is tested on MIMIC-II that is a database of ICU records collected at BID Hospital, affiliated with Harvard Medical School, and post-processed and de-identified by researchers at the Massachusetts Institute of Technology (MIT) [10], [12]. Data in MIMIC-II are collected from the intensive care unit (ICU) from 2001 to 2008.

This paper is organized as follows. In the next section, we present the database and describe the procedure of patients' selection, and then we introduce our method. After that, the outcome of this work is displayed and interpreted in the results section. Finally, a conclusion summarizes the methods and the results obtained, and proposes possible perspectives.

II. MATERIALS & METHODS

A. The Multi-Parameter Intelligent Monitoring of Intensive Care-II Database (MIMIC-II)

MIMIC-II database consists of two types of data: bedside monitor waveforms and clinical data [10], [12]. The waveform database includes high resolution physiological signals such as electrocardiograms (ECG) and continuous blood pressure, as well as numeric records extracted from physiological measurements such as systolic and diastolic pressure, heart rate and respiratory rate. The clinical database was recorded from the CareVue Clinical Information Systems less frequently than

978-1-5090-5281-3/16/$31.00 ©2016 IEEE

bedside monitor data. It includes ventilator settings, laboratory results, ICD-9 codes, nursing progress notes, patient monitoring data, etc. Some of the records in the waveform database were matched with their records in the clinical database and formed a third database called MIMIC-II waveform Database Matched Subset.

B. Patients Selection

The patients' selection was done in order to distinguish between two groups of subjects, the "Stable" group composed of the subjects who did not develop the ARDS and the "Unstable" group composed of the subjects who started at a stable state and then developed the ARDS. Since the aim of the study is to predict ARDS, all subjects must have high-duration signals with the same situation of ventilation. To this end, the selected subjects for both groups were mechanically-ventilated for a duration higher than 48 hours, and did not have ARDS at the beginning of ventilation. As previously noted, the ARDS is diagnosed if the ratio P/F of the partial pressure of oxygen in the patient's arterial blood (PaO_2) to the fraction of oxygen in the inspired air (FiO_2) is less than 200. Therefore, we included mechanically ventilated subjects from the MIMIC-II clinical database. For these subjects, the P/F ratio was calculated. The "Stable" group is defined by subjects having the P/F ratio higher than 200 all over the signals recording, and the "Unstable" group is defined by subjects who started with a P/F higher than 200, then experienced a drop to less than 200. Then, we matched the selected subjects in both groups with the MIMIC-II waveform database, which leads to 18 "Stable" subjects and 38 "Unstable" ones. Figure 1 shows the applied criteria for patients' selection. Once the groups are defined, we extracted the signals records from the waveform database as heart rate and breathing rate.

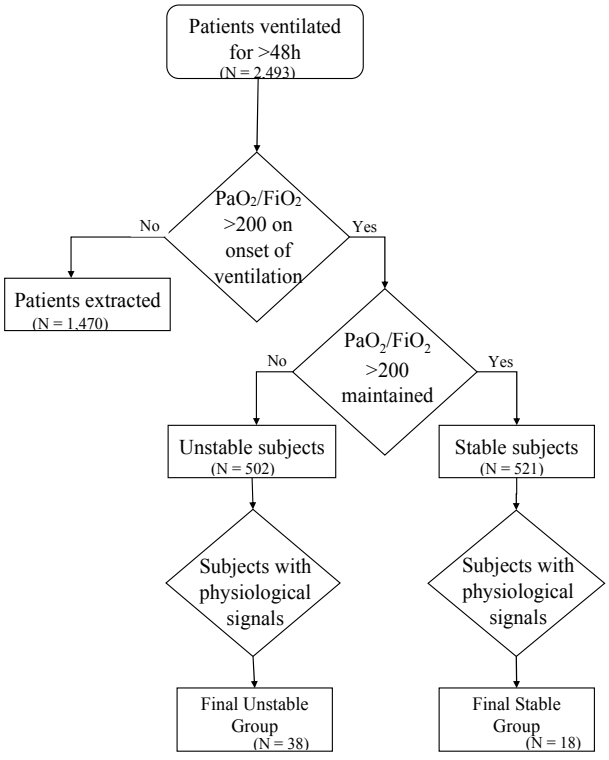

Figure 1: Patients' selection procedure

These signals have a frequency of one sample per minute. For Stable subjects, data were studied for the entire ventilation time, whereas for unstable subjects, data were studied until ARDS is detected (P/F < 200).

C. Statistical Analysis

Being a subject-centered approach, the proposed method aims to detect for each subject whether it deviates from its initial state. Let $\{X_i(k)\}_{k=1\dots K}$ denote the extracted signals from the database for a certain subject, with K being the length in minutes of the signals of this subject, and i denoting the signals of heart rate ($i = 1$) and breathing rate ($i = 2$). The Kolmogorov-Smirnov test is applied to check whether the data are normally distributed or not [13]. If the data are not normal, a Gaussianization phase is required, as described in [14]. In the following, X_i would refer to normally distributed data.

Now, to define the initial state of a subject, an initial segment of duration T is selected from each signal, with $T = 720$ minutes (12 hours) for instance. Statistical characteristics of these segments are then computed for $i = 1, 2$:

- The empirical mean $\mu_{0,i} = \frac{1}{T}\sum_{k=1}^{k=T} X_i(k)$,
- The empirical standard deviation $\sigma_{0,i} = \sqrt{\frac{1}{T}\sum_{k=1}^{k=T}\left(X_i(k) - \mu_{0,i}\right)^2}$.

A decision rule based on these parameters is constructed afterwards using statistical tests [15], [16]. Indeed, since the data are normally distributed, then 1% of the data is outside the interval $\left[\mu_{0,i} - 3\sigma_{0,i}, \mu_{0,i} + 3\sigma_{0,i}\right]$. This sets up the null hypothesis for each signal i:

$$H_{0,i}: Less\ than\ 1\%\ of\ data\ \notin \left[\mu_{0,i} - 3\sigma_{0,i}, \mu_{0,i} + 3\sigma_{0,i}\right].$$

Let $\{X_i(k)\}_{k=T+1\dots K}$ denote the remaining segment of the i^{th} signal. Then $\{X_i(k)\}_{k=T+1\dots K}$ is tested to verify if it satisfies the hypothesis $H_{0,i}$. Practically, the number of the points of $\{X_i(k)\}_{k=T+1\dots K}$ outside the interval $\left[\mu_{0,i} - 3\sigma_{0,i}, \mu_{0,i} + 3\sigma_{0,i}\right]$ is divided by the total number of points and the ratio is compared to 1%. If it is less than the threshold, the hypothesis $H_{0,i}$ is accepted and the subject is assumed stable; otherwise, $H_{0,i}$ is rejected and the subject is supposed unstable. Performance is evaluated using the classical sensitivity (Se) which measures the proportion of positives (unstable) that are correctly identified and specificity (Sp) which measures the proportion of negatives (stable) that are correctly identified.

Two other conjunctive and disjunctive fusion rules are also investigated. According to the first one, a subject is assumed to be unstable if he is considered unstable in the hypothesis testing of both signals. According to the second one, a subject is assumed unstable if he is unstable in one or the other of the hypothesis tests. If the first one increases the specificity of the test, the second one decreases it with an improvement in the sensitivity. Now, in order to evaluate the prediction performance of the developed approach, the signals of unstable subjects are not entirely considered. Instead, a "lead time" is defined as the time between the last values used in the algorithm and the time K at which the ARDS is detected. Let K' be the difference between K and the considered lead time.

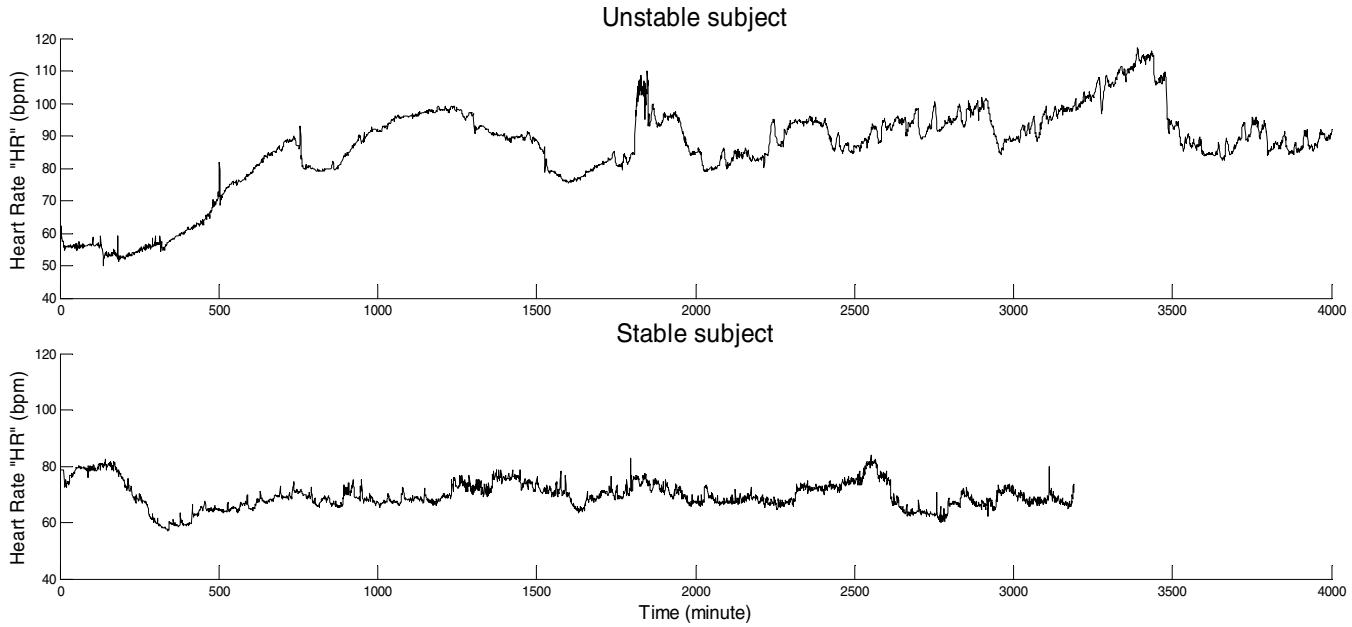

Figure 2: An example of the heart rate signal for an unstable subject (top plot) and a stable subject (bottom plot)

Hence, $\{X_i(k)\}_{k=T+1\ldots K}$, is considered in the hypothesis testing to check whether a prediction could be made. The lead times would be taken equal to 360 minutes (6 hours), 720 minutes (12 hours) and 1440 minutes (24 hours) in the algorithm.

III. Results

A. Patients selection

2,493 out of 32,267 ICU patients in the clinical database were mechanically ventilated for more than 48 hours during their hospital stay. Acute Respiratory Distress Syndrome was not present for 1023 subjects on the first day of ventilation; from these, 502 patients have worsening P/F values, and have developed ARDS. Then, we manually matched the subjects to the numerical records, to obtain 56 subjects out of 5,266 numeric records in the waveform database. Then, we extracted heart rate (HR) and breathing rate (BR) signals for these subjects. Figure 2 shows the HR signals for an unstable subject and a stable subject in the top plot and the bottom plot respectively. A pre-processing step was required to remove artifacts and fill the gaps in the signals. Data points placed away from the artifact rejection limits were removed [17], and missing data were reconstructed by average substitution with timing [18]. Additionally, we eliminated the onset of ventilating effect by removing the first six hours of the signals.

B. Statistical Analysis

As mentioned before, we are studying the state evolution of each patient separately using the signals HR and BR. The null hypothesis was defined for segments with 12 hours duration. As said before, the performance of the developed hypothesis was evaluated using sensitivity (Se) and specificity (Sp). The results are presented in Table 1, using HR and BR with zero lead time. As shown, we have reached a sensitivity of 69%, and a specificity of 55% for heart rate, and 64% and 44% respectively for breathing rate. The results show that the heart rate signals are more informative than the breathing rate signals, however the number of subjects remains too low to set a final conclusion.

To evaluate the predictive capability of our algorithm, we considered lead times equal to 6 hours, 12 hours and 24 hours and observed the changes in sensitivity. In this part, we only included the unstable subjects having signals duration higher than 72 hours (to cover the 24 hours lead time), leaving us with 14 unstable subjects, instead of 38. By working on these 14 subjects signals, the sensitivities with no lead time increased to 85.71% and 78.57% for HR and BR respectively. All sensitivity results are shown in Table 2. Specificities are unaffected since the stable subjects are not modified. The increase of the lead time for BR does not affect the ARDS prediction power of the algorithm ($Se = 78.57\%$), while the sensitivity of HR decreased with the increasing of lead time. This is mainly due to the occurrence of abnormality in BR early in the signals, which is not the case in HR signals. The prediction before 24 hours remains possible with either of the signals.

Afterwards, we applied the conjunctive and the disjunctive rules to test the prediction ability of combined signals. In conjunctive rule, a subject is considered unstable only if both signals show instability. Whereas, in disjunctive rule only one

Table 1: Preliminary results (subjects: 38 unstable/18 stable)

Signal \ Performance	Sensitivity (%)	Specificity (%)
HR	69	55
BR	64	44

Table 2: Changes in sensitivities with increased lead time (subjects: 14 unstable/18 stable)

Lead time	Heart rate	Breathing rate
0h	85.71%	78.57%
6h	78.57%	78.57%
12h	71.42%	78.57%
24h	64.28%	78.57%

signal needs to show instability for the subject to be considered unstable.

The results are presented in Table 3. As expected, the conjunctive rule raised the specificity of the test, while it decreased its sensitivity. This is not the case of the disjunctive rule that improved the sensitivity, but it reduced the specificity. To recall, we are comparing our study to that developed by Ennet et al. [9], they applied combined rule sets on data extracted from the MIMIC-II clinical database, they got a sensitivity of 60% or 50% and a specificity of 80% or 90% for single rules or refined rules respectively. Our approach outperformed their results in terms of sensitivity, especially with increasing lead time. One could then choose between both rules applied in this work depending on its interest in increasing the sensitivity or the specificity of the prediction process.

Table 3: Sensitivities and specificities of fusion rules (subjects: 14 Unstable/18 stable)

Lead time	Conjunctive Rule		Disjunctive Rule	
	Se (%)	Sp (%)	Se (%)	Sp (%)
0h	71.43	61.11	92.8	38.89
6h	64.29	61.11	92.8	38.89
12h	57.14	61.11	92.8	38.89
24h	57.14	61.11	85.71	38.89

IV. Conclusion

This paper proposed a decision rule for predicting the Acute Respiratory Distress Syndrome (ARDS). Applying a statistical analysis, the method was tested on subjects from the MIMIC II waveform database. The competency of this approach is the prediction of ARDS using physiological signals such as heart rate and breathing rate. Our results have shown an improvement in sensitivity compared to the previous study. The information of both signals was then combined using disjunctive and conjunctive rules to improve respectively the sensitivity to 93% or the specificity to 61%. Such results encourage us to expand our approach to improve the performance of the prediction of the respiratory distress using more physiological signals. Our future work will include efforts on exploring other signals and developing weighted multivariate signal analysis in order to enhance both sensitivity and specificity.

ACKNOWLEDGMENT

Authors thank the Lebanese University, Lebanon and the Université de Technologie de Troyes, France for their funding of this research.

REFERENCES

[1] L. B. Ware and M. A. Matthay, "The Acute Respiratory Distress Syndrome," *N. Engl. J. Med.*, vol. 342, no. 18, pp. 1334–1349, May 2000.

[2] K. P. Steinberg, L. D. Hudson, R. B. Goodman, C. L. Hough, P. N. Lanken, R. Hyzy, B. T. Thompson, M. Ancukiewicz, and and B. I. A. R. D. S. (ARDS) C. T. N. National Heart, Lung, "Efficacy and Safety of Corticosteroids for Persistent Acute Respiratory Distress Syndrome," *N. Engl. J. Med.*, vol. 354, no. 16, pp. 1671–1684, Apr. 2006.

[3] G. R. Bernard, A. Artigas, K. L. Brigham, J. Carlet, K. Falke, L. Hudson, M. Lamy, J. R. Legall, A. Morris, and R. Spragg, "The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination.," *Am. J. Respir. Crit. Care Med.*, vol. 149, no. 3 Pt 1, pp. 818–24, Mar. 1994.

[4] The Acute Respiratory Distress Syndrome Network, "Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome.," *N. Engl. J. Med.*, vol. 342, no. 18, pp. 1301–8, May 2000.

[5] M. B. Amato, C. S. Barbas, D. M. Medeiros, R. B. Magaldi, G. P. Schettino, G. Lorenzi-Filho, R. A. Kairalla, D. Deheinzelin, C. Munoz, R. Oliveira, T. Y. Takagaki, and C. R. Carvalho, "Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome.," *N. Engl. J. Med.*, vol. 338, no. 6, pp. 347–54, Feb. 1998.

[6] A. Esteban, A. Anzueto, F. Frutos, I. Alía, L. Brochard, T. E. Stewart, S. Benito, S. K. Epstein, C. Apezteguía, P. Nightingale, A. C. Arroliga, M. J. Tobin, and Mechanical Ventilation International Study Group, "Characteristics and outcomes in adult patients receiving mechanical ventilation: a 28-day international study.," *JAMA*, vol. 287, no. 3, pp. 345–55, Jan. 2002.

[7] X. Jia, A. Malhotra, M. Saeed, R. G. Mark, and D. Talmor, "Risk factors for ARDS in patients receiving mechanical ventilation for > 48 h," *Chest*, vol. 133, no. 4, pp. 853–861, 2008.

[8] G. Boverman and S. Genc, "Prediction of Mortality from Respiratory Distress Among Long-Term Mechanically Ventilated Patients," pp. 3464–3467, 2014.

[9] C. M. Ennett, K. P. Lee, L. J. Eshelman, B. Gross, L. Nielsen, J. J. Frassica, and M. Saeed, "Predicting respiratory instability in the ICU," in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, 2008, pp. 2848–51.

[10] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark, "Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database.," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–60, May 2011.

[11] H. Ravishankar, A. Saha, G. Swamy, and S. Genc, "An early respiratory distress detection method with Markov models," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, 2014, pp. 3438–3441.

[12] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.

[13] F. J. Massey, “The Kolmogorov-Smirnov Test for Goodness of Fit,” *J. Am. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, Mar. 1951.

[14] S. S. Chen and R. A. Gopinath, “Gaussianization,” in *Nips*, 2001, p. 7.

[15] M. G. Kendall, “Elementary Statistics,” *Nature*, vol. 155, pp. 531–531, 1945.

[16] T. L. Lai, Lai, and T. L., “Sequential Analysis,” in *Encyclopedia of Biostatistics*, Chichester, UK: John Wiley & Sons, Ltd, 2005.

[17] A. Hann, “Multi-parameter monitoring for early warning of patient deterioration.” University of Oxford, 2008.

[18] P. Langley, S. King, K. Wang, D. Zheng, R. Giovannini, M. Bojarnejad, and A. Murray, “Estimating of missing data in multi-channel physiological time-series by average substitution with timing from a reference channel,” in *2010 Computing in Cardiology.*, 2010, pp. 309–312.