Optimization techniques to detect early ventilation extubation in Intensive Care Units

Pedro Oliveira[1] Filipe Portela[1,2], Manuel F. Santos[1], José Machado[1], António Abelha[1], Álvaro Silva[3], Fernando Rua[3]

[1]Algoritmi Research Centre, University of Minho, Portugal
[2]ESEIG, Porto Polytechnic, Porto, Portugal
[3] Intensive Care Unit, Centro Hospitalar do Porto, Porto, Portugal
a61565@alunos.uminho.pt, {cfp, mfs}@dsi.uminho.pt, {jmac, abelha}@di.uminho.pt
moreirasilva@me.com; fernandorua.sci@chporto.min-saude.pt

Abstract. The decision support models in intensive care units are developed to support medical staff in their decision making process. However, the optimization of these models is particularly difficult to apply due to dynamic, complex and multidisciplinary nature. Thus, there is a constant research and development of new algorithms capable of extracting knowledge from large volumes of data, in order to obtain better predictive results than the current algorithms. To test the optimization techniques a case study with real data provided by INTCare project was explored. This data is concerning to extubation cases. In this dataset, several models like Evolutionary Fuzzy Rule Learning, Lazy Learning, Decision Trees and many others were analysed in order to detect early extubation. The hydrids Decision Trees Genetic Algorithm, Supervised Classifier System and KNNAdaptive obtained the most accurate rate 93.2%, 93.1%, 92.97% respectively, thus showing their feasibility to work in a real environment.

Keywords. Optimization techniques; Decision Support Systems; Machine Learning; Heuristics; Intensive Care Units Extubation.

1 Introduction

This work is part of an evaluative study integrated in INTCare project. INTCare is an Intelligent Decision Support Systems (IDSS). It is implemented in the Intensive Care Unit (ICU) of the Hospital Centre of Porto (CHP). This IDSS is based on intelligent agents and data mining models [1] that seek to automate tasks and to predict clinical events in order to help the diagnosis and treatment of patients by providing new knowledge able to help them to recover to his previous health state. This type of patients is usually admitted to the ICU where it is possible to monitoring and maintaining their physiological functions through life support devices. These units provide a continuous patient monitoring of vital signs as well as a constant analysis of each organ system: neurological, respiratory, hepatic, haematological, cardiovascular and kidney [2]. At the same time, many of those patients are mechanically ventilated. Around 75% of the

admitted patients to ICU require assisted ventilation to support breath function, with an automatic mechanical control that can improve the quality of patient care. A well-done extubation process can reduce the risk in provoking lung injuries and reduce the morbidity and mortality rates associated with provision of inappropriate cases of ventilation. To enhance this process urges the use of optimization techniques to improve the models already developed by Oliveira et al. [3]. An exhaustive literature review for this work was made and several optimization techniques applied to local and global solutions were found. These techniques are used to interpret the collective behaviours and evolution of species or adapting both making them hybridizing to overcome some or other weakness from one to another.

The results take in consideration the performance of the method present in the system, statistical classification tests, the predictive acuity and representation derivation of the solution more user friendly. To assess the efficiency of the optimization techniques in this case study, 63 techniques were explored to insight the viability of improving the INTCare system. Decision Trees Genetic Algorithm (DT_GA), sUpervised Classifier System (UCS) and K-Nearest Neighbours Adaptive algorithm (KNNAdaptive), with results 93.2%, 93.1% and 92.97% respectively, achieved the best accuracy rates. These models are important to determine the best time to extubation, through the characteristics of these patients.

The present paper is divided in five sections, Introduction of the theme, explanation of the Background scenario, Materials and Methods taken in consideration; CRISP-DM sub phases to guide the work, and Conclusion and future work.

2 Background

2.1 Intensive Care Units

People who have serious illnesses are usually admitted to the intensive care units (ICU) so that it is possible to keep their vital physiological functions through various media devices, such as medical devices (ventilators, vital signs machines, etc.) until they have their organs functioning independently again [4].

The bulk heterogeneous information processing becomes critical, which makes the extremely complex environments in intensive care units taking into account the amount of data that it is necessary consider by medical teams [5]. This reality exposes the fragility of the teams and a few existing decision support models in situations where the decision time on the analysis of multiple variables is critical. The optimization techniques have interdisciplinary characteristics able to optimize the solutions and decisions. For example, these techniques are capable of anticipating the decision making process on a patient while maintaining a monitoring of its vital functions, ensuring their safety and helping to realize the best treatment.

2.2 Mechanical ventilation weaning and extubation

A patient intubation case happens when the respiratory system fails in oxygenation, carbon dioxide elimination or both [6]. The goal of artificial ventilation intubation is to reduce lung injury due to over distention. In current days, mechanical ventilation is very important to treat many different illnesses, but is relatively costly [7].

In ICU, the weaning process is gradual and it is considered successful when a patient can breathe by himself for a period upper than one hour. Actually there is a set of IDSS to ventilators, an expert advisory system or an automatic control of ventilation [3], however, most of them are rule-based, turning them into static models and not adaptive.

The extubation process is based on clinical knowledge and medical assumption [8] creating a tentative-error procedure. This happens because the ventilators are only used to consult the patient values; the unused data is not stored in any database or file, resulting in a waste of data that could be transformed in information and up ways into knowledge to help the decision-making process.

2.3 Optimization

In medicine, optimization has become ubiquitous [9]. The use of computing power for applications in the medical field has opened up many questions and challenging problems inherent in these communities. The mathematical techniques (continuous and discrete) play an increasingly important role in understanding of several critical problems in medicine. Of course, optimization is one fundamental tool due to the limitation of the resources involved and the need for better decision-making in the shortest time possible [9]. An optimization algorithm is an iterative process where after a certain number of iterations may converge to a solution that ideally will be the optimal solution to the problem. During the interactions that occur in the process, the solutions that emerge are state of evolution that are drawn according to mathematical equations or set of rules for convergent solutions of a self-organizing system. As a result, their ability to represent a self-organizing system shows us some emerging features and its ability to evolve.

Briefly, course, discontinuous, Single-solution based, Population-based, Guided search, Unguided search and hybrid methods are described as some of the most important features for this type of optimization methods [10]. If it is possible to provide a balance between diversification and intensification, the metaheuristics techniques will be successful in a given optimization problem. The increase is needed in the search for parts in space with high quality solutions. It is important finding some promising areas on the accumulated research experience. The main differences between the existing meta-heuristics relate to their particular way of achieving this balance [11]. The classification criteria can be used for the meta-heuristics, in terms of the features that follow in the search, memory feature, kind of neighbour holding used or the number of current solutions created through iterations.

2.4 Population based methods

Dealing with a set (population) solutions instead of an initial solution. Most studies based on these methods are related to Evolutionary Computation (EC) inspired by Darwin's theory, where the population of individuals is modified through recombination, genetic operators and swarm intelligence (SI). Here the idea is to create computational intelligence to explore simple analogies of social interactions rather than purely individual cognitive abilities [12] in how the simulation of evolution structure their subjects, objectives, through processes of selection, recombination and mutation breeding in order to develop better solutions [13]. In general, the following algorithm serves as an example for the methods addressed in this study.

Algorithm 1 – Evolutionary Computation

```
Algorithm - EC
Initialize the population with random individuals
Evaluate each individual
Repeat
    Select country
    Recombine pairs of parents
    Mutate descendants
    Evaluate new individuals
    Select individuals for the next generation
Until the stop criterion is satisfied
```

3 Materials and Methods

A literature review about optimization techniques to support decision models already elaborated was taken in consideration [14]. To explore optimization KEEL (Knowledge Extraction based on Evolutionary Learning) [15] algorithms were used. This tool allow evaluating evolutionary algorithms in Data mining problems, but also in Machine Learning questions. In this work, methods like Evolutionary Fuzzy Rule Learning, Lazy Learning, Evolutionary Crisp Rule Learning, Prototype Generation, Fuzzy Instance Based Learning, Decision Trees, Crisp Rule Learning, Neural Networks and Evolutionary Prototype Selection were analysed and explored.

Even without any explicit Data Mining technique been used, the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology was chosen to guide this study. CRISP-DM is divided in six phases, Business Understanding, Data Understanding, Data preparation, Modelling, Evaluation and Deploy.

4 Knowledge Discover Process

The following sections were divided according to CRISP-DM phases.

4.1 Business Understanding

This work was proposed to find optimization techniques in order to improve the INTCare system on predicting the best time to a successful extubation. It is also expected creating knowledge to help clinical teams to make better decisions in ICU scenario. The data used for these experiments were from other work already carried out, in the Centro Hospitalar do Porto (CHP) [3]. The study aimed to identify a set of features / variables associated to a successful extubation. Most of the data provided by INTCare system was already characterized and transformed. The dataset created resulted from the application of DM techniques, more specifically the creation of clusters, using the k-means algorithm and k-medoids based the partition principle and sensitivity difference extreme values [3].

4.2 Data Understanding

The dataset used correspond to 50 patients in a total of 24295 records with fifteen fields from the ventilators in the ICU of CHP, collected between 2014-09-19 and 2015-02-03. Each record is composed by the following variables:

- Pressure support setting in cmH2O;
- End inspiratory pressure in cmH2O;
- Dynamic compliance (CDYN) in mL/cmH2O;
- Static compliance (CSTAT) from inspiratory pause manoeuvre in mL/cmH2O;
- Mean airway pressure in cmH2O;
- Maximum circuit pressure in cmH2O;
- Static resistance (RSTAT) from inspiratory pause manoeuvre in cmH2O/L/s;
- O2% setting;
- Peak flow setting in litters per minute;
- Tidal volume setting in litters;
- Respiratory rate setting in breaths per minute;
- Exhaled tidal volume in litters;
- PEEP or PEEP Low setting in cmH2O.

4.3 Data preparation

To prepare the data, after a brief check, a feature selection algorithm present in KEEL was used. This algorithm allowed determining the most preponderant variables in the dataset. Reaching the results, a new dataset with 14139 records to use on the techniques as inputs was created. The following

Table 1 describes what each input variable means their average, standard deviation and variance.

Table 1. Description and distribution of the inputs

Description	$\overline{x}$	σ	s^2
Peak flow setting in litters per minute (F12)	24.00	23.55	554.86
O2% setting (F13)	49.75	7.65	58.56
PEEP or PEEP Low setting in cmH2O (F15)	5.09	1.05	1.11
Pressure support setting in cmH2O (F26)	14.11	3.60	12.98
Exhaled tidal volume in litters (F35)	0.52	0.16	0.02
Maximum circuit pressure in cmH2O (F38)	20.53	3.86	14.95
Mean airway pressure in cmH2O (F39)	10.50	1.84	3.40
End inspiratory pressure in cmH2O (F40)	19.98	3.85	14.86
Dynamic compliance in mL/cmH2O (F65)	45.75	40.18	1615.23

As output, the Target variable translates three events. When the value is 0, means that it was not possible to extubate the patient, when the value is 1 the patient was extubated but after some time, he needed to be ventilated again and when the value is 2, the patient was extubated with success. As can be observed Fig. 1 shows a limitation in successes cases.

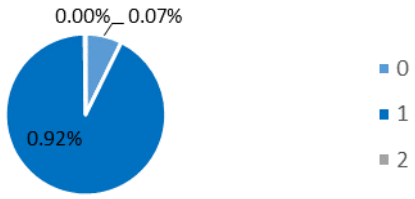

Fig. 1. Output representation from the received data

After that, the dataset was partitioned using cross validation presented by Dietterich in 1998 [16].

4.4 Modelling

In this work only the algorithms proposed after the year 2000 were considered. After elaborate the experiments some algorithms were excluded due to their huge process time consuming and unsuccessful results. Several experiments were compromised, making them inconclusive, despite being tested. Altogether 63 algorithms tests were conducted, which included 10 models Evolutionary Fuzzy Rule Learning, 16 Evolutionary Crisp Rule Learning, 4 Lazy Learning, 7 Prototype Generation 5 Fuzzy Instance based Learning, 7 Decision Trees, 5 Crisp Rule Learning, 4 Artificial Neural Networks and 5 Evolutionary Prototype Selection with selection criteria as the global average acuity (the set of test data). To the case study, the configurations presented in Table 2 were used.

Table 2. Parameter settings of the algorithms

Algorithm	Parameters
DT_GA	Confidence: 0.25, instances per leaf: 2, genetic algorithm approach: GA-LARGE-SN, threshold S to consider a small disjunt:10, number of total generations for the GA: 50, number of chromosomes in the population: 200, crossover probability: 0.8, mutation probability: 0.01
UCS	Number of explores: 100.000, population size: 6400, delta: 0.1, nu: 10, Acc: 0.99, pX: 0.8, pM: 0.04, theta_ga: 50, theta_del: 50, theta_sub: 500, doGASubsumption: true, type of selection: RWS, tournament size: 0.4, type of mutation: free, type of crossover: 2PT, r: 0.6, m: 0.1,
KNNAdaptive	K value: 1, distance function: euclidean

4.5 Evaluation

After an exhaustive test made in KEEL the top 3 techniques / results was selected. The hybrid algorithm Decision Trees Genetic Algorithm (DT_GA) [17], the fuzzy rules, learning classifier systems and data mining supervised Classifier System (UCS) [18] and K-Nearest Neighbours Adaptive algorithm (KNNAdaptive) [19]. These algorithms are the best optimization techniques to improve the decision making process to determine the conditions for successful extubation in intensive care units.

To DT_GA, the parameters defined in KEEL obtained a 93.2% accuracy rate, to UCS, the set parameters achieved one accuracy rate of 93.1% and finally to the KNNAdaptive algorithm yielded a 92.97% accuracy rate.

Table 3 show the average results in the Friedman test [20], Multiple test [20] and Contrast Estimation [21]. Table 4 presents confidence matrix for three trust levels: 99%, 95% and 90%.

Table 3. Classification test of the best techniques

Median error in classification			
	Friedman test	Multiple test	Contrast Estimation
DT_GA	0.06	0.06	0.06
UCS	0.07	0.07	0.07
KNNAdaptive	0.07	0.07	0.07

Table 4. Confidence matrix of the best techniques

p-value confidence matrix			
	DT_GA	UCS	KNNAdaptive
DT_GA	0.00	0.09	0.25
UCS	0.00	0.00	0.96
KNNAdaptive	0.00	0.00	0.00

DT_GA technique have high interpretability of results by the user. Figure 2 show an example of the sheets (rules) for the case 2 (successful extubation) and table 5 the execution time need to achieve the global average by algorithm.

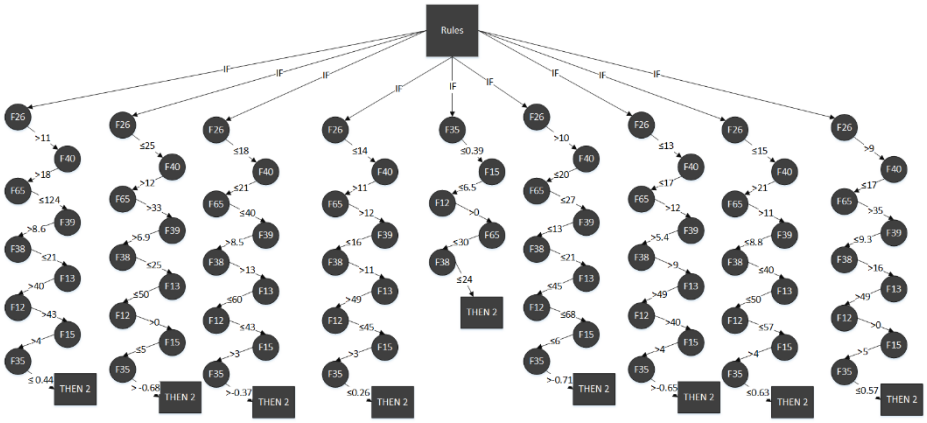

Fig. 2. Decision trees rules created by DT_GA algorithm

Table 5. Global average acuity and execution time of the best techniques

Algorithm	Global average	Execution time
DT_GA	93.2%	4 minutes
UCS	93.1%	4 minutes
KNNAdaptive	92.97%	2 minutes

4.6 Deploy

The best optimization technique to apply in the ICU for this case study was the DT_GA, not only because the predictive accuracy, but also for the interpretability of the results. It shows the value range of the variables to consider in detecting an early extubation and medical teams can take into consideration with the representation of these variables value. These optimization models will be added to INTCare system in order to improve the Data Mining models.

5 Conclusion and future work

Continuing the studies already made in INTCare [3, 22, 23, 24, 25, 26, 27, and 28] in the field of respiratory system and data mining a new approach was explored.

Throughout this work, the study has been extended by a quantitative research based on a benchmarking process, where several existing optimization techniques able to be included in these models were explored. After an extensive and appropriate literature review [14] the most adequate techniques were explored in KEEL. A total of 63 experiments were made.

Evolutionary models: Crisp Rule Learning, Lazy Learning and Decision Trees were selected. The best results were achieved by DT_GA techniques, UCS and KNNAdaptive with the accuracy rate of 93.2%, 93.1% and 92.97% respectively.

In the future, these models will be incorporated in INTCare system in order to improve the results and optimize the support system already implemented to predict

ventilator extubation. The technical aspects of introducing new procedures should be described in detail in order to provide necessary instructions for medical assistance which have to able to interpret the collected data. Moreover, there is a need to analyze the implementation effectiveness of optimization techniques presented in the article and further research on feedback given by medical professionals.

Acknowledgments

This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013. The authors would like to thank FCT for the financial support through the contract PTDC/EEI-SII/1302/2012 (INTCare II).

References

1. Santos, M.F., Portela, F., Vilas-Boas, M.. INTCARE -Multi-agent Approach for Real-time Intelligent Decision Support in Intensive Medicine. ICAART 2011 - International Conference on Agents and Artificial Intelligence. pp 364-369. ISBN: 978-989-8425-40-9. Rome, Italy. SciTePress. (2011).
2. Filipe Portela, Filipe Pinto, Manuel Filipe Santos. Data Mining Predictive Models For Pervasive Intelligent Decision Support In Intensive Care Medicine. KMIS 2012 - Knowledge Management and Information Sharing. pp 81-88. ISBN: 978-989-8565-31-0. Barcelona, Spain. SciTePress (2012).
3. Oliveira, S., Portela, F., Santos, M. F., Neves, J., Silva, Á. Rua, F.: Feature selection for detecting patients with weaning failures in Intensive Medicine, Mathematics and Computers in Sciences and Industry. Volume 50, pp 195-200. ISBN: 978-1-61804-327-6. CPS. (2015)
4. Ramon, J., Fierens, D., Güiza, F., Meyfroidt, G., Blockeel, H., Bruynooghe, M., & Van Den Berghe, G.: Mining data from intensive care patients. Advanced Engineering Informatics, 21(3), 243–256. doi:10.1016/j.aei.2006.12.002. (2007)
5. De Turck, F., Decruyenaere, J., Thysebaert, P., Van Hoecke, S., Volckaert, B., Danneels, C., De Moor, G.: Design of a flexible platform for execution of medical decision support agents in the intensive care unit. Computers in Biology and Medicine, 37, 97–112. doi:10.1016/j.compbiomed.2005.10.004. (2007)
6. Kaynar, A. and Sharma, S.: Respiratory Failure. 39. Available: http://emedicine.medscape.com/article/167981-print. (2010)
7. Tehrani, F. T.: Automatic control of mechanical ventilation. Part 2: the existing techniques and future trends, Journal of clinical monitoring and computing, vol. 22, pp. 417-424. (2008)
8. Stawicki, S. P.: Mechanical ventilation: weaning and extubation (2007)
9. Alves, C. J. S., Pardalos, M. P., Vicente, L. N.: Optimization in Medicine, Springer Optimization and its Applications Series, Vol. 12. Springer (2008)
10. Gilli, M., & Winker, P.: A review of heuristic optimization methods in econometrics. Heuristic Optimization Methods in Econometrics (2008).
11. Birattari, M., Paquete, L., Stützle, T., Varrentrapp, K.: Classification of Metaheuristics and Design of Experiments for the Analysis of Components, Technical Report AIDA-01-05, FG Intellektik, FB Informatik, Technische Universität Darmstadt, Darmstadt, Germany. (2001)
12. Boussaïd, I., Lepagnot, J., & Siarry, P.: A survey on optimization metaheuristics. Information Sciences, 237, 82–117. doi:10.1016/j.ins.2013.02.041. (2013)

13. Bäck, T.: Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms, Oxford University Press, Oxford, UK. (1996)
14. Oliveira, P., Portela, C.F., Santos, M.F., Silva, Á., Machado, J., Abelha, A.: Machine Learning: an overview of optimization techniques. Recente Advances in Computer Science, INASE (2015)
15. Alcalá-Fdez, J., Sánchez, L., Garcia, S., del Jesus, M.J., Ventura, S. Garrel, J.M., Otero, J., Romero, C., Bacardit, J., Rivas, V.M., Fernández, J.C., Herrera, F.: KEEL: A Software Tool to Assess Evolutionary Algorithms to Data Mining Problems. Soft Computing 13:3 (2009)
16. Dietterich, T.G.: Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Computation 10:7 (1998)
17. Carvalho, D.R., Freitas, A., A.: A hybrid decision tree/genetic algorithm method for data mining. Information Sciences 163:1, 13-35. (2004)
18. Bernadó-Mansilla, E., Garrel, J., M.: Accuracy-Based Learning Classifier Systems: Models, Analysis and Applications to Classification Tasks. Evolutionary Computation 11:3, 209-238. (2003)
19. Wang, J., Neskovic, P., Cooper, L., N.: Improving nearest neighbor rule with a simple adaptative distance measure. Pattern Recognition Letters 28, 207-213. (2007)
20. Sheskin, D.: Handbook of parametric and nonparametric statistical procedures. Chapman and Hall/CRC (2003)
21. Doksum, K.: Robust procedures for some linear models with one observation per cell. Annals of Mathematical Statistics 38, 878-883 (1967)
22. Filipe Portela, Manuel Filipe Santos, José Machado, Álvaro Silva and António Abelha. Pervasive and Intelligent Decision Support in Critical Health Care using Ensemble. Lecture Notes in Computer Science (LNCS) - Information Technology in Bio- and Medical Informatics. Volume 8060, 2013, pp 1-16. ISBN: 978-3-642-40093-3. Springer. (2013).
23. Filipe Portela, Manuel Filipe Santos, Álvaro Silva, António Abelha, José Machado. Pervasive Ensemble Data Mining Models to Predict Organ Failure and Patient Outcome in Intensive Medicine. Communications in Computer and Information Science. Volume 415, 2013, pp 410-425. ISBN: 978-3-642-54104-9. Springer. (2013).
24. Portela, F., Santos, M., Vilas-Boas, M., Rua, F., Silva, Á., Neves, J.: Real-time Intelligent decision support in intensive medicine. In: KMIS 2010-International Conference on Knowledge Management and Information Sharing, Valência, Espanha, p. 7 (2010)
25. Sérgio Oliveira, Filipe Portela, Manuel Filipe Santos, José Machado, António Abelha, Álvaro Silva and Fernando Rua. Predicting Plateau Pressure in Intensive Medicine for Ventilated patients. Advances in Intelligent Systems and Computing (WorldCist 2015 - Healthcare Information Systems: Interoperability, Security and Efficiency Workshop). Volume 354, 2015, pp 179-188. ISBN: 978-3-319-16527-1. Springer. (2015).
26. Portela, C.F., Santos, M.F., Silva, Á., Machado, J., Abelha, A.: Enabling a Pervasive Approach for Intelligent Decision Support in Critical Health Care. In: Cruz-Cunha, M.M., Varajão, J., Powell, P., Martinho, R. (eds.) CENTERIS 2011, Part III. CCIS, vol. 221, pp. 233–243. Springer, Heidelberg (2011)
27. Portela, F., Santos, M.F., Vilas-Boas, M.: A Pervasive Approach to a Real-Time Intelligent Decision Support System in Intensive Medicine. In: Fred, A., Dietz, J.L.G., Liu, K., Filipe, J. (eds.) IC3K 2010. CCIS, vol. 272, pp. 368–381. Springer, Heidelberg (2013)
28. Portela, F., Santos, M. F., Machado, J., Abelha, A., Silva, Á., & Rua, F.: Pervasive and intelligent decision support in intensive medicine--the complete picture. Information Technology in Bio-and Medical Informatics, 87–102. http://doi.org/10.1007/978-3-319-10265-8_9 (2014)