

Article

# Patients' Admissions in Intensive Care Units: A Clustering Overview

Ana Ribeiro <sup>1</sup>, Filipe Portela <sup>1,\*</sup>, Manuel Santos <sup>1</sup>, António Abelha <sup>1</sup>, José Machado <sup>1</sup> and Fernando Rua <sup>2</sup>

<sup>1</sup> Centro ALGORITMI, University of Minho, Campus Azurém, 4800-058 Guimarães, Portugal; a64886@alumni.uminho.pt (A.R.); mfs@dsi.uminho.pt (M.S.); abelha@di.uminho.pt (A.A.); jmac@di.uminho.pt (J.M.)

<sup>2</sup> Intensive Care Unit, Centro Hospitalar do Porto, Largo do Prof. Abel Salazar, 4099-001 Porto, Portugal; fernandorua.sci@chporto.min-saude.pt

\* Correspondence: cfp@dsi.uminho.pt; Tel.: +351-253-510-180

Academic Editor: Willy Susilo

Received: 20 November 2016; Accepted: 14 February 2017; Published: 17 February 2017

**Abstract:** Intensive care is a critical area of medicine having a multidisciplinary nature requiring all types of healthcare professionals. Given the critical environment of intensive care units (ICUs), the need to use information technologies, like decision support systems, to improve healthcare services and ICU management is evident. It is proven that unplanned and prolonged admission to the ICU is not only prejudicial to a patient's health, but also such a situation implies a readjustment of ICU resources, including beds, doctors, nurses, financial resources, among others. By discovering the common characteristics of the admitted patients, it is possible to improve these outcomes. In this study clustering techniques were applied to data collected from admitted patients in an intensive care unit. The best results presented a silhouette of 1, with a distance to centroids of  $6.2 \times 10^{-17}$  and a Davies–Bouldin index of  $-0.652$ .

**Keywords:** data mining; decision support systems; clustering; intensive care; admissions; INTCare system

---

## 1. Introduction

Health organizations generate and store large volumes of data every day. Over the years we have observed that the evolution of technology has had an impact on how these data volumes are treated. In recent years, several methods to automate the entire data management and knowledge discovery process has appeared [1]. The emergence of techniques such as data mining (DM) and its application in the healthcare industry has enabled the improvement of services provided to patients. Saving lives using these methods can be possible now.

One of the main problems found in intensive care units (ICUs) are related to unplanned admissions. There are some solutions to predict readmissions and other critical events, but there is a lack of solutions directed to ICU admissions. The poor selection of patients to be admitted to the ICU is a major cause of hospital resource occupation (beds, doctors, nurses, financial resources, etc.) which could be used with other patients in need. Additionally, ICU extended stays expose patients to infections and inflammations that can aggravate their conditions [2,3].

Patients can be admitted urgently if they need special care for maintaining the function of their vital organs or because they only need to be monitored continuously for a given period of time and being treated according to their clinical status [3].

Through the predictive feature of data mining techniques, there are projects that allow the anticipation of critical events [4], such as a patient's readmission [5,6], among others. Another possibility

is the use of clustering techniques to find patterns in data by creating natural groups with similar characteristics. This is the main goal of the paper.

Clustering techniques were applied to data extracted from a Clinical Decision Support System called INTCare being used in the ICU of Centro Hospitalar do Porto, in Portugal. These data were structured in different scenarios, using two different data mining tools (Orange, version 2.7, Bioinformatics Lab at University of Ljubljana, Slovenia; RapidMiner, version 7.4, RapidMiner, Inc., Boston, MA, USA) and different evaluation methods: silhouette, inter-cluster distance, and distance to centroids, *K*-means, and Davies–Bouldin index. Clustering techniques were used in this work, which originated from a prior application of classification techniques. In addition, it allowed identifying some useful variables for the application of classification techniques [7]. The data used are the same before admission. It was intended to create patterns that allow understanding of the patient condition at admission (using data collected during the patient stay in hospital). This situation allows the clinician to easily understand if a patient has, or does not have, similar values that were presented by the clusters created. These data are stored in the electronic health record during patient admission in other services. The values are combined and humans cannot make all of the combinations without the use of a machine. Another goal of this work is alerting the clinicians to the patient's condition. For example, the patient can have five days before ICU admission due to transplant or surgery. These data are stored in the database and, when the patient is admitted to the ICU, they are considered. When a patient is outside the ICU and has some of these values, the clinician will be alerted when there is a match of some patient conditions (outside the ICU) with another patient admitted earlier to the ICU.

All of this work is framed in INTCare research work. Several studies were performed using other tools and algorithms. This paper presents a particular part of that research work.

The best scenario used only three attributes, since the others had a negative impact on the results. The best results had a silhouette of 1, inter-cluster distance of 1.5, and distance to centroids of  $6.2 \times 10^{-17}$ .

This document is divided into five sections. The first is the introduction of the problem and the topics that will be discussed during the paper are described. The second section provides a background, presenting the basic concepts involved in this work. The third section is the Study Description where the tools and techniques used in this work are identified, and the business understanding, data understanding, and preparation, modeling, and evaluation are presented. The fourth section is a discussion of the findings. This section presents some interesting analytical points about the results achieved within this work. The fifth section provides the achieved conclusions, with the results obtained, and where further work is introduced.

## 2. Background

### 2.1. Intensive Medicine

Intensive medicine arose during the 20th century [8] and it is composed by several other fields of healthcare that meet with the objective of diagnose and treating patients with critically ill conditions, potentially reversible. In these patients, there is a possibility of being verified multiple organ failures. By this reason they need to be continuously monitored [4]. This type of patients is admitted to a specialized place called Intensive Care Unit (ICU) characterized by its critical environment [9]. Bersten and Soni [10] consider two types of patients that can be admitted to ICU: Patients who require continuous monitoring and treatment and patients with organ failure, with recovering hypotheses. These units are responsible for the admitted patients, with continuous monitoring (24-hour) of the patient's condition by medical equipment and physicians of different specialties prepared to act in any critical situation [11].

### 2.2. Admissions

In order to provide better healthcare to patients it is necessary to have an efficient way to properly admit patients to intensive care [12].

Several authors [12,13] suggested the identification of two types of patients: high risk and low risk. The first type of patients are those who are too sick to benefit from intensive care and would have no improvement in their health. The second category are those who are not ill enough to require intensive care and that can be treated in other hospital units.

The poor selection of patients to be admitted to the ICU causes the overcrowding of beds and other resources, such as doctors and nurses, or medical equipment that could be used correctly with other patients [2]. Furthermore, a prolonged stay in the ICU exposes patients to infections and inflammations which complicate their health condition [2,3].

This work aims to show the importance of categorizing the type of patients admitted to the ICU in order to be able to optimize resources, reduce costs, and realize in advance the characteristics of a hospitalized patient.

### 2.3. Clustering

Clustering consists in apportioning data objects into a number of groups called clusters [14] where all elements are similar within the group [15]. These clusters should be internally similar but externally different [14]. This technique has a variety of applications areas, like engineering, economics, medical science, and astronomy, among others [16].

The main objective of the cluster analysis is discovering hidden structure and relations between the data [16]. There are some approaches to clustering, but the most popular are the agglomerative clustering and prototype-based clustering. In the first approach, each element of the data is initially partitioned into single clusters. The most similar clusters are united into one, and so on, until a determined number of clusters is reached. In the second approach, the number of clusters is the first thing to be defined, and then part the data into those cluster until they are similar intra-cluster and different inter-clusters [15].

In this work, two methods were used with different evaluation metrics for when it is not possible to make a comparison between the two techniques.

### 2.4. INTCare

The INTCare system is an intelligent decision support system with pervasive features developed for the area of intensive care medicine. It is currently installed in the ICU of the Centro Hospitalar do Porto (CHP), Porto, Portugal [5,17,18]. The purpose of introducing such a system was to modify the response time from reactive to proactive [5] and improve the quality of treatment of patients by suggesting treatments, appropriate therapies and procedures, and also to predict the clinical conditions of the patient in the next hours. Data is provided from five data sources: bedside monitor, electronic health record, electronic nursing record, laboratory results, and drugs system. With INTCare it is possible to improve the knowledge base on which health professionals base their decisions [19].

### 2.5. Related Work

There are some projects in the field of prediction of admissions to the ICU, but most of them do not use data mining to achieve this. Instead, it is statistical analysis with the objective of predicting events is commonly used. The difference between data mining and statistical analysis is that statistics are the central part of the data mining process. Data mining also involves data exploration and analysis, cleaning and visualization, and provides means to achieve patterns and predictions, while statistical analysis is concerned with probabilistic models and quantifying numbers.

One of the existing projects, held in New York, is aimed at the triage of high-risk hospital patients in the ICU after total arthroplasty of the hip (ATA). It is a statistical model to identify preoperative risk, which can predict whether a patient is more likely to be admitted to the ICU after operation [20].

A study conducted in France, using statistical analysis, allows stratifying patients with pneumonia in four admissions to the ICU risk categories within three days after their presentation in a hospital emergency (HE). The categories 1 and 2 are moderate risk, although it is required some monitoring

they do not need to be admitted to the ICU. In categories 3 and 4, patients should be transferred to the ICU after its entry as HE [21].

There's also a medical score named "Medical Early Warning Score" (MEWS), which helps with the identification of patients at risk of being admitted to the ICU and classifies them by illness severity [22]. However this score is not, by itself, the most efficient way to predict admissions because it only uses data from vital signs [23].

There are many others studies like the previous ones [23–26] using statistical analysis.

From the studies found there is only one that allows predictions of readmissions in the ICU using Data mining techniques. The INTCare system uses data mining techniques to forecast critical events and readmissions [27,28]. One of the studies found also uses clustering techniques to find patterns in the data and assemble them into groups with similar features of patients who were readmitted to the ICU [6].

We conducted a classification to predict patient admissions. In this work it only categorized patient admissions data was used. This data, as it was said before, will be used by clinicians as an alert point. When a patient outside of the ICU has some of these values the clinician can make a decision based in their expertise. The clusters are computed in real-time and when a patient is framed in some of the conditions the clinicians will receive an alert from the system.

Compared to the projects found, this work provides many contributions to the data mining community, particularly with respect to patient admissions to the ICU.

### 3. Study Description

#### 3.1. Methods and Tools

Cross Industrial Standard Process for Data Mining (CRISP-DM) was followed as the data mining methodology. It is divided into six phases: business understanding, business, data understanding, data preparation, modeling, and evaluation and deployment. The tools used were Oracle SQL Developer, for data exploration and preparation, and Orange and RapidMiner for building scenarios and clustering data. In Orange, the following tests were used: silhouette, inter-cluster distance and distance to centroids. RapidMiner used a k-means algorithm and Davies–Bouldin Index.

#### 3.2. Business Understanding

The main goal of this work, as already mentioned, is to characterize patients admitted to the ICU from their medical data and discovering if it has groups of characteristics that make sense and are helpful. Clustering these data is important to determine all of the hidden features and relations before further analysis.

The data mining goal is to create useful models, able to identify groups of patients in the ICU with similar characteristics through clustering techniques.

The solution should be able to support medical decision in the ICU, presenting to the intensivists information about the characteristics of admitted patients and providing new knowledge in this field.

#### 3.3. Data Understanding and Data Preparation

The data were extracted from the ICU of Centro Hospitalar do Porto, using the Agency for Integration, Diffusion and Archive of Medical Information (AIDA) platform, and comes from three tables. One is composed of 12 attributes and contains data about the hospital admission of patients between 30 June 2006 and 18 February 2016. The second table is composed of eight attributes and contains information about admissions to the ICU, for example, the identification of the bed and the patient identification number, among others. The third table is composed of 53 attributes and contains clinical data about the patients admitted to the ICU. In order to understand the data, an exploration of all attributes was made. The attributes were characterized by function, type, and value ranges.

From the table containing data from hospital admissions, 86.6% of the patients were previously admitted to the ICU.

Patients' gender analysis is represented in Table 1. By using the sex attribute, it was possible to determine that the most predominant patients were male, both at hospital admission and ICU admission.

**Table 1.** Percentage distribution by gender.

		Gender	
		Man	Woman
Local	Hospital	39%	61%
	ICU	62%	38%

Analyzing patient's birth dates, it was possible to make a statistical analysis. As can be understood through Table 2, both at hospital and ICU admission 13 was the minimum patient's age and the mode is also 75 for both locales. Analyzing the others measures, the ages do not differ much.

**Table 2.** Statistical measurements of age attribute.

		Measure				
		Max	Min	Mode	Avg	Median
Local	Hospital	96	13	75	62.19	64
	ICU	101	13	75	59.07	60

Considering the ICU admitted patients, most of them were admitted after surgery from an operating room and with urgency.

For patients who were hospitalized more than one day, the maximum number of days of hospitalization was 894, probably for some comatose patients; 1281 patients were hospitalized for one day, this being the most frequent number of days of hospitalization. The average number of days of hospitalization is about five days.

The dataset used in the modeling phase include only clinical data about patients admitted to the ICU, which is composed of several attributes that were used in the modeling phase:

- VA: GLAGOW\_HOSPITAL: classify the patients accordingly to Glasgow Coma Scale at the hospital;
- VB: GLASGOW\_SERVICE: classify the patients accordingly to Glasgow Coma Scale at the service;
- VC: DEAF: indicates if the patient is deaf;
- VD: MUTE: indicates if the patient is mute;
- VE: BLIND: indicates if the patient is blind;
- VF: Allergy: indicates if the patient has any allergy;
- VG: RESPIRATORY: indicates if the patient has some respiratory problems;
- VH; PACEMAKER: indicates if the patient has a pacemaker;
- VI: PSYCHIC DESABILITY: indicates if the patient has physical disability;
- VJ: PHYSIC DESABILITY: indicates if the patient has any physical disability;
- VK: ALCOHOLISM indicates if the patient has alcoholism problems;
- VL: DRUG ADDICT: indicates if the patient has some problem with drugs;
- VM: ATTRIBUTE H: hospital attribute (confidential);
- VN: ATTRIBUTE S: hospital attribute (confidential);
- VO: LIVER CHRONIC INSUFFICIENCY: indicates if the patient is suffering from chronic liver insufficiency;
- VP: CHRONIC RENAL FAILURE: indicates if the patient is suffering from chronic renal failure;

- VQ: CARDIAC INSUFFICIENCY: indicates if the patient is suffering from cardiac insufficiency;
- VR: CHRONIC RESPIRATORY FAILURE: indicates if the patient is suffering from chronic respiratory failure;
- VS: COPD: indicates if the patient has chronic obstructive pulmonary disease
- VT: HEMATOLOGIC DISEASE: indicates if the patient has some kind of hematologic disease;
- VU: CORTICOSTEROID THERAPY: indicates if the patient is having some corticosteroid therapy;
- VV: HTN: indicates if the patient has hypertension;
- VW: AVC sequelae: indicates if the patient has any AVC sequelae;
- VX: DIABETES: indicates if the patient has a treated or untreated diabetes;
- VY: PROVENANCE: indicates where the patient comes from;
- VZ: TYPEOFADMISSION: is the admission type and indicates if the patient were admitted urgently or if it was planned;
- VAA: TYPEOFADMISSIONSURGERY: indicates if the patient was admitted after surgery or not;
- VAB: TRANSPLANTED: indicates if the patient was admitted after transplantation;
- VAC: RISK: indicates if the patient had Stroke sequelae, transplantations, arterial hypertension, chronic renal insufficiency, chronic, cardiac, or chronic respiratory insufficiency or not;
- VAD: NEOPLASMS: indicates if the patient has any tumor or cancer, metastasized or not;
- VAE: CHEMOTHERAPY: indicates if the patient had any chemotherapy treatment or not;
- VAF: RADIOTHERAPY: indicates if the patient had any radiotherapy treatment or not; and
- VAG: OTHERIMUNOSSUPRESSANT: indicates if the patient had any other immunosuppressant treatment or not.

Due to the large number of used attributes, only the attributes of the best results of the modeling phase will be demonstrated. The distribution of values from the attributes with the best results are presented in Table 3.

**Table 3.** Best attributes distribution.

Attributes	Value 1	Value 0
TYPEADMISSION	68%	32%
TYPEADMISSIONSURGERY	70%	30%
TRANSPLANTED	8%	82%
RISK	48%	52%
NEOPLASMS	4%	96%
CHEMOTHERAPY	3%	97%
RADIOTHERAPY	1%	99%
OTHERIMUNOSSUPRESSANT	4%	96%

To verify data quality, a search for errors, data omissions, and data integrity was conducted, and then several solutions to correct the errors were proposed. The errors encountered were blank spaces, where data was not filled in by doctors or writing errors occurred.

Next step it was to convert false and true values to 0 and 1, correspondingly, to simplify the clustering process in the used tools. After this process, the best solutions were applied and the data were corrected, so they would be ready for the data mining process.

In addition to data correction, the attributes were analyzed and weighted for their importance for clustering analyses. From this point, some of the attributes were not included into the clustering due to their insignificance. For example, data referring to the bed identification number or the patient identification number assigned by the hospital are meaningless attributes. It is important to note that only data from admitted patients was used.

The total of attributes that composed the dataset after the data preparation phase were 34.



### 3.4. Modeling

The modelling phase began by the scenarios construction to generate varied models. Different attributes were grouped into different scenarios and clustering techniques was applied. The criteria used to build these scenarios were defined by the authors, according to the analyses made to the data structure. Being that this work is based on clinical evidence, the scenarios also were created using clinical input provided by the clinicians. The first set of attributes created was Case Mix, which contained all the attributes from the table used. The next scenarios were created by using some criteria, like the attributes targeting patients that had surgery, or patients that had neoplasms. Other scenarios (like (4) and (8)) were created to understand if any attribute had positive or negative impact in the results. By applying these criteria 10 scenarios were created. These scenarios were introduced into RapidMiner and Orange using “Admission” as the target and are listed below:

S1 = {All attributes};  
 S2 = {VZ; VAA; VA; VB; VAC; VY};  
 S3 = {VA; VY; VZ; VW; VAB};  
 S4 = {VAA; VZ; VAB; VAC};  
 S5 = {VO; VP; VQ; VR; VY};  
 S6 = {VAC; VA; VY};  
 S7 = {VO; VP; VQ; VR; VAA; VAB; VZ; VAC; VY};  
 S8 = {VZ; VAA; VAB};  
 S9 = {VAA; VAB; VZ; VM; VN; VT};  
 S10 = {VAD; VAE; VAF; VAG; VZ}.

These scenarios were constructed in RapidMiner which generated 10 models by applying Davies–Bouldin Index as the technique. These models can be represented by:

$$\text{DMM} = \{10 \text{ Scenarios}, 1 \text{ Technique}, 1 \text{ Representation Method}, 1 \text{ Target}\}$$

The same process of scenario construction was made with Orange, but it was applied the *K*-means algorithm using three techniques: silhouette, distance to centroids, and inter-cluster distance. The value of “*K*” was found through tests made with RapidMiner and Orange functionality which allows finding the optimum number of “*K*” clusters. With this tool 30 models were generated that can be represented by:

$$\text{DMM} = \{10 \text{ Scenarios}, 3 \text{ Techniques}, 1 \text{ Representation Method}, 1 \text{ Target}\}$$

For this paper the three scenarios that had better results were chosen (scenario (8), (4), and (10)). That is, the scenarios in which the silhouette value is closer to the value 1 and distance to centroids and inter-cluster distance values are smaller. Table 4 presents the attributes used on the three best scenarios, with the distinct values of each attributes, the average of the values, the minimum and maximum values of the corresponding value.

It is possible to see that two of the scenarios are very similar and that the only difference is the absence of one attribute. This difference was made to understand the impact of different attributes in the data mining process.

**Table 4.** Best three scenarios' attributes.

Scenario	Attribute	Distinct Values
4, 8, and 10	TYPEADMISSION	2 (0 or 1)
4 and 8	TYPEADMISSIONSURGERY	2 (0 or 1)
4 and 8	TRANSPLANTED	2 (0 or 1)
4	RISK	2 (0 or 1)
10	NEOPLASMS	2 (0 or 1)
10	CHEMOTHERAPY	2 (0 or 1)
10	RADIOTHERAPY	2 (0 or 1)
10	OTHERIMUNOSSUPRESSANT	2 (0 or 1)

### 3.5. Evaluation

This phase started with the assessment of results obtained in the previous phase. By analyzing each result achieved by applying the techniques mentioned before it was possible to make some useful observations. It should be noted that these results and the conclusions achieved are not supposed to make predictions but, instead, they are useful for intensivists to act in a proactive manner and being alert by knowing which type of patients are usually admitted to the ICU and prevent their condition from getting worse.

Table 5 presents the results of the top three scenarios corresponding to the silhouette, inter-cluster distance, and distance to centroids values. Silhouette values are between  $[-1, 1]$ , with 1 being the best value; in the case of inter-cluster distance and distance to centroids values that are the smaller the better.

**Table 5.** Silhouette, inter-cluster distance and distance to centroids results.

Scenarios	Silhouette	Inter-Cluster Distance	Distance to Centroids
8	1	1.5	$6.2 \times 10^{-17}$
10	0.93	1.4	0.04264
4	0.81	1.4	0.067

The results using Davies–Bouldin Index are presented in Table 6 and as it is possible to see the best results correspond to tests with two clusters. This table presents the three best scenarios, the best Davies–Bouldin index, and the corresponding best number of clusters.

**Table 6.** RapidMiner results.

Scenario	Davies–Bouldin Index	Number of Clusters
8	−0.652	2
4	−1.348	2
10	−0.369	2

Note for the Davies–Bouldin Index on RapidMiner that uses negative values. The justification for this is that it happens because when the density is low, the value is automatically assigned as negative. Therefore, the lower the value, the better the result.

Although, in Orange, the best results were verified in scenario (8), in RapidMiner, the best results with the Davies–Bouldin Index were verified in scenario (4).

## 4. Discussion

From the models created three scenarios were selected with the best results. By analyzing the previous results from the clustering tests, it was possible to conclude that TYPEADMISSION has a positive impact in the clusters formation due to its values distribution among clusters. Additionally,



it was discovered that the attribute RISK has a negative impact on the results, which can be verified by the improvement of scenarios (4) and (8) results when withdrawing the attribute. As can be seen among all of the scenarios, there are two major types of patient’s characteristics that should be analyzed when treating patients. Those types are related with patients with surgery or that received a transplant and patients with tumors/cancer. With these scenarios, it was possible to group patients into two groups (clusters) that can help intensivists to pay attention to their patients and prevent unplanned admissions to the ICU. For the best scenario, scenario (8), the possible values of each attribute that are presented were analyzed in the two clusters, which are exhibited in Table 7. As can be seen, the attribute TYPEADMISSION has its values clearly divided between the two clusters. Cluster 1 only has value 0, and Cluster 2 only has value 1.

**Table 7.** Values about groups of admitted patients.

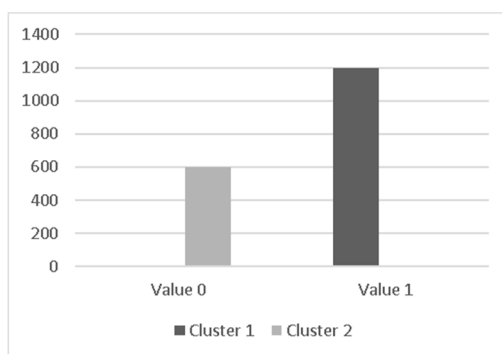
Attribute	Cluster 1	Cluster 2
TYPEADMISSION	0	1
TYPEADMISSIONSURGERY	0 and 1	0 and 1
TRANSPLANTED	0 and 1	0 and 1

Analyzing the distribution of values between the clusters, Cluster 1 has 1255 values, while Cluster 2 has the remaining 604 values. In Table 8 are shown the percentage of values (count) for the best scenario, scenario (8), for each cluster attribute and values. As can be seen, the attribute TYPEADMISSION is highly divided between the two clusters created, showing its importance to characterize the type of patient.

**Table 8.** Number of values from scenario (8) attributes for each cluster.

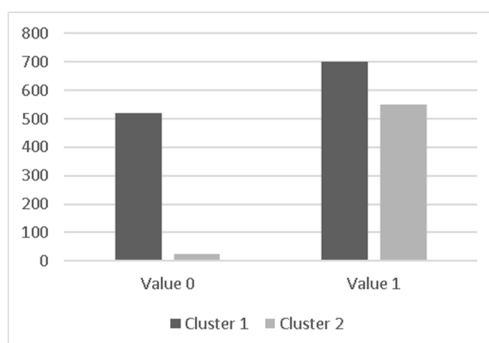
Attribute	Values	Cluster 1	Cluster 2
TYPEADMISSION	0	0%	100%
	1	100%	0%
TYPEADMISSIONSURGERY	0	42.87%	4.64%
	1	57.13%	95.36%
TRANSPLANTED	0	90.84%	95.20%
	1	9.16%	4.80%

Figure 1 shows the separation of TYPEADMISSION attributes with values 0 (scheduled) and 1 (urgent) between Cluster 1 and Cluster 2. As can be seen, there is a clear separation between values for the two clusters, showing that this is an important attribute to categorize admitted patients to ICU.



**Figure 1.** Distribution of values from TYPEADMISSION for each cluster.

Still in the analysis of the distribution of the attributes in clusters, it can be seen that the value 0 (scheduled) for the attribute TYPEADMISSIONSURGERY is not so present in Cluster 2, and the majority of this value is found in Cluster 1. This distribution is illustrated in Figure 2.



**Figure 2.** Distribution of values from TYPEADMISSION for each cluster.

## 5. Conclusions

This work provided useful results which help to characterize the type of patients that are admitted to the ICU. The clusters developed cannot ensure which patients will be admitted, but they give information about which type of patients need intensive care. This work is based on the study of the data collected from the patient admissions. The results achieved will require a clinical analysis from the experts (doctors) in order accomplish any clinical result. During the development, the models and the work were followed by an intensivist (expert in intensive care).

Analyzing the best scenario (8), characterized by patients that could have had surgery before admission to the ICU, the most important attributes are TYPEADMISSION, TYPEADMISSIONSURGERY, and TRANSPLANTED. The attribute RISK affects the results negatively, so it was not entered in this scenario.

The results and models obtained will be implemented and used to strengthen the Pervasive Clinical Decision Support System (also known as INTCare) and the Pervasive Business Intelligence System that are running in the ICU of Centro Hospitalar do Porto [29,30]. One of the main features of the INTCare system is presenting data mining results. Future work will include more studies about admitted patients to the ICU based on the data extracted. Other data mining techniques will be applied to predict if a patient will be admitted to the ICU, or not, to create valuable knowledge to hospital units and the data mining community.

Finally, in the future these data and the respective scenarios can be inserted into other tools that are more commonly used, such as R or Python, and, for example, the Pervasive Data Mining Engine [31].

**Acknowledgments:** This work has been supported by Compete: POCI-01-0145-FEDER-007043 and FCT within the Project Scope UID/CEC/00319/2013.

**Author Contributions:** This work was developed under the Ana Ribeiro master degree Work under the supervision and collaboration of Filipe Portela, Manuel Santos, and Fernando Rua. José Machado and António Abelha contributed to this work through the development of some features in AIDA platform. All authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Milovic, B.; Milovic, M. Prediction and decision making in Health Care using Data Mining. *Kuwait Chapter Arab. J. Bus. Manag. Rev.* **2012**, *1*, 126. [[CrossRef](#)]

2. Arabi, Y.; Venkatesh, S.; Haddad, S.; Malik, S.A.; Shimemeri, A.A. The characteristics of very short stay ICU admissions and implications for optimizing ICU resource utilization: The Saudi Experience. *Int. J. Qual. Health Care* **2004**, *16*, 149–155. [CrossRef] [PubMed]
3. Ramon, J.; Fierens, D.; Güiza, F.; Meyfroidt, G.; Blockeel, H.; Bruynooghe, M.; Van Den Berghe, G. Mining data from intensive care patients. *Adv. Eng. Inform.* **2007**, *21*, 243–256. [CrossRef]
4. Silva, Á.; Cortez, P.; Santos, M.F.; Gomes, L.; Neves, J. Rating organ failure via adverse events using data mining in the intensive care unit. *Artif. Intell. Med.* **2008**, *43*, 179–193. [CrossRef] [PubMed]
5. Braga, P.; Portela, F.; Santos, M.F.; Rua, F. Data Mining to Predict Patient's Readmission in Intensive Care Units. In Proceedings of the 6th International Conference on Agents and Artificial Intelligence (ICAART 2014), Loire Valley, France, 6–8 March 2014.
6. Veloso, R.; Portela, C.F.; Santos, M.F.; Silva, Á.; Rua, F.; Abelha, A.; Machado, J. Categorize readmitted patients in Intensive Medicine by means of Clustering Data Mining. *Int. J. E-Health Med. Commun.* **2015**, in press.
7. Ribeiro, A.; Portela, F.; Santos, M.F.; Machado, J.; Abelha, A.; Martins, F.R. Predicting Patients admission in Intensive Care Units using Data Mining. *POLIBITS* **2017**, in press.
8. Avaliação da Situação Nacional das Unidade de Cuidados Intensivos—Relatório final. Available online: <https://www.sns.gov.pt/wp-content/uploads/2016/05/Avalia%C3%A7%C3%A3o-nacional-da-situa%C3%A7%C3%A3o-das-unidades-de-cuidados-intensivos.pdf> (accessed on 16 February 2017).
9. Portela, F.; Santos, M.F.; Machado, J.; Abelha, A.; Silva, Á.; Rua, F. Pervasive and Intelligent Decision Support in Intensive Medicine—The Complete Picture. In *Information Technology in Bio- and Medical Informatics*; Springer: Cham, Switzerland, 2014.
10. Bersten, A.D.; Soni, N. *Oh's Intensive Care Manual*; Elsevier Health Sciences: London, UK, 2013.
11. de Saúde, D.-G.; de Planeamento, D.S. *Cuidados Intensivos: Recomendações para o seu Desenvolvimento*; Direcção-Geral da Saúde: Lisboa, Portugal, 2003. (In Portuguese)
12. "Medicine" in Oxford Dictionaries. Available online: <https://en.oxforddictionaries.com/definition/medicine> (accessed on 17 February 2017).
13. Caldeira, V.M.; Silva Júnior, J.M.; Oliveira, A.M.; Rezende, S.; Araújo, L.A.; Santana, M.R.; Amendola, C.P.; Rezende, E. Criteria for patient admission to an intensive care unit and related mortality rates. *Rev. Assoc. Med. Bras.* **1992**, *56*, 528–534. [CrossRef]
14. Xu, R.; Wunsch, D. *Clustering*; John Wiley & Sons: Hoboken, NJ, USA, 2008; p. 364.
15. Basu, S.; Davidson, I.; Wagstaff, K. *Constrained Clustering: Advances in Algorithms, Theory, and Applications*; CRC Press: Boca Raton, FL, USA, 2008; p. 472.
16. Anderberg, M.R. *Cluster Analysis for Applications: Probability and Mathematical Statistics: A Series of Monographs and Textbooks*; Academic Press: Amsterdam, The Netherlands, 2014.
17. Portela, F.; Santos, M.; Machado, J.; Abelha, A.; Silva, Á. Pervasive and Intelligent Decision Support in Critical Health Care Using Ensembles. In *Information Technology in Bio- and Medical Informatics*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 1–16.
18. Marins, F.; Cardoso, L.; Portela, F.; Santos, M.; Abelha, A.; Machado, J. Intelligent Information System to Tracking Patients in Intensive Care Units. In *Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction*; Springer: Cham, Switzerland, 2013; Volume 8276, pp. 54–61.
19. Veloso, R.; Portela, F.; Santos, M.F.; Silva, Á.; Rua, F.; Abelha, A.; Machado, J. A Clustering Approach for Predicting Readmissions in Intensive Medicine. *Proced. Technol.* **2014**, *16*, 1307–1316. [CrossRef]
20. Kamath, A.F.; Gutsche, J.T.; Kornfield, Z.N.; Baldwin, K.D.; Kosseim, L.M.; Israelite, C.L. Prospective Study of Unplanned Admission to the Intensive Care Unit after Total Hip Arthroplasty. *J. Arthroplast.* **2013**, *28*, 1345–1348. [CrossRef] [PubMed]
21. LabarÛre, J.; Schuetz, P.; Renaud, B.; Claessens, Y.-E.; Albrich, W.; Mueller, B. Validation of a Clinical Prediction Model for Early Admission to the Intensive Care Unit of Patients with Pneumonia. *Acad. Emerg. Med.* **2012**, *19*, 993–1003. [CrossRef] [PubMed]
22. Subbe, C.P.; Kruger, M.; Rutherford, P.; Gemmel, L. Validation of a modified Early Warning Score in medical admissions. *QJM* **2001**, *94*, 521–526. [CrossRef] [PubMed]

23. Tsai, J.C.-H.; Weng, S.-J.; Huang, C.-Y.; Yen, D.H.-T.; Chen, H.-L. Feasibility of using the predisposition, insult/infection, physiological response, and organ dysfunction concept of sepsis to predict the risk of deterioration and unplanned intensive care unit transfer after emergency department admission. *J. Chin. Med. Assoc.* **2014**, *77*, 133–141. [[PubMed](#)]
24. Van den Bosch, G.E.; Merkus, P.; Buysse, C.M.; Vaessen-Verberne, A.A.; Hop, W.C.; de Hoog, M. Risk Factors for Pediatric Intensive Care Admission in Children With Acute Asthma. *Respir. Care* **2012**, *57*, 1391–1397. [[CrossRef](#)] [[PubMed](#)]
25. Goldman, L.; Cook, E.F.; Johnson, P.A.; Brand, D.A.; Rouan, G.W.; Lee, T.H. Prediction of the need for intensive care in patients who come to emergency departments with acute chest pain. *New Engl. J. Med.* **1996**, *334*, 1498–1504. [[CrossRef](#)] [[PubMed](#)]
26. Brunelli, A.; Ferguson, M.K.; Rocco, G.; Pieretti, P.; Vigneswaran, W.T.; Morgan-Hughes, N.J.; Zanello, M.; Salati, M. A Scoring System Predicting the Risk for Intensive Care Unit Admission for Complications after Major Lung Resection: A Multicenter Analysis. *Ann. Thorac. Surg.* **2008**, *86*, 213–218. [[CrossRef](#)] [[PubMed](#)]
27. Portela, F.; Santos, M.F.; Machado, J.; Abelha, A.; Rua, F.; Silva, Á. Real-time Decision Support using Data Mining to predict Blood Pressure Critical Events in Intensive Medicine Patients. In *Ambient Intelligence for Health*; Springer: Cham, Switzerland, 2015.
28. Portela, F.; Santos, M.F.; Silva, Á.; Rua, F.; Abelha, A.; Machado, J. Preventing Patient Cardiac Arrhythmias by using Data Mining Techniques. In Proceedings of the IEEE Conference on Biomedical Engineering and Sciences (IECBES 2014), Sarawak, Malaysia, 8–10 December 2014.
29. Pereira, A.; Portela, F.; Santos, M.F.; Abelha, A.; Machado, J. Pervasive Business Intelligence: A New Trend in Critical Healthcare. *Proced. Comput. Sci.* **2016**, *98*, 362–367. [[CrossRef](#)]
30. Pereira, A.; Portela, F.; Santos, M.F.; Rua, F. Pervasive Business Intelligence in Intensive Medicine—An overview of clinical solution. 2017; In Submission.
31. Peixoto, R.; Portela, F.; Santos, M.F. Towards a Pervasive Data Mining Engine—Architecture Overview. In *New Advances in Information Systems and Technologies*; Springer: Cham, Switzerland, 2016; Volume 445, pp. 557–566.



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).