

Article

Predicting Intensive Care Unit Patients' Discharge Date with a Hybrid Machine Learning Model That Combines Length of Stay and Days to Discharge

David Cuadrado , Aida Valls * and David Riaño

Department of Computer Science and Mathematics, Universitat Rovira i Virgili, 43007 Tarragona, Spain; david.cuadrado@estudiants.urv.cat (D.C.)

* Correspondence: aida.valls@urv.cat

Abstract: Background: Accurate planning of the duration of stays at intensive care units is of utmost importance for resource planning. Currently, the discharge date used for resource management is calculated only at admission time and is called length of stay. However, the evolution of the treatment may be different from one patient to another, so a recalculation of the date of discharge should be performed, called days to discharge. The prediction of days to discharge during the stay at the ICU with statistical and data analysis methods has been poorly studied with low-quality results. This study aims to improve the prediction of the discharge date for any patient in intensive care units using artificial intelligence techniques. Methods: The paper proposes a hybrid method based on group-conditioned models obtained with machine learning techniques. Patients are grouped into three clusters based on an initial length of stay estimation. On each group (grouped by first days of stay), we calculate the group-conditioned length of stay value to know the predicted date of discharge, then, after a given number of days, another group-conditioned prediction model must be used to calculate the days to discharge in order to obtain a more accurate prediction of the number of remaining days. The study is performed with the eICU database, a public dataset of USA patients admitted to intensive care units between 2014 and 2015. Three machine learning methods (i.e., Random Forest, XGBoost, and lightGBM) are used to generate length of stay and days to discharge predictive models for each group. Results: Random Forest is the algorithm that obtains the best days to discharge predictors. The proposed hybrid method achieves a root mean square error (RMSE) and mean average error (MAE) below one day on the eICU dataset for the last six days of stay. Conclusions: Machine learning models improve quality of predictions for the days to discharge and length of stay for intensive care unit patients. The results demonstrate that the hybrid model, based on Random Forest, improves the accuracy for predicting length of stay at the start and days to discharge at the end of the intensive care unit stay. Implementing these prediction models may help in the accurate estimation of bed occupancy at intensive care units, thus improving the planning for these limited and critical health-care resources.



Citation: Cuadrado, D.; Valls, A.; Riaño, D. Predicting Intensive Care Unit Patients' Discharge Date with a Hybrid Machine Learning Model That Combines Length of Stay and Days to Discharge. *Mathematics* **2023**, *11*, 4773. <https://doi.org/10.3390/math11234773>

Academic Editor: Carmen Patino-Alonso

Received: 19 October 2023

Revised: 18 November 2023

Accepted: 22 November 2023

Published: 26 November 2023

Keywords: days to discharge prediction; intensive care unit; machine learning; classification

MSC: 68T05; 62H30; 62H12; 62P10



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Intensive care units (ICU) are hospital services that the World Federation of Societies of Intensive and Critical Care Medicine define as [1] “organized systems for the provision of care to critically ill patients that provide intensive and specialized medical and nursing care, and enhanced capacity for monitoring, and multiple modalities of physiologic organ support to sustain life during a period of life-threatening organ system insufficiency”.

ICU patients may have a wide variety of pathologies affecting one or more vital functions, which are potentially reversible. The Working Group on Quality Improvement

of the European Society of Intensive Care Medicine classified these patients [2] into two groups: those “requiring monitoring and treatment because one or more vital functions are threatened by an acute (or an acute on chronic) disease [...] or by the sequel of surgical or other intensive treatment [...] leading to life-threatening conditions” and those “already having failure of one of the vital functions [...] but with a reasonable chance of a meaningful functional recovery”. Patients in the end-stages of untreatable terminal diseases were left out of these groups.

More specific classifications distinguish between ICU patients, such as those requiring close monitoring, patients facing critical lung issues, patients with severe cardiac problems, and patients with serious infections. All this variability in admission extends throughout the patient’s stay in the ICU and is reflected in the significant disparity in patient evolution, treatments [3], outcomes, and costs [4,5]. Moreover, ICU resources, such as beds, are usually limited. In cases of unexpected increases in demand (e.g., during the COVID-19 pandemic, earthquakes, or other catastrophes), thorough understanding and planning of patient occupancy (i.e., days to discharge) become crucial for effective healthcare services. The days to discharge prediction is also essential for the proper management of an ICU in terms of bed occupancy, pharmacological and non-pharmacological stock availability, staff provision, flow of patients to and from other hospital units, etc. [6].

ICU patients are assessed in terms of demographic parameters such as gender or age at the ICU admission, in addition to some clinical measures. During their ICU stage, some other clinical parameters such as temperature (T), heart rate (HR), mean arterial pressure (MAP), or peripheral oxygen saturation (SpO₂) are systematically monitored, some of them continuously, some others at different discrete times during the day. These measurements are collected in the health information systems and are used for medical decision making [7]. The hypothesis of this work is that these values can also be used to foresee the days to discharge (DTD) of the patient.

DTD is closely related to the concept of length of stay (LOS) but, unlike this, DTD is not a constant parameter and it is not predicted on the patient condition in the first 24–48 h after admission, but with the evolving condition of the patient. A systematic review analyzed 31 LOS predictive models and concluded that they suffer from serious limitations [8]. Statistical and machine learning approaches (e.g., [9]) provide moderate predictions of short term LOS (1–5 days), but are unable to correctly predict long-term LOS (>5 days). Therefore, while LOS is essential for resource allocation during a patient’s initial admission, it becomes less informative as their stay progresses and their medical needs change.

Some of these methods predict the days of stay only on concrete patient types [10–17] (e.g., postsurgical or coronary diseases); others perform a classification into a few LOS intervals, such as distinguishing ICU discharges in less than two days in front of patients who stay longer [11,18,19].

The problem of estimating the LOS for any patient admitted at an ICU has also been approached using different kinds of statistical and machine learning methods. The oldest works used some statistical techniques and different kinds of regression models [20–23]. Some other approaches use traditional machine learning methods, such as Random Forest, Support Vector Machine, or Neural Network [12,18,20,24–27]. Some more recent works have also included advanced neural networks and deep learning techniques [28–30]. However, these LOS prediction methods only reached root mean square errors (RMSE) of 0.47–8.74 days and mean absolute errors (MAE) of 0.22–4.42 days [13,20,21,26]. It is worth noting that the study that obtained the best results [26] applied the concept of tolerance, meaning that errors which were proportionally below the tolerance level were discarded in the calculation of the average errors (i.e., LOS errors below $0.4 \times \text{LOS}$ did not count). The study with the second-best results [21] (RMSE 0.88 and MAE 0.87) worked with a dataset with multiple features with 44–50% missing values, whose management forcing a replacement value could have a high impact in LOS predictions.

In fact, some recent works [9] question the capacity of computer-based predictive models based only on the condition of the patient in the first 24–48 h after admission. Therefore, there is still a need to produce good, robust, and generic mathematical models to dynamically predict the days to discharge in ICUs. A good DTD model must have a low average error and must be robust with respect to ICU patient heterogeneity.

This diversity in the patients’ data was analyzed in [31], where four measures to characterize the ICU patient heterogeneity with respect to the DTD were described and applied in a small in-house hospital database. First, a graphical representation of the means and standard deviations of clinical parameters and severity scales was shown in a DTD_i group over time, serving as a valuable tool for analyzing patient heterogeneity and evolution in terms of complexity. Second, a cluster analysis was conducted, observing that it was difficult to distinguish groups of similar patients on each day of evolution. Finally, the DTD confusion matrix method was shown to be able to determine the number of patients discharged in i days who were clinically indistinguishable from other patients who were discharged in j days ($j \neq i$). The results on 3973 ICU patients with a mean stay of 8.56 days admitted to a tertiary hospital in Spain showed that, on average, 37% of the patients were clinically very similar to other patients who were discharged before and 26% to patients who were discharged later [31]. A preliminary work on constructing a DTD prediction model using machine learning was conducted in [32] with the same small in-house hospital dataset. The limitation on the number of features and the short length of stay of the examples (less than 14 days) hampered the quality of the models obtained, which had average error around 1.5 days, which is too large for proper personnel and resource planning at an ICU.

The goal of this paper is to improve the date of discharge prediction up to obtaining a mean absolute error below one day (i.e., almost perfect prediction) for a population of highly heterogeneous ICU patients. In this paper, we propose a new methodology that first divides patients into different classes in order to build different prediction models for each group using machine learning techniques. A hybrid group-conditioned model that combines LOS and DTD predictor models is defined and evaluated.

2. Methods and Technologies

The research work process of this study is depicted in Figure 1. It comprises two main stages: data preparation and prediction model construction. Details on all steps are given in this section.

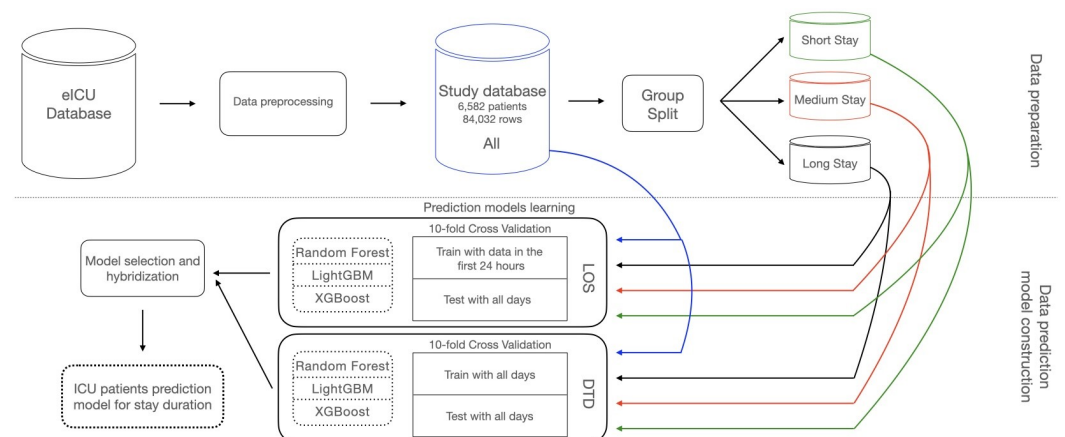


Figure 1. Research work process for building an ICU date of discharge prediction model.

2.1. ICU Data Cohort

The eICU Collaborative Research Database [33] is a public dataset that includes patients admitted in the ICUs across the United States between 2014 and 2015. Only patients discharged alive were considered. Patients discharged on the same day of their admission were not considered. In eICU, this cohort encompasses 16,585 patients with a total of

84,032 rows corresponding to each of the days of treatment. From the data available, due to the need to have daily values, some features were discarded. The rest were selected based on previous studies [32].

Table 1 gives some descriptors of the variables, for the whole dataset in the column All and for each of the three subsets. For numerical (type N) and for scale features (type S), we give the mean, standard deviation, minimum, and maximum. For example, for All data, we have an average age of 63.2 years, with a standard deviation of 17, a minimum value of 18, and a maximum value of 90 years. For categorical features (type C), we give the percentage of each of the possible categories of the variable. For instance, in the All dataset, 54.4% are male and 45.61% are female, and, regarding UT, 46.07% of patients are admitted in Medical–Surgical Intensive Care Unit (MSICU), 13.67% in Neurological Intensive Care Unit (NICU), 11.07% in Medical Intensive Care Unit (MICU), and 29.19% are under other ICUs (e.g., CCU–CTICU, SICU, Cardiac ICU, etc.). In addition to the information in that table, regarding the duration of the stay, 26.30% of patients remain one day, 11.08% two days, 9.98% three days, 9.41% four days, 8.37% five days, 7.91% six days, 6.31% seven days, and 20.65% remain eight days or more. The average length of stay is 5 days.

Table 1. eICU dataset features description. For numerical features (N) we have Age, Temperature (T), SpO2, Heart rate (HR), and Mean Arterial Pressure (MAP), and each cell has its mean, stdev, minimum, and maximum values. For categorical data (C), each cell has the percentage of each category for Gender, Unit Type (UT = (MSICU/NICU/MICU/Other)), and mechanical ventilation invasive and non-invasive (MVI/MVNI). The two scale features (S) are Glasgow Coma Score (GCS) and Pain Score; cells show mean, stdev, minimum, and maximum values.

Type	Freq.	Name	All	Short Stay	Medium Stay	Long Stay
N	E	Age	63.2; 17.0; 18.0; 90.0	63.8; 16.9; 18.0; 90.0	63.6; 16.0; 18.0; 90.0	63.1; 16.8; 18.0; 90.0
	D	Avg. T	36.9; 0.5; 32.6; 40.0	36.9; 0.5; 32.6; 39.9	36.9; 0.5; 32.6; 39.9	36.9; 0.5; 32.6; 40.0
		Min. T	36.5; 0.7; 25.0; 39.9	36.4; 0.6; 25.0; 39.3	36.5; 0.6; 25.0; 39.3	36.5; 0.7; 25.0; 39.4
		Max. T	37.4; 0.7; 32.6; 43.6	37.3; 0.6; 33.0; 42.0	37.3; 0.7; 33.0; 43.6	37.4; 0.7; 33.0; 43.6
		Avg. SpO2	96.7; 2.1; 80.4; 100.0	96.6; 2.1; 80.0; 100.0	96.7; 2.1; 80.1; 100.0	96.7; 2.1; 81.5; 100.0
		Min. SpO2	92.2; 4.2; 80.0; 100.0	92.2; 4.1; 80.2; 100.0	92.1; 4.2; 81.1; 100.0	92.1; 4.2; 81.4; 100.0
		Max. SpO2	99.2; 1.4; 81.1; 100.0	99.0; 1.5; 80.2; 100.0	99.1; 1.5; 80.9; 100.0	99.1; 1.4; 81.9; 100.0
		Avg. MAP	79.6; 9.6; 36.3; 125.0	79.7; 10.1; 36.3; 124.3	79.6; 9.8; 36.3; 124.3	79.6; 9.7; 36.3; 125.9
		Min. MAP	68.5; 11.2; 29.0; 125.0	69.0; 11.8; 29.0; 124.3	68.6; 11.4; 29.0; 124.3	68.5; 11.3; 29.0; 124.2
		Max. MAP	92.7; 11.5; 36.3; 129.7	92.5; 12.0; 36.3; 129.3	92.7; 11.7; 36.3; 129.7	92.8; 11.6; 36.3; 129.7
		Avg. HR	86.2; 15.7; 36.0; 140.0	85.2; 15.6; 37.0; 139.0	85.8; 15.6; 37.0; 139.3	86.0; 15.6; 37.0; 139.3
		Min. HR	73.4; 15.3; 36.0; 137.4	73.0; 15.4; 36.0; 138.1	73.2; 15.4; 36.0; 139.1	73.3; 15.3; 36.0; 141.8
		Max. HR	102.7; 18.9; 36.0; 139.4	101.0; 18.9; 38.0; 141.9	102.2; 18.9; 38.0; 139.2	102.6; 18.8; 38.0; 142.4
		C	E	Gender(M/F) %	54.4/45.6	53.0/47.0
UT %	46.1/13.7/11.1/29.2			46.4/11.3/11.7/30.6	46.8/12.4/11.5/29.3	46.5/13.5/11.4/30.0
D	MVI (0/1) %		91.8/8.2	94.0/6.0	92.4/7.6	92.0/8.0
	MVNI (0/1) %		98.1/1.9	97.8/2.2	97.9/2.1	98.1/1.9
S	D	Avg. GCS	12.5; 3.3; 3.0; 15	13.4; 2.7; 3.0; 15.0	12.9; 3.1; 3.0; 15.0	12.6; 3.3; 3.0; 15.0
		Min. GCS	13.4; 2.6; 3.0; 15.0	14.1; 2.0; 3.0; 15.0	13.7; 2.4; 3.0; 15.0	13.5; 2.6; 3.0; 15.0
		Max. GCS	12.8; 2.9; 3.0; 15.0	13.7; 2.4; 3.0; 15.0	13.1; 2.7; 3.0; 15.0	12.9; 2.9; 3.0; 15.0
		Avg. Pain	0.4; 1.4; 0.0; 10.0	0.5; 1.5; 0.0; 10.0	0.4; 1.4; 0.0; 10.0	0.4; 1.4; 0.0; 10.0
		Min. Pain	1.5; 2.9; 0.0; 10.0	1.8; 3.2; 0.0; 10.0	1.6; 3.0; 0.0; 10.0	1.6; 3.0; 0.0; 10.0
		Max. Pain	0.7; 1.7; 0.0; 10.0	0.9; 1.9; 0.0; 10.0	0.8; 1.8; 0.0; 10.0	0.7; 1.7; 0.0; 10.0

2.2. Patient Grouping

In a previous work [32], we confirmed the complexity of estimation of the date of discharge, in the general case of any admission at ICU, due to the high heterogeneity in the data. Different grouping mechanisms were studied, based on biomarkers and clinical-based phenotypes, but none of them generated a partition that reduced heterogeneity in the groups. In this work, we make the hypothesis that a grouping of the data based on

the length of stay in the ICU could generate more accurate prediction models. A weekly time span is considered as appropriate for this problem. Consequently, three subgroups are defined: short, medium, and long stays. Short stays encompass patients with a length of stay up to seven days. Medium stays encompass patients with a length of stay up to 14 days (which includes all patients from short stays and also patients with a length of stay between 8 and 14 days). Long stays encompass patients with a length of stay up to 21 days (which includes all patients from short and medium stays and also patients with a length of stay between 15 and 21 days). Patients with a discharge on the same day as the admission (DTD = 1) were excluded for not being clinically relevant.

For the eICU database, the subgroup information is given in Table 2. Short stays encompass 8,799 patients, medium stays 11,432 patients, and long stays 11,981 patients. The average LOS for short stays is 4.21 days, for medium stays is 5.54 days, and for long stays is 6.07 days. The statistical values of the used features for each of these groups are given in Table 2.

Table 2. Subgroups based on LOS. N is the total number of days for the patients in the group. The last column shows the average LOS.

Group	Interval	Patients	N	Avg. LOS
Short stay	LOS > 1 and LOS ≤ 7	8799	37,014	4.21
Medium stay	LOS > 1 and LOS ≤ 14	11,432	63,298	5.54
Long stay	LOS > 1 and LOS ≤ 21	11,981	72,761	6.07

2.3. Measuring Patients' Heterogeneity

An heterogeneity analysis is performed in order to study the characteristics of the proposed grouping with respect to the data available for each patient on each of his/her days of stay at the ICU. Several measures defined in a previous paper [31] are applied to the whole dataset and to each of the three proposed groups.

These heterogeneity indicators are based on a similarity measure that is defined as the root mean square error between two patients' conditions on different days in Equation (1), where m is the number of clinical parameters and $d_{ij} = (v_{ij1}, \dots, v_{ijm})$ is the vector of normalized values of the j -th patient on his/her i -th day before discharge. The similarity measure for two values in a given k feature, $s_k(v_{ijk}, v_{i'j'k})$, is calculated as the Euclidean distance if the values are numerical or scale, and with Manhattan distance if they are categorical.

$$sim(d_{ij}, d_{i'j'}) = \sqrt{\frac{1}{m} \cdot \sum_{k=1}^m (s_k(v_{ijk}, v_{i'j'k}))^2} \tag{1}$$

Patient conditions with a similarity value of 99% or higher are considered clinically equivalent. Then, we calculate two confusion ratios: premature discharge ($p_\delta(i)$), Equation (2), which is the risk of discharging patients before day i , and overdue discharge ($o_\delta(i)$), Equation (3), which is the risk of discharging after day i . Both are calculated as the average number of patients discharged in i days who have an equivalent clinical condition to other patients that were discarded from the ICU in more (or fewer) than i days. Both are calculated as a proportion, where $n_\delta(i, j)$ is the number of patient descriptions with a similarity greater than 99%, and ℓ being the maximum number of days of stay in the analyzed group.

$$p_\delta(i) = \frac{\sum_{j < i} n_\delta(i, j)}{\sum_{j=1}^{\ell} n_\delta(j, i)} \tag{2}$$

$$o_\delta(i) = \frac{\sum_{j > i} n_\delta(i, j)}{\sum_{j=1}^{\ell} n_\delta(j, i)} \tag{3}$$

Afterwards, three measures of cluster cohesion/separation are used for each of the groups: Davies–Bouldin (DB), Dunn (D), and average silhouette (S)—Equations (4)–(6).

Clusters of patients discharged in i days are taken within each group, in order to quantify the degree of heterogeneity in the clinical conditions of the patients that are discharged in the same number of days. In this DTD-based clustering, let us denote as C_i the representative patient of the i -th cluster (i.e., the one with greatest similarity to the rest of patients with the same number of days to discharge), and let us define $m_i(d)$ as the average similarity sim of a patient with description d to any other patient, as defined in Equation (1). Then, the indices can be defined as

$$DB = \frac{1}{\ell} \sum_{i=1}^{\ell} \max_{1 \leq j \neq i \leq \ell} \frac{2 - (m_i(C_i) + m_j(C_j))}{1 - sim(C_i, C_j)} \tag{4}$$

$$D = \frac{1 - \max_{1 \leq i < j \leq \ell} \{sim(C_i, C_j)\}}{1 - \min_{1 \leq i \leq \ell} \{m_i\}} \tag{5}$$

$$S = \frac{1}{\sum_{i=1}^{\ell} n_i} \cdot \sum_{i=1}^{\ell} \sum_{d \in RD_i} \frac{m_i(d) - m(d)}{1 - \min\{m(d), m_i(d)\}} \tag{6}$$

These are cluster validity indices that are aimed at the quantitative evaluation of the results of a clustering, mainly focusing on the fact that different clusters must be distinguishable because of a small similarity between them [34]. A characterization of such indices can be found in [35]. The Davies–Bouldin index provides a positive value, which is higher as heterogeneity increases. The Dunn index takes positive values, where the higher the value, the more compact and separated the groups. Silhouette values are in the range $[-1, +1]$, with values below 0.25 considered to reflect a high heterogeneity in the data.

2.4. Method Proposed for Date of Discharge Prediction

The most widely used approach for predicting patient stays in ICU with the use of artificial intelligence models that consider only the data captured in the first 24–48 h after admission does not give enough accuracy. Without undermining the importance of this type of prediction, it does not seem justified that, as the clinical conditions of the patients evolve, the new information about the patients’ states is not taken into account to dynamically predict the DTD of the corresponding patients.

The first hypothesis of this study is that LOS can be an initial good estimate of the date of discharge, but, after a certain number of days ($\alpha_{g(p)}$), clinicians should make a dynamic prediction of the DTD in order to obtain a new (more accurate) prediction date, since the DTD model is trained using the daily clinical condition of a patient along their full ICU stay.

The second hypothesis is that building different prediction models for 3 different subgroups of patients, based on the number of days of stay of an ICU patient (short, medium, and long), the differentiation of the day of discharge should be more clear and, thus, it may decrease the overall RMSE and MAE errors.

Consideration of these different group-conditioned prediction models should enable the determination when the DTD model prediction obtains better results than LOS model prediction. Therefore, we propose a hybrid prediction model for ICU patients using conditioned LOS models at the beginning of stay and DTD models at the end of the stay. It is formalized as follows. Let us denote the following:

- $Y_{LOS} = f(x_1, \dots, x_n)$ is the prediction model of length of stay using the n data values of the variables at admission time, created with data from all the ICU patients;
- $Y_{LOS|g_j} = f(x_1, \dots, x_n)$ is the prediction model of length of stay using the n data values of the variables at admission time, created with data from the ICU patients that belong to group $g_j = \{Short, Medium, Long\}$;
- $Y_{DTD|g_j} = f(x_1, \dots, x_n)$ is the prediction model of days to discharge using the n data values of the variables during all the stay, created with data from the ICU patients that belong to group $g_j = \{Short, Medium, Long\}$.

The proposed hybrid system divides patients based on weeks. This criterion was determined by the planning organizational strategy of medical personnel. The proposed procedure is the following:

1. When a patient p is admitted at ICU, we calculate $Y_{LOS}(p)$ and assign the patient to one of these three groups as follows:
 - If $Y_{LOS}(p) \leq 7$, then he/she belongs to the group $g(p) = Short$;
 - If $Y_{LOS}(p) > 7$ and ≤ 14 , then he/she belongs to the group $g(p) = Medium$;
 - If $Y_{LOS}(p) > 14$, then he/she belongs to the group $g(p) = Long$;
2. Make the prediction $Y_{LOS|g(p)}$ and keep it for a number of days $Y_{LOS|g(p)}(p) - \alpha_{g(p)}$;
3. Afterwards, in the final $\alpha_{g(p)}$ days, make a prediction with $Y_{DTD|g(p)}$.

The first step is used to assign the patient to one of the subgroups, so that we can use specific conditioned models of prediction, built with the data from patients with a similar length of stay. Steps 2 and 3 use these group-conditioned models. During the first days of stay, we assume that the dynamic information may not be sufficiently accurate and representative to be used in the prediction model (due to high changes on clinical conditions among the ICU patients until stabilized). Therefore, the dynamic models of DTD are only used for the last α_{g_j} days of the stay, where the date must be more accurately calculated to be closer to the real date. This threshold value α_{g_j} for each group is obtained empirically from the analysis of the training dataset.

2.5. Machine Learning Methods for Constructing DTD and LOS Prediction Models

In order to construct the mentioned DTD and LOS prediction models, we selected three of the most successful machine learning algorithms. These algorithms were Random Forest [36], XGBoost [37], and lightGBM [38]. The prediction methods were implemented using the libraries sklearn, lightgbm, and xgboost. All models were constructed using a 10-fold cross-validation method for testing and training with the features described in Table 1.

For LOS predictions, the methods were trained with the data from the patients in their first 24 h in the ICU for all patients together and also for the three subgroups separately. However, for DTD prediction, models were trained not only on the patient conditions 24–48 h after ICU admission, but also on the daily clinical condition of patients along their full ICU stay. Each day of stay in the database is treated as an independent patient with a set of attributes; therefore, temporal patterns have not been captured.

A hyperparameter optimization was performed for all algorithms using the Grid Search algorithm. The parameters used for the prediction models are described in Table 3.

Table 3. Parameters optimized for the prediction models.

Models	Parameters
Random Forest	n.estimators = 170, max.depth = 80, max.features = 5, min.samples split = 5
lightGBM	boosting type = 'gbdt', num.leaves = 131, max.depth = -1, learning rate = 0.1, n.estimators = 100, subsample for bin = 2000, min.split gain = 0.0, min.child weight = 0.001, min.child samples = 20, subsample = 1.0, subsample freq. = 0, colsample bytree = 1.0, importance type = 'split'
XGBoost	base score = 0.3, booster = 'gbtree', n.estimators = 208, max.depth = 5, learning rate = 0.1, n.jobs = 1, objective = 'reg: squarederror', verbosity = 1

2.6. Evaluation Measures

To evaluate the quality of the prediction, we calculated the mean absolute error (MAE) and the root mean square error (RMSE) obtained with the three different training algorithms.

A 10-fold cross-validation method was used to obtain the corresponding mean predictive errors RMSE and MAE for each i -th day before discharge ($i = 2, \dots, 21$). MAE and RMSE were calculated for both the LOS and DTD models, for the whole data set, and also for each of the three subgroups separately.

An aggregated RMSE and MAE for each model $X = \{LOS, DTD\}$ was then obtained by averaging the errors of each day, weighted by the number of patients' data on the i -th day in the corresponding j -th group $w_{i,j}$.

$$RMSE(Y_{X|g_j}) = \frac{\sum_{i=1}^{\ell} RMSE(Y_{X|g_j}(i)) * w_{i,j}}{\sum_{i=1}^{\ell} w_{i,j}} \tag{7}$$

The error in the hybrid prediction method H , which combines LOS and DTD models on different days, is evaluated by taking the error in the model used in each of the days of stay of the patients in a group g_j . It is defined as follows:

$$RMSE^H(g_j) = \frac{\sum_{i=1}^{\alpha_j} RMSE(Y_{DTD|g_j}(i)) * w_{i,j} + \sum_{i=\alpha_j+1}^{\ell} RMSE(Y_{LOS|g_j}(i)) * w_{i,j}}{\sum_{i=1}^{\ell} w_{i,j}} \tag{8}$$

3. Results

This section analyzes the results of the different methodological steps presented in the previous section. First, the values in Table 1 about the analysis of the values of the features are discussed. Second, the indicator values of patient heterogeneity are presented in Table 4. Third, the results of training the models for calculating length of stay (Y_{LOS} and $Y_{LOS|g_j}$) are presented. Next, the results of the models obtained for days to discharge (Y_{DTD} and $Y_{DTD|g_j}$) are discussed for each subgroup. A comparison of LOS and DTD models is made to establish the threshold parameter α_{g_j} for each group. Finally, the results of the proposed hybrid group-conditioned method are given, showing its good performance in terms of small error in prediction.

Table 4. Data and subgroup heterogeneity values.

Metric	All	Short Stay	Medium Stay	Long Stay
Premature discharge	9.10%	17.36%	13.01%	9.84%
Overdue discharge	7.40%	12.43%	10.00%	9.81%
Davies–Bouldin	63.61	49.93	67.07	81.34
Dunn	0.0007	0.0012	0.0011	0.001
Silhouette	−0.26	−0.02	−0.04	−0.07

3.1. Groups of Patients

The description of the features for the whole dataset (All) shown in Table 1 can be taken as reference. They consist of the mean, standard deviation, min, and max values of the 13 numerical variables (type N), the 6 scale variables (type S), and the percentages of the most frequent values of the 4 categorical (type C) variables for the 84,032 days of treatment of all the patients. These values allow the cohort comparison of the different groups of ICU patients (short, medium, and long stays), with the whole set of patients in the ICU, from a population point of view.

In general, the numerical variables of all the groups have similar values to the ones observed for the whole population of patients. Scale features show the same similarities, with some exceptions in the average values of GCS_avg, GCS_min, and GCS_max (higher than the rest of the subgroups and column All).

Categorical features show more variations between column All and subgroups. The three subgroups present a lower number of patients admitted in Medical–Surgical Intensive Care Units (MICUs) and a higher number admitted into Neurological Intensive Care Units (Neuro ICUs) and Medical Intensive Care Units (MICUs) with respect to column All. Mechanical ventilation invasive (MVI) also shows a difference between the subsets. Short stays

present higher values (3 points above column All), while medium and long stays present lower values (2 and 4 points below column All, respectively). Non-invasive mechanical ventilation (MVNI) also shows lower values for medium and long stays (2 points below column All in both cases).

3.2. Patients' Heterogeneity

This study was carried out for the whole data set and also within each one of the three subgroups defined in Table 2. The heterogeneity values obtained are included in Table 4.

For the whole (All) dataset, we have that, in 9.10% of the days, patients are 99% similar to other patients that were discharged later, and 7.40% of the days patients are similar to other patients discharged earlier. A Davies–Bouldin value of 63.61 and a Silhouette value of -0.26 confirms the high heterogeneity in ICU patients.

When considering the three subgroups, we have that, for the short stay, in 17.36% of the days, patients are similar to other patients discharged later, and 12.43% to patients discharged before. Heterogeneity in terms of the Davies–Bouldin, Dunn, and silhouette is smaller in short stays than in the rest of the groups. Silhouette index is the one that finds more cohesion when working with three subgroups in comparison to having the dataset as a whole. As expected, we obtained the highest heterogeneity in the long stay group, as it includes any patient with an LOS below 22 days, with scores of 81.34 in Davis–Boulding and silhouette of -0.07 . The higher compactness of the short and medium groups is encouraging for finding appropriate prediction models for these groups.

3.3. LOS Prediction

LOS predictive models for patients in ICU were obtained using the patient descriptions of 16,585 days of treatment in the eICU dataset and also for the three subgroups (see Table 2). In these LOS models, training is made using only the data from the patients in their first 24 h. Results are gathered in Tables 5–7. In bold, we highlighted the best predictions for each day of stay (i.e., each row).

Table 5. RMSE and MAE of the Random Forest model for LOS prediction for each number of remaining days (R.D.). Best predictions are highlighted in bold.

R.D.	All		Short Stay		Medium Stay		Long Stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	1.2	0.9	0.95	0.84	1.45	1.26	1.63	1.4
3	1.16	0.8	0.59	0.49	1.18	0.98	1.43	1.16
4	0.99	0.69	0.29	0.21	0.87	0.65	1.14	0.84
5	0.99	0.72	0.41	0.31	0.67	0.47	1.01	0.69
6	0.95	0.73	0.79	0.68	0.54	0.42	0.82	0.57
7	1.13	0.87	1.25	1.12	0.78	0.6	0.81	0.62
8	1.31	0.98			1.06	0.83	0.98	0.75
9	1.63	1.25			1.44	1.2	1.27	0.98
10	1.95	1.57			1.84	1.6	1.59	1.29
11	2.27	1.88			2.25	1.99	1.95	1.63
12	2.77	2.37			2.71	2.43	2.44	2.11
13	3.27	2.84			3.25	2.93	2.94	2.58
14	3.46	3.04			3.53	3.19	3.16	2.79
15	4.04	3.58					3.73	3.33
16	4.36	3.9					4.09	3.67
17	4.7	4.22					4.41	3.99
18	5.03	4.52					4.83	4.36
19	5.77	5.22					5.52	5
20	6.25	5.68					5.89	5.35
21	6.47	5.85					6.19	5.62

Table 6. RMSE and MAE of the lightGBM model for LOS prediction for each number of remaining days (R.D.). Best predictions are highlighted in bold.

R.D.	All		Short Stay		Medium Stay		Long Stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	1.45	1.14	0.94	0.8	1.62	1.44	1.87	1.66
3	1.36	0.99	0.62	0.46	1.27	1.03	1.53	1.24
4	1.24	0.94	0.39	0.29	0.96	0.71	1.23	0.91
5	1.3	1.02	0.5	0.37	0.8	0.59	1.09	0.78
6	1.38	1.1	0.81	0.64	0.77	0.6	0.99	0.75
7	1.67	1.32	1.12	0.89	1.06	0.83	1.14	0.9
8	1.94	1.5			1.36	1.06	1.36	1.07
9	2.28	1.78			1.72	1.38	1.68	1.32
10	2.65	2.13			2.11	1.76	2.02	1.6
11	2.99	2.45			2.49	2.14	2.36	1.92
12	3.63	3.08			3.02	2.64	2.92	2.44
13	4.28	3.69			3.69	3.28	3.54	3.02
14	4.29	3.69			3.83	3.39	3.5	2.95
15	5.09	4.48					4.25	3.67
16	5.37	4.77					4.55	3.94
17	5.57	5.01					4.75	4.16
18	6.19	5.41					5.24	4.61
19	7.11	6.51					6.15	5.5
20	7.59	7					6.48	5.82
21	7.88	7.1					6.88	6.12

Table 7. RMSE and MAE of the XGBoost model for LOS prediction for each number of remaining days (R.D.). Best predictions are highlighted in bold.

R.D.	All		Short Stay		Medium Stay		Long Stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	2.04	1.69	1.64	1.59	2.52	2.42	2.84	2.7
3	1.92	1.43	0.98	0.87	1.95	1.76	2.37	2.1
4	1.67	1.28	0.5	0.4	1.39	1.1	1.83	1.46
5	1.68	1.38	0.76	0.62	1.07	0.83	1.54	1.13
6	1.76	1.47	1.42	1.28	1.02	0.85	1.26	1.02
7	2.19	1.81	2.13	1.97	1.51	1.26	1.49	1.26
8	2.6	2.09			2.03	1.71	1.9	1.55
9	3.14	2.58			2.68	2.36	2.45	2.05
10	3.78	3.28			3.36	3.1	3.1	2.75
11	4.3	3.81			4	3.75	3.68	3.31
12	5.2	4.74			4.8	4.55	4.54	4.2
13	6.15	5.69			5.76	5.48	5.47	5.11
14	6.3	5.85			6.05	5.77	5.71	5.34
15	7.49	7.07					6.77	6.42
16	8.01	7.56					7.35	6.96
17	8.29	7.83					7.68	7.31
18	8.92	8.48					8.44	8.05
19	10.33	9.92					9.66	9.29
20	10.98	10.53					10.17	9.81
21	11.09	10.34					10.12	9.47

From the previous tables we can see that in the short stay subgroup LightGBM slightly outperforms the rest for the last day of stay in ICU, while Random Forest shows the best results for every day in the other subgroups. The average and deviation of the RMSE and MAE for each of the 4 groups is given in Table 8. We can see that Random Forest method generally outperforms the other methods for medium and long stays, and remains with similar results for short stays.

Table 8. RMSE and MAE average and standard deviation for LOS prediction for each model.

Model	Measure	All		Short Stay		Medium Stay		Long Stay	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Random Forest	avg.	2.0	1.6	0.7	0.6	1.3	1.1	1.8	1.5
	st.dev.	1.4	1.3	0.6	0.4	0.9	0.8	1.3	1.2
LightGBM	avg.	2.5	2.1	0.7	0.6	1.6	1.3	2.1	1.7
	st.dev.	1.7	1.5	0.4	0.3	1.0	0.9	1.4	1.3
XGBoost	avg.	3.6	3.2	0.7	0.6	2.1	1.8	3.2	2.9
	st.dev.	2.5	2.4	0.7	0.8	1.7	1.5	2.3	2.3

3.4. DTD Prediction

We also obtained DTD predictors by training the Random Forest, the XGBoost, and the lightGBM algorithms using all the data of the patients (84,032 days of treatment) during their stay in the ICU. The average errors (RMSE and MAE) after 10-fold cross-validation are shown in Tables 9–11. In bold, we highlighted the best predictions for each day of stay (i.e., each row).

Random Forest is the algorithm that obtains the best DTD predictors in all subgroups, with the exception of short stays, where MAE and RMSE values in the lightGBM model outperform for 2, 5, 6 and 7 days. XGBoost is the worst for DTD in all subgroups, producing models with an MAE and RMSE above 1 day for short stays, above 6 days in medium stays, and above 10 days for long stays. For the RMSE values, the average difference between Random Forest and lightGBM is 0.06 for short stays (with lightGBM outperforming above 5 days), 0.22 for medium stays, and 0.30 for long stays (with Random Forest outperforming every day in both groups). For the MAE values, the average difference between Random Forest and lightGBM is 0.09 for short stays (with lightGBM outperforming above 5 days), 0.19 for medium stays, and 0.27 for long stays (with Random Forest outperforming every day in both groups). The whole data set (group All) shows MAE and RMSE values below 1 day between days 3 and 6, but there is always an MAE and RMSE value in some of the other subgroups with better performance than the error obtained in the whole data set.

Table 9. RMSE and MAE of the Random Forest model for DTD prediction for each number of remaining days (R.D.). Best predictions are highlighted in bold.

R.D.	All		Short Stay		Medium Stay		Long Stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	1.15	0.9	0.56	0.4	0.96	0.74	1.23	0.97
3	0.98	0.68	0.41	0.31	0.77	0.54	1.04	0.75
4	0.85	0.58	0.48	0.37	0.65	0.47	0.89	0.62
5	0.79	0.59	0.71	0.57	0.66	0.51	0.8	0.59
6	0.86	0.66	0.98	0.84	0.83	0.64	0.81	0.62
7	1.06	0.82	1.25	1.12	1.11	0.87	0.99	0.76
8	1.32	1.02			1.44	1.18	1.26	0.96
9	1.66	1.33			1.81	1.54	1.6	1.27
10	2.02	1.69			2.16	1.89	1.95	1.62
11	2.4	2.05			2.53	2.24	2.33	1.99
12	2.82	2.46			2.91	2.6	2.76	2.4
13	3.22	2.84			3.33	2.99	3.15	2.78
14	3.55	3.15			3.54	3.2	3.49	3.09
15	3.95	3.51					3.88	3.45
16	4.37	3.91					4.3	3.84
17	4.7	4.22					4.64	4.16
18	5.21	4.69					5.15	4.63
19	5.75	5.19					5.64	5.09
20	6.23	5.62					6.02	5.46
21	6.42	5.83					6.22	5.64

Table 10. RMSE and MAE of the LightGBM model for DTD prediction for each number of remaining days (R.D.). Best predictions are highlighted in bold.

R.D.	All		Short Stay		Medium Stay		Long Stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	1.4	1.13	0.53	0.37	1.12	0.89	1.49	1.22
3	1.17	0.86	0.44	0.32	0.89	0.66	1.22	0.92
4	1.08	0.82	0.5	0.37	0.82	0.63	1.08	0.82
5	1.14	0.9	0.68	0.49	0.92	0.72	1.09	0.85
6	1.32	1.05	0.88	0.67	1.13	0.88	1.22	0.96
7	1.61	1.27	1.12	0.89	1.44	1.14	1.48	1.17
8	1.92	1.53			1.77	1.42	1.78	1.44
9	2.3	1.86			2.15	1.76	2.17	1.81
10	2.68	2.22			2.51	2.1	2.54	2.15
11	3.09	2.62			2.87	2.49	2.96	2.51
12	3.59	3.08			3.25	2.88	3.44	2.96
13	3.96	3.45			3.72	3.3	3.8	3.32
14	4.3	3.77			3.87	3.48	4.15	3.62
15	4.75	4.18					4.58	4.02
16	5.18	4.6					5.01	4.43
17	5.45	4.88					5.29	4.73
18	6.02	5.43					5.9	5.31
19	6.79	6.15					6.57	5.96
20	7.11	6.46					6.88	6.28
21	7.39	6.64					7.08	6.36

Table 11. RMSE and MAE of the XGBoost model for DTD prediction for each number of remaining days (R.D.). Best predictions are highlighted in bold.

R.D.	All		Short Stay		Medium Stay		Long Stay	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2	2.09	1.75	0.98	0.77	1.76	1.46	2.24	1.91
3	1.78	1.34	0.73	0.60	1.39	1.07	1.88	1.45
4	1.59	1.21	0.88	0.72	1.21	0.96	1.63	1.24
5	1.57	1.28	1.30	1.09	1.32	1.09	1.52	1.24
6	1.78	1.49	1.74	1.55	1.68	1.39	1.67	1.39
7	2.22	1.85	2.13	1.97	2.21	1.87	2.07	1.71
8	2.73	2.29			2.79	2.45	2.58	2.16
9	3.36	2.92			3.44	3.12	3.23	2.8
10	4.02	3.59			4.07	3.78	3.88	3.46
11	4.72	4.31			4.72	4.43	4.58	4.18
12	5.5	5.09			5.39	5.11	5.36	4.97
13	6.23	5.84			6.07	5.77	6.1	5.73
14	6.79	6.38			6.13	5.79	6.68	6.27
15	7.53	7.1					7.36	6.95
16	8.26	7.83					8.1	7.67
17	8.74	8.32					8.58	8.16
18	9.6	9.16					9.48	9.05
19	10.5	10.03					10.29	9.84
20	11.02	10.51					10.6	10.09
21	10.96	10.14					10.54	9.77

Broadly speaking, results show that Random Forest and lightGBM are good in producing DTD predictors for patients with a length of stay up to 7 days, which are optimal in the sense that their root mean square error (RMSE) and their mean absolute predictive error (MAE) are always below 1 day (with the exception of the seventh day). Random Forest is also good in DTD predictions for patients with a length of stay up to 14 and up to 21 days, with RMSE and MAE values below 1 day in the last 7 days before discharge. This last week

is the most crucial in the planning at the ICUs because it gives the opportunity to know in advance, with quite small error, the date of discharge and, therefore, to properly schedule beds, personnel, and other resources, in addition to having the possibility of advancing the planning of the transfer of the patient to any other hospital unit.

The average and deviation of the RMSE and MAE for each of the 4 groups is given in Table 12, taking into account all the days of each group (blue) and also considering only the remaining days below 7 in each group (green). We can see that Random Forest gives the lowest errors. The best average of the RMSE is 1.4 with a deviation of 1.0 for the All and long groups for the Random Forest model, but it is much smaller for the short stay (mean of 0.5, stdev of 0.4) and the medium stay (mean of 1.1, stdev of 0.9). This indicates that using different prediction models for each case would lead to better results in general. Considering only the last week of stay at ICU, the predictions made are much better. The mean MAE obtained for Random Forest is between 0.4 and 1.1 with a maximum standard deviation of 0.9.

Table 12. RMSE and MAE average and standard deviation for DTD prediction for each model considering all the days of each group (blue) or only the last 6 days of stay (green).

Model	Measure	All		Short Stay		Medium Stay		Long Stay	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Random Forest	avg.	1.4	1.1	0.5	0.4	1.1	0.9	1.4	1.1
	st.dev.	1.0	0.9	0.4	0.3	0.7	0.7	1.0	0.9
	avg. 2–6	1.0	0.7	0.6	0.5	0.8	0.6	1.0	0.7
	st.dev. 2–6	0.3	0.2	0.5	0.3	0.2	0.1	0.2	0.1
LightGBM	avg.	1.8	1.5	0.5	0.3	1.3	1.0	1.8	1.5
	st.dev.	1.2	1.1	0.5	0.3	0.8	0.7	1.1	1.0
	avg. 2–6	1.2	1.0	0.6	0.4	1.0	0.8	1.2	1.0
	st.dev. 2–6	0.4	0.2	0.3	0.2	0.5	0.2	0.2	0.2
XGBoost	avg.	2.7	2.3	0.9	0.7	2.0	1.7	2.7	2.3
	st.dev.	1.9	1.7	0.7	0.4	1.3	1.1	1.8	1.6
	avg. 2–6	1.9	1.4	1.1	0.9	1.5	1.2	1.8	1.5
	st.dev. 2–6	0.4	0.3	0.5	0.4	0.3	0.3	0.3	0.2

3.5. Hybrid Model

In general, results for DTD prediction models are better than LOS prediction models when applied in the last days of stay. Differences between DTD and LOS prediction models become more evident when they are used to predict the discharge times of patients in their i -th day before discharge ($i = 2, \dots, 21$) separately. Figures 2–4 show that DTD models have better performance (with lower values in both RMSE and MAE) when the stay arrives at its end. The inflection points give us the information to establish the value of the threshold parameter α_{g_i} associated with each DTD subgroup, which are the following: 3.5 for short stays, 5 days for medium-length stays, and 6 days for longer stays.

We observed that these values correspond to the average length of stay for every subgroup (i.e., 3.51 for short stays, 5.54 for medium stays, and 6.07 for long stays).

Since DTD models have been trained not only on the patient conditions 24–48 h after ICU admission, but also on the daily clinical condition of patients along their full ICU stay, the amount of data to train the model is larger than in LOS models, resulting in better results for DTD models at the end of stay.

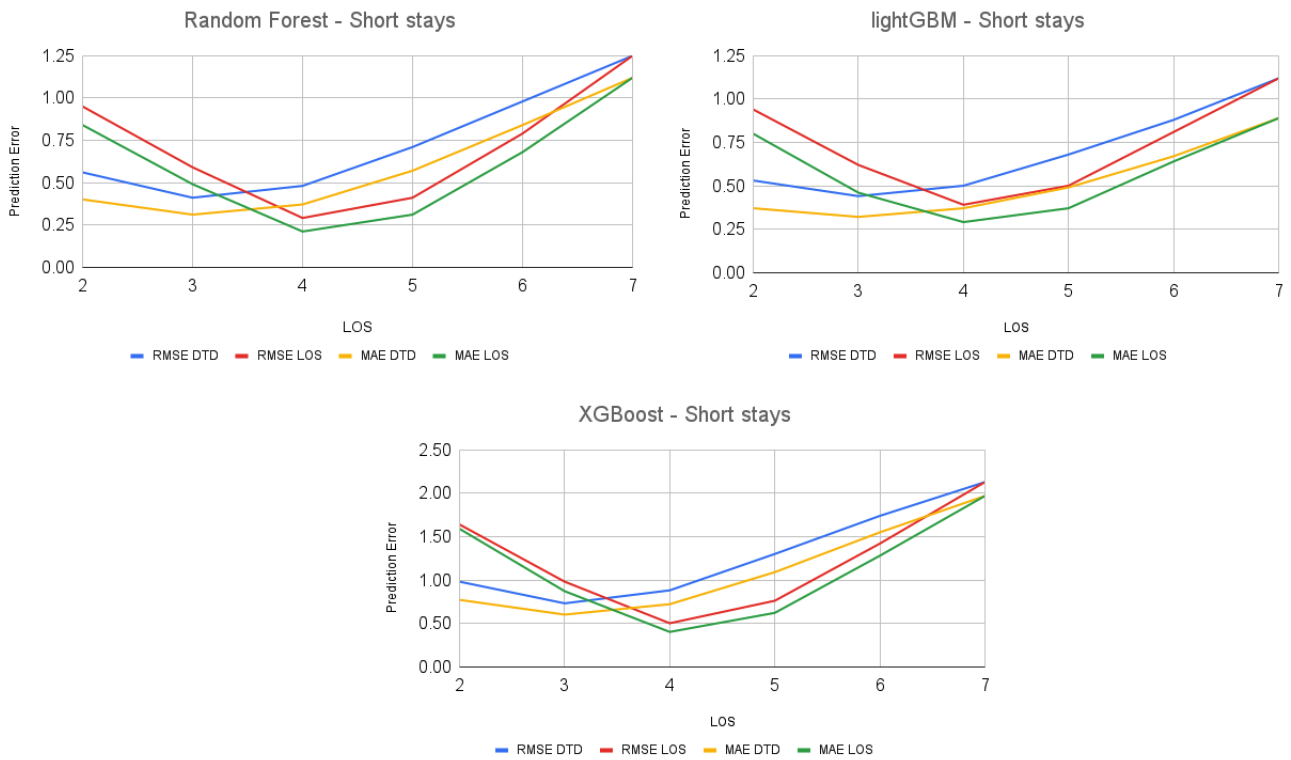


Figure 2. Prediction errors for short stays.

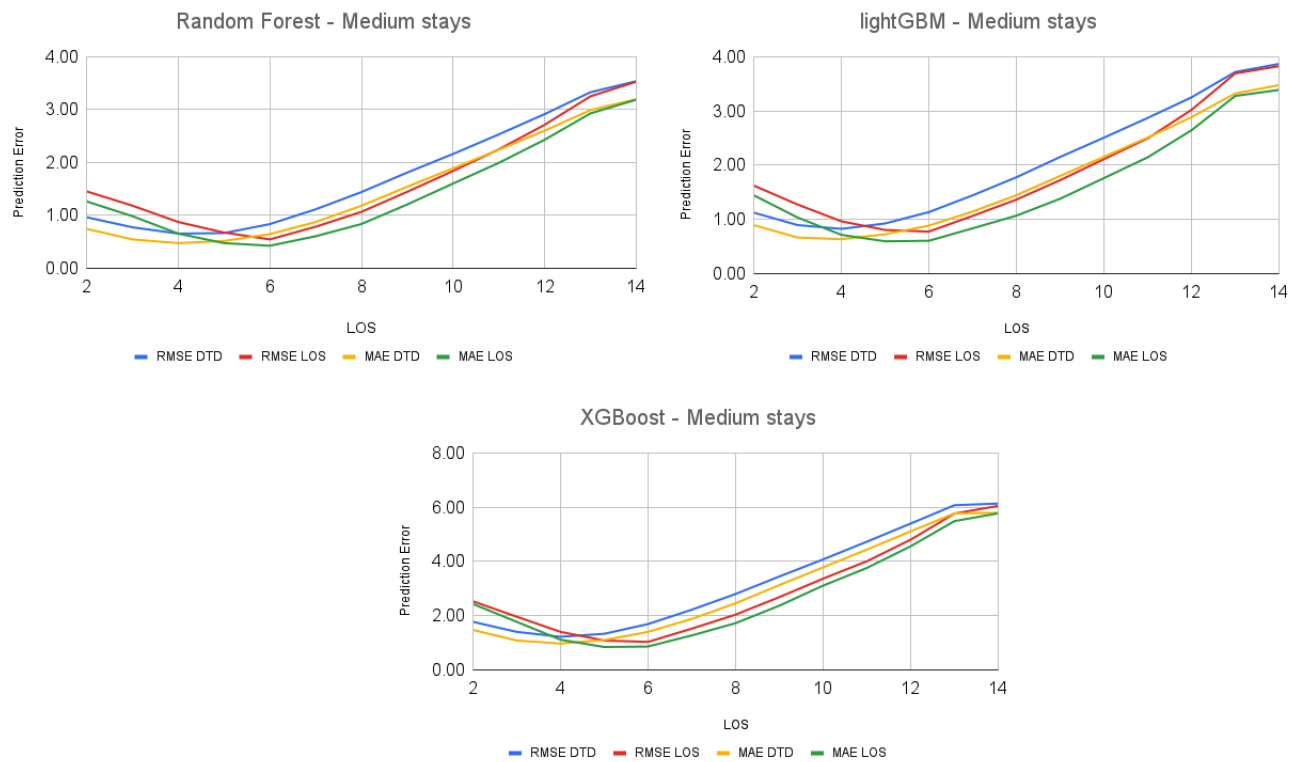


Figure 3. Prediction errors for medium stays.

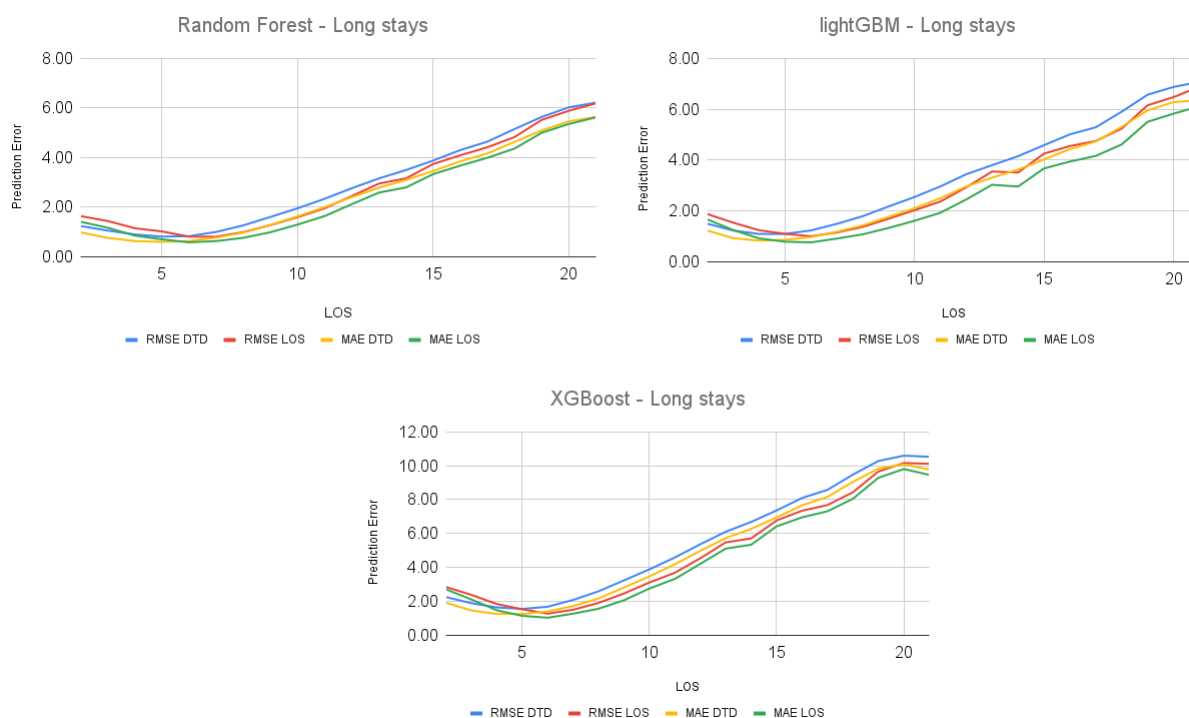


Figure 4. Prediction errors for long stays.

By combining DTD and LOS models, we can improve prediction outcomes for all subgroups. Table 13 shows the weighted average of RMSE values when applying the hybrid proposed method to each of the groups, in comparison to using a single prediction of LOS for any patient (first column), which is the current approach used at ICUs nowadays. It can be clearly seen that the results of the proposed model give a considerably lower RMSE error in all the methods, including below one day when using Random Forest.

Table 13. Average RMSE for hybrid group models, compared with RMSE of LOS for All data.

Model	LOS All	Short Stay	Medium Stay	Long Stay
Random Forest	2.00	0.50	0.52	0.88
lightGBM	2.50	0.90	1.09	1.64
XGBoost	3.60	1.21	1.40	2.22

4. Conclusions and Future Work

Predicting patient stays in the ICU is normally addressed with studies that foresee the LOS with data captured in the first 24–48 h after admission. Without undermining the importance of this type of prediction, it does not seem justified that, as the clinical condition of the patients evolve, this new information is not taken into account to dynamically predict the DTD of the corresponding patients. Statistical techniques have proven insufficient for addressing this problem. Therefore, other data analysis methods based on the construction of prediction models using machine learning, have been considered in this work.

As a first contribution, the definition and characterization of three classes based on the number of days of stay of an ICU patient (short, medium, and long) allow us to identify different subsets of data that help in the construction of more accurate prediction models.

As a second contribution, the use of dynamic DTD prediction models that exploit the daily data collected at ICUs demonstrates a clear improvement in the quality of the results.

Thirdly, the paper proposes a hybrid model whose results show that it drastically improves the results in prediction models for ICU patients. It consists of using group-specific LOS models at the beginning of the stay and DTD models at the end of the stay. Results obtained when comparing the proposed method with the current prediction of LOS

at admission date show a clear improvement in the prediction error, reducing from 2 days to less than 1 day for all the groups. Therefore, the distinction of these three groups and the construction of specific models for each of them permits us to adjust the predictions to the different patients' situations. It is worth noting that the proposed methodology is applicable to all the patients that are treated in the ICU, and it is not restricted to certain diagnoses as in other works.

The proposed models do not exploit temporal patterns, as each day of stay is treated as an independent patient. As future work, techniques that consider some temporal progression of the medical conditions could be used in order to capture those patterns and study if the prediction date can be further improved. Future work will also aim to enhance the capability of this hybrid prediction model to provide understandable insights into their decision-making processes, in order to improve explainability in clinical scenarios, in line with other recent works [39–42].

Author Contributions: Conceptualization: D.C. and D.R.; methodology: D.C. and D.R.; software: D.C.; validation: D.C., D.R. and A.V.; formal analysis: D.C., D.R. and A.V.; resources: D.C.; data curation: D.C. and D.R.; writing-original draft: D.C.; writing-review and editing: D.R. and A.V.; supervision: D.R. and A.V.; funding acquisition: A.V. All authors have read and agreed to the published version of the manuscript.

Funding: Work partially supported by URV project 2022PFR-URV-41.

Data Availability Statement: No new data were created in this study.

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

References

1. Marshall, J.; Bosco, L.; Adhikari, N.; Connolly, B.; Diaz, J.; Dorman, T.; Fowler, R.; Meyfroidt, G.; Nakagawa Pelosi, P.; Vincent, J.; et al. What is an intensive care unit? A report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine. *J. Crit. Care* **2017**, *37*, 270–276. [[CrossRef](#)] [[PubMed](#)]
2. Valentin, A.; Ferdinande, P.; ESICM Working Group on Quality Improvement. Recommendations on basic requirements for intensive care units: Structural and organizational aspects. *Intensive Care Med.* **2011**, *37*, 1575–1587. [[CrossRef](#)] [[PubMed](#)]
3. Kross, E.; Engelberg, R.; Downey, L.; Cuschieri, J.; Hallman, M.; Longstreth, W.; Tirschwell, D.; Curtis, J. Differences in End-of-Life Care in the ICU Across Patients Cared for by Medicine, Surgery, Neurology, and Neurosurgery Physicians. *Chest* **2014**, *145*, 313–321. [[CrossRef](#)] [[PubMed](#)]
4. Jacobs, K.; Roman, E.; Lambert, J.; Moke, L.; Scheys, L.; Kesteloot, K.; Roodhooft, F.; Cardoen, B. Variability drivers of treatment costs in hospitals: A systematic review. *Health Policy* **2022**, *126*, 75–86. [[CrossRef](#)] [[PubMed](#)]
5. Rossi, C.; Simini, B.; Brazzi, L.; Rossi, G.; Radrizzani, D.; Iapichino, G.; Bertolini, G. Variable costs of ICU patients: A multicenter prospective study. *Intensive Care Med.* **2006**, *32*, 545–552. [[CrossRef](#)]
6. Bai, J.; Fügner, A.; Schoenfelder, J.; Brunner, J. Operations research in intensive care unit management: A literature review. *Health Care Manag. Sci.* **2018**, *21*, 1–24. [[CrossRef](#)]
7. McKenzie, M.; Auriemma, C.; Olenik, J.; Cooney, E.; Gabler, N.; Halpern, S. An Observational Study of Decision Making by Medical Intensivists. *Crit. Care Med.* **2015**, *43*, 1660–1668. [[CrossRef](#)]
8. Verburg, I.; Atashi, A.; Eslami, S.; Holman, R.; Abu-Hanna, A.; de Jonge, E.; Peek, N.; de Keizer, N.F. Which Models Can I Use to Predict Adult ICU Length of Stay? A Systematic Review. *Crit. Care Med.* **2017**, *45*, 222–231. [[CrossRef](#)]
9. Kramer, A. Are ICU Length of Stay Predictions Worthwhile? *Crit. Care Med.* **2017**, *45*, 379–380. [[CrossRef](#)]
10. Hachesu, P.; Ahmadi, M.; Alizadeh, S.; Sadoughi, F. Use of data mining techniques to determine and predict length of stay of cardiac patients. *Healthc. Inform. Res.* **2013**, *19*, 121–129. [[CrossRef](#)]
11. Mollaei, N.; Londral, A.; Cepeda, C.; Azevedo, S.; Santos, J.P.; Coelho, P.; Fragata, J.; Gamboa, H. Length of Stay Prediction in Acute Intensive Care Unit in Cardiothoracic Surgery Patients. In Proceedings of the 2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 25–27 March 2021; pp. 1–5.
12. Gholipour, C.; Rahim, F.; Fakhree, A.; Ziapour, B. Using an Artificial Neural Networks (ANNs) Model for Prediction of Intensive Care Unit (ICU) Outcome and Length of Stay at Hospital in Traumatic Patients. *J. Clin. Diagn.* **2015**, *9*, 19–23. [[CrossRef](#)] [[PubMed](#)]
13. Muhlestein, W.; Akagi, D.; Davies, J.; Chambless, L.B. Predicting Inpatient Length of Stay After Brain Tumor Surgery: Developing Machine Learning Ensembles to Improve Predictive Performance. *Neurosurgery* **2019**, *85*, 384–393. [[CrossRef](#)]
14. Su, L.; Xu, Z.; Chang, F.; Ma, Y.; Liu, S.; Jiang, H.; Wang, H.; Li, D.; Chen, H.; Zhou, X.; et al. Early Prediction of Mortality, Severity, and Length of Stay in the Intensive Care Unit of Sepsis Patients Based on Sepsis 3.0 by Machine Learning Models. *Front. Med.* **2021**, *8*, 664966. [[CrossRef](#)] [[PubMed](#)]

15. Rowan, M.; Ryan, T.; Hegarty, F.; O'Hare, N. The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors. *Artif. Intell. Med.* **2007**, *40*, 211–221. [[CrossRef](#)] [[PubMed](#)]
16. Van Houdenhoven, M.; Nguyen, D.; Eijkemans, M.; Steyerberg, E.W.; Tilanus, H.W.; Gommers, D.; Wullink, G.; Bakker, J.; Kazemier, G. Optimizing intensive care capacity using individual length-of-stay prediction models. *Crit. Care* **2007**, *11*, 42. [[CrossRef](#)] [[PubMed](#)]
17. Jayamini, W.; Mirza, F.; Naeem, M.; Chan, A. State of Asthma-Related Hospital Admissions in New Zealand and Predicting Length of Stay Using Machine Learning. *Appl. Sci.* **2022**, *12*, 9890. [[CrossRef](#)]
18. Alghatani, K.; Ammar, N.; Rezugui, A.; Shaban-Nejad, A. Predicting Intensive Care Unit Length of Stay and Mortality Using Patient Vital Signs: Machine Learning Model Development and Validation. *JMIR Med. Inform.* **2021**, *9*, e21347. [[CrossRef](#)]
19. Nassar, J.; Caruso, P. ICU physicians are unable to accurately predict length of stay at admission: A prospective study. *Int. J. Qual. Health Care* **2015**, *28*, 99–103. [[CrossRef](#)]
20. Verburg, I.; Keizer, N.; Jonge, E.; Peek, N. Comparison of Regression Methods for Modeling Intensive Care Length of Stay. *PLoS ONE* **2014**, *9*, e109684. [[CrossRef](#)]
21. Li, C.; Chen, L.; Feng, J.; Wu, D.; Wang, Z.; Liu, J.; Xu, W. Prediction of Length of Stay on the Intensive Care Unit Based on Least Absolute Shrinkage and Selection Operator. *IEEE Access* **2019**, *7*, 110710–110721. [[CrossRef](#)]
22. Huang, Z.; Juarez, J.; Duan, H.; Li, H. Length of stay prediction for clinical treatment process using temporal similarit. *Expert Syst. Appl.* **2013**, *40*, 6330–6339. [[CrossRef](#)]
23. Moran, J.; Solomon, P. A review of statistical estimators for risk-adjusted length of stay: Analysis of the Australian and new Zealand intensive care adult patient data-base, 2008–2009. *BMC Med. Res. Methodol.* **2012**, *12*, 68. [[CrossRef](#)] [[PubMed](#)]
24. Abd-ElrazekaAhmed, M.; Eltahawi, A.; Elaziz, M.; Abd-Elwhab, M. Predicting length of stay in hospitals intensive care unit using general admission features. *Ain Shams Eng. J.* **2021**, *12*, 3691–3702. [[CrossRef](#)]
25. Chrusciel, J.; Girardon, F.; Roquette, L.; Laplanche, D.; Duclos, A.; Sanchez, S. The prediction of hospital length of stay using unstructured data. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 351. [[CrossRef](#)]
26. Caetano, N.; Laureano, R.; Cortez, P. A Data-driven Approach to Predict Hospital Length of Stay—A Portuguese Case Study. In Proceedings of the 16th International Conference on Enterprise Information Systems, Lisbon, Portugal, 27–30 April 2014; Volume 3, pp. 407–414.
27. Houthoofd, R.; Ruyssinck, J.; Herten, J.; Stijven, S.; Couckuyt, I.; Gadeyne, B.; Ongenaes, F.; Colpaert, K.; Decruyenaere, J.; Dhaene, T.; et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artif. Intell. Med.* **2015**, *63*, 191–207. [[CrossRef](#)] [[PubMed](#)]
28. Ma, X.; Si, Y.; Wang, Z.; Wang, Y. Length of stay prediction for ICU patients using individualized single classification algorithm. *Comput. Methods Programs Biomed.* **2020**, *186*, 105224. [[CrossRef](#)] [[PubMed](#)]
29. Wu, J.; Lin, Y.; Li, P.; Hu, Y.; Zhang, L.; Kong, G. Predicting Prolonged Length of ICU Stay through Machine Learning. *Diagnostics* **2021**, *11*, 2242. [[CrossRef](#)]
30. Ayyoubzadeh, S. A study of factors related to patients' length of stay using data mining techniques in a general hospital in southern Iran. *Health Inf. Sci. Syst.* **2020**, *8*, 9. [[CrossRef](#)]
31. Cuadrado, D.; Riaño, D.; Gómez, J.; Rodríguez, A.; Bodí, M. Methods and measures to quantify ICU patient heterogeneity. *J. Biomed. Inform.* **2021**, *117*, 103768. [[CrossRef](#)]
32. Cuadrado, D.; Riaño, D. ICU Days-to-Discharge Analysis with Machine Learning Technology. In *Artificial Intelligence in Medicine, Proceedings of the 19th International Conference on Artificial Intelligence in Medicine, AIME 2021, Virtual, 15–18 June 2021*; Springer: Cham, Switzerland, 2021; pp. 103–113.
33. Pollard, T.J.; Johnson, A.E.W.; Raffa, J.D.; Celi, L.A.; Mark, R.G.; Badawi, O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci. Data* **2018**, *5*, 180178. [[CrossRef](#)]
34. Sevilla, B.; Gibert, K.; Sánchez-Marrè, M. Using CVI for Understanding Class Topology in Unsupervised Scenarios. In *Advances in Artificial Intelligence, Proceedings of the 17th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2016, Salamanca, Spain, 14–16 September 2016*; Luaces, O., Gámez, J.A., Barrenechea, E., Troncoso, A., Galar, M., Quintián, H., Corchado, E., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; pp. 135–149.
35. Sevilla-Villanueva, B. A Methodology for Pre-Post Intervention Studies: An Application for a Nutritional Case Study. Ph.D. Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, 2016.
36. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
37. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
38. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149–3157.
39. Moreno-Sánchez, P.A. Improvement of a prediction model for heart failure survival through explainable artificial intelligence. *Front. Cardiovasc. Med.* **2023**, *10*, 1219586. [[CrossRef](#)] [[PubMed](#)]
40. de Moura, L.V.; Mattjie, C.; Dartora, C.M.; Barros, R.C.; Marques da Silva, A.M. Explainable Machine Learning for COVID-19 Pneumonia Classification With Texture-Based Features Extraction in Chest Radiography. *Front. Digit. Health* **2022**, *3*, 662343. [[CrossRef](#)]

41. Stenwig, E.; Salvi, G.; Rossi, P.S.; Skjærvold, N.K. Comparative analysis of explainable machine learning prediction models for hospital mortality. *BMC Med. Res. Methodol.* **2022**, *22*, 53. [[CrossRef](#)] [[PubMed](#)]
42. Du, Y.; Rafferty, A.R.; McAuliffe, F.M.; Wei, L.; Mooney, C. An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. *Sci. Rep.* **2022**, *12*, 1170. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.