



Article

Global and Local Interpretable Machine Learning Allow Early Prediction of Unscheduled Hospital Readmission

Rafael Ruiz de San Martín ¹, Catalina Morales-Hernández ², Carmen Barberá ², Carlos Martínez-Cortés ³, Antonio Jesús Banegas-Luna ³ , Francisco José Segura-Méndez ⁴ , Horacio Pérez-Sánchez ³ , Isabel Morales-Moreno ² and Juan José Hernández-Morante ^{2,*}

¹ Servicio Murciano de Salud, Hospital Universitario Virgen de la Arrixaca, Crta. El Palmar, 30120 Murcia, Spain; rafael.ruiz2@carm.es

² Faculty of Nursing, Universidad Católica de Murcia (UCAM), Avd. de los Jerónimos, 30107 Murcia, Spain; cmorales@ucam.edu (C.M.-H.); mcbarbera@ucam.edu (C.B.); imorales@ucam.edu (I.M.-M.)

³ Structural Bioinformatics and High Performance Computing (BIO-HPC), Universidad Católica de Murcia (UCAM), Avd. de los Jerónimos, 30107 Murcia, Spain; cmartinez1@ucam.edu (C.M.-C.); ajbanegas@ucam.edu (A.J.B.-L.)

⁴ Hydrological Modeling and Research Lab, Universidad Católica de Murcia (UCAM), Avd. de los Jerónimos, 30107 Murcia, Spain; fsegura@ucam.edu

* Correspondence: jjhernandez@ucam.edu; Tel.: +34-968-278-809

Abstract: Nowadays, most of the health expenditure is due to chronic patients who are readmitted several times for their pathologies. Personalized prevention strategies could be developed to improve the management of these patients. The aim of the present work was to develop local predictive models using interpretable machine learning techniques to early identify individual unscheduled hospital readmissions. To do this, a retrospective, case-control study, based on information regarding patient readmission in 2018–2019, was conducted. After curation of the initial dataset ($n = 76,210$), the final number of participants was $n = 29,026$. A machine learning analysis was performed following several algorithms using unscheduled hospital readmissions as dependent variable. Local model-agnostic interpretability methods were also performed. We observed a 13% rate of unscheduled hospital readmissions cases. There were statistically significant differences regarding age and days of stay ($p < 0.001$ in both cases). A logistic regression model revealed chronic therapy (odds ratio: 3.75), diabetes mellitus history (odds ratio: 1.14), and days of stay (odds ratio: 1.02) as relevant factors. Machine learning algorithms yielded better results regarding sensitivity and other metrics. Following, this procedure, days of stay and age were the most important factors to predict unscheduled hospital readmissions. Interestingly, other variables like allergies and adverse drug reaction antecedents were relevant. Individualized prediction models also revealed a high sensitivity. In conclusion, our study identified significant factors influencing unscheduled hospital readmissions, emphasizing the impact of age and length of stay. We introduced a personalized risk model for predicting hospital readmissions with notable accuracy. Future research should include more clinical variables to refine this model further.

Keywords: hospital readmission; chronic patient; machine learning; prediction; shap-values



Citation: Ruiz de San Martín, R.; Morales-Hernández, C.; Barberá, C.; Martínez-Cortés, C.; Banegas-Luna, A.J.; Segura-Méndez, F.J.; Pérez-Sánchez, H.; Morales-Moreno, I.; Hernández-Morante, J.J. Global and Local Interpretable Machine Learning Allow Early Prediction of Unscheduled Hospital Readmission. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1653–1666. <https://doi.org/10.3390/make6030080>

Academic Editors: Quynh C. Nguyen and Thu T. Nguyen

Received: 28 June 2024

Revised: 11 July 2024

Accepted: 16 July 2024

Published: 17 July 2024



Copyright: © 2024 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

Patients with higher comorbidity levels experience a greater number of treatment complications, longer clinical stays, heightened surgical risks, predictable readmissions, and a substantial consumption of healthcare resources and services [1]. Therefore, prediction of these kind of complications could significantly improve their health care. Nowadays, there is a wealth of data on patients admitted to hospitals that may help. Unfortunately, hospital departments, particularly emergency and surgical departments, are sometimes overwhelmed, leading to an impaired quality of care due to excessive demand. To improve

the management of these services, it would be highly beneficial to predict which individuals will require these services again.

Currently, several tools have been developed to assess the probability of hospital readmission, with the Charlson Comorbidity Index [2,3] being the most extensively studied. According to the methodology used in the National Health System for analysing hospitalization, unexpected hospital readmission (UHR) refers to unexpected admissions (emergency admissions) after a previous discharge from the same hospital [4].

Readmissions are influenced by patient comorbidity and the morbidity treated at the hospital. They can indicate clinical stability in the course of the disease, where readmission may be prompted by complications after discharge, reflecting inadequate patient follow-up. However, readmissions can also be related to patient clinical stability, where early discharge from the hospital may lead to a readmission [5]. For these reasons, the Minimum Basic Data Set Register studies classify hospital readmissions into two periods: those occurring during the first week after discharge and between the second week and the 30th day after discharge, on which this work focuses. Since the first week will be associated with bad clinical care, there will be many factors implied in unexpected hospital readmission in the second period.

Data analysis has recently met significant challenges due to insufficient statistical power and the lack of advanced computing methods, limiting the effective analysis of large datasets and accurate predictive modelling [6]. However, the development of modern computational technologies allows for the use of diverse algorithms and machine learning techniques aided by Artificial Intelligence methods, enabling the processing of complex data with greater efficiency and scalability [7]. This evolution in computing not only enhances the depth of analysis but also supports the development of more accurate predictive models by using advanced machine learning procedures. The healthcare sector has been working for years to include machine learning procedures in data management and decision-making [8]. This sector is being transformed rapidly thanks to machine learning, which makes it possible to predict, treat, and diagnose illnesses with better accuracy and reliability [9,10].

Considering the above comments, the main objective of this study was to assess the effectiveness and efficiency of several machine learning algorithms in predicting unexpected hospital readmission for a University Clinic Hospital. Additionally, this study aimed to identify the characteristics of the population at higher risk of hospital readmission through the emergency department to propose individualized risk criteria for personalized prediction models.

2. Materials and Methods

2.1. Design

This was a retrospective, case-control study, based on information regarding patient readmission in 2018–2019. First, a data collection template was developed based on previous studies related to unscheduled hospital readmissions (unexpected hospital readmission). After obtaining approval from the Clinical Committee of the Hospital (date 30 April 2019) and those legally responsible for the centre, a process of anonymisation of the data, which allowed to collect the variables required for the study excluding personal information, was performed. In a subsequent phase, a preliminary dataset was received. Data curation and statistical procedures are described below. All information obtained was treated with strict confidentiality, in accordance with the Spanish Organic Law on Data Protection (Organic Law 15/1999) and other relevant legislation in force.

2.2. Data Collection

The population selected for this study was focused on all persons who have been attended by the emergency department of the University Clinic Hospital of Murcia (Spain) and who have been admitted to the hospital through this department, which may be considered as a readmission or non-readmission. The data recorded by the emergency department during 2018 and 2019 were used for this study. Data for the years 2020–2022

were not included due to COVID-19, as it could represent a deviation from the current real situation.

From the collected dataset, all records belonging to the maternity and children's hospital, as well as all those related to obstetrics, gynaecology, paediatrics, neonatology, schoolchildren, etc., services were excluded. Children under 14 years of age were also excluded for this study. The same applies to some adolescents under 18 years of age who usually receive medical care in special hospital units, for example, oncology and child psychiatry.

The dataset included information regarding sociodemographic data, like sex, age, and ZIP code, as well as other information regarding admission and discharge services. The patient conditions were recorded by two different systems, the International Classification of Diseases (ICD-10) and the Clinical Classification Software (CCS). Several variables of the clinical history were also collected. Due to the high heterogeneity in the clinical records, we were able to obtain reliable information only for hypertension, diabetes mellitus, dyslipidaemia, and allergies history. Enolic and smoking habits were also recorded, and chronic treatment, as well as any adverse drug reaction, were also recorded. Finally, exploratory tests (radiography, ultrasound, or tomography) were also included in the data set.

Initially, 76,210 patients were admitted in the hospital during the study period and comprised the initial dataset. Two patients were excluded due to unknown sex. The lack of information regarding admission service made us withdraw 883 patients. In this line, the lack of information of the discharge service occurred in 424 patients. The lack of other personal information (age or ZIP code) also excluded 51 individuals. As the present study focused on those patients admitted from the emergencies service, after selecting these patients, we obtained a revised dataset of $n = 50,729$ patients. The last steps to debug the dataset were performed to exclude those patients with an age lower than 18 years ($n = 8176$) and those derived from maternal services ($n = 13,527$). Finally, the curated dataset that represents the data showed in this work included 29,026 individuals. Baseline information about admission and discharge services is shown in Supplementary Table S1.

2.3. Outcome Variable

The primary outcome was unexpected hospital readmission between the second week and the 30th day after discharge, which was computed based on patients' record for 2018–2019. The information included only patients admitted directly from the emergencies service.

2.4. Statistical Analysis

Descriptive statistics for the sociodemographic and the other predictive variables were calculated. Chi-squared tests were performed to test the association between the unexpected hospital readmission and sex and the other categorical features. An unpaired *t*-test was performed to analyse possible differences on age and days of stay attending to the presence of an UHR or not. A multivariable logistic regression model was fitted to examine the factors associated with UHR. Considering the high number of variables, to avoid collinearity, this procedure was carried out in three steps. In a first model, the association between UHR and the admission and discharge services was analysed. In a second model, the influence of the patient's disease was evaluated. In the third model, the other clinical history data were evaluated. Finally, a combined model was determined by backward steps with those variables with a significant influence on the previous models. Adjusted odds ratios and 95% confidence intervals were reported. All data analyses were performed using SPSS software, version 27.0 (IBM SPSS software, Armonk, NY, USA). Graphs and figures were produced with GraphPad Prism 9.0 (Graphpad Software, Boston, MA, USA) Probability maps were produced with Qgis 3.22 (<https://qgis.org/>, accessed on 7 July 2024).

2.5. Machine Learning Procedure

2.5.1. Data Cleaning and Pre-Processing

The data pre-processing stage is described in Figure 1. Initially, the dataset was organized into a target variable (UHR) and various input features. This study was conducted using Python library scikit-learn and RStudio (<https://posit.co/download/rstudio-desktop/>, accessed on 7 July 2024). For nominal variables with no inherent order, such as the Admission Service and Major Diagnostic Categories, we utilized OneHotEncoder, converting them into multiple binary features. Ordinal variables like triage level were encoded using LabelEncoder, maintaining their natural order which could be informative for the predictive model. Numerical data underwent standardization via StandardScaler. Standardization ensured that each feature contributes equally to the model's predictions, thereby enhancing overall performance.

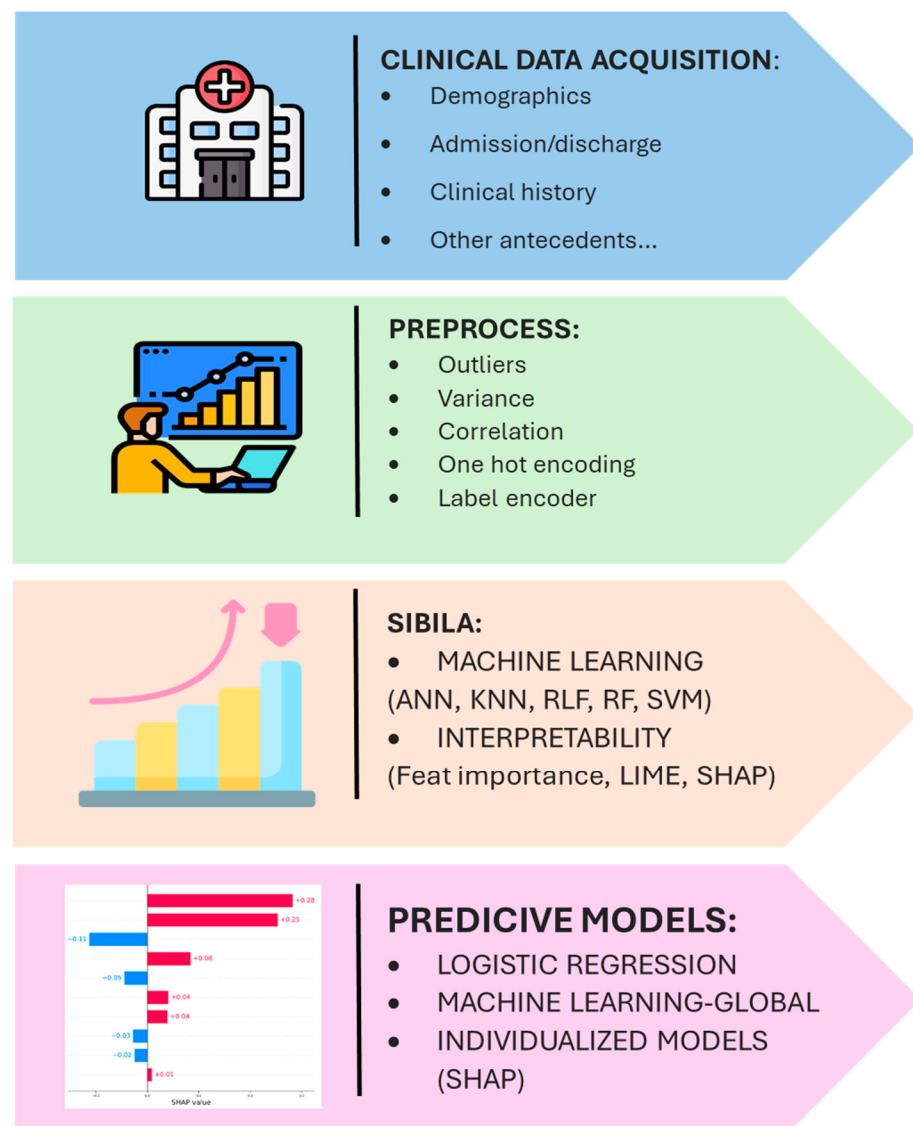


Figure 1. Overview of the study design.

The dataset, consisting of 29,026 records post-curation, was divided into training (80%) and validation (20%) subsets using `.model_selection`. This allocation, resulting in 23,205 training and 5821 validation records, was chosen to balance model training with adequate validation assessment while considering computational efficiency.

Anonymization of patient data was a critical step in our methodology, conducted to protect patient privacy while maintaining data integrity. Techniques employed for anonymization included the removal of direct identifiers and the aggregation of potentially identifiable information.

2.5.2. Machine Learning Analysis

In the quantitative exploration of the dataset, this study operationalized the SIBILA computational framework, a computational tool designed for leveraging High-Performance Computing architectures in the construction and assessment of various advanced machine learning models (Available at <https://github.com/bio-hpc/sibila>, accessed on 5 July 2024). This framework facilitates the deployment of Artificial Neural Networks, Random Forests, and eXtreme Gradient Boosting machines. These models were selected for their proven efficacy in analysing complex, multidimensional datasets prevalent in healthcare research.

A distinctive feature of the SIBILA platform is the integration of model interpretability techniques, including feature importance analysis, Local Interpretable Model-agnostic Explanations, and SHapley Additive exPlanations. These interpretability techniques are paramount for deciphering the operational mechanisms of the employed models. This is essential for the elucidation of both the global influence and the local impact of features within the predictive models, providing an exhaustive understanding of the determinants influencing hospital readmission outcomes.

Furthermore, SIBILA generates comprehensive performance reports for each algorithm, with a particular emphasis on the AUC of the ROC curve. In addition, other metrics like sensitivity, accuracy, etc., were also evaluated. The interpretability outputs provided by SIBILA are instrumental in providing empirical evidence for clinical application.

3. Results

3.1. Preliminary Analysis

Finally, after curating the initial dataset, the number of patients that were included in the present study was 29,026 individuals. An initial assessment of the unscheduled hospital readmissions (UHR) for the years 2018 and 2019 revealed an incidence of 14.3% ($n = 4145$) in the population under the study. The age of those patients that were readmitted was statistically significantly higher than those not readmitted (64.1 vs. 62.9, $p < 0.001$) (Figure 2a). As age increased, the incidence of UHR also increased, although it is interesting to highlight that in those people who were aged 80 or more years, the incidence of UHR was lower than in those of 70–79 years (Figure 2b). Men tended to show a higher incidence of unscheduled hospital readmissions than women, although in this case, the differences did not reach the significance level ($p = 0.139$) (Figure 2c). The hospital stay was also longer in the UHR group.

The hospital service with the higher number of UHR cases was internal medicine (Supplementary Figure S1 and Table S1). This situation was also observed regarding discharge service. Of note, several services like pneumology, maxillofacial surgery, and rheumatology showed less than 5% of UHR. Attending to the ICD-10, the factors influencing health status (ICD section XXI) and the diseases of the blood (ICD section III) showed the higher UHR incidence. Following the CCS, those admitted with endocrine disorders showed the higher UHR incidence.

Figure S2 shows the incidence of UHR attending to the date and time of admission and discharge. Wednesday and Sunday were the days with higher UHR. The UHR incidence was quite similar along the different months, although during summer and December, an increase on UHR incidence was observed. When the admission was produced from 07:00 to 10:00 h, the incidence was higher, while the discharge time with higher UHR incidence was observed at 05:00 h. As expected, those with worse clinical history tended to be readmitted in higher proportion (Figure S3). Smoking and drinking habits were also related with the unscheduled hospital readmission. Interestingly, former drinker and smoker patients

showed higher UHR than current drinker/smoker and never drinker/smoker patients ($p < 0.001$ in all cases) (Figure S3).

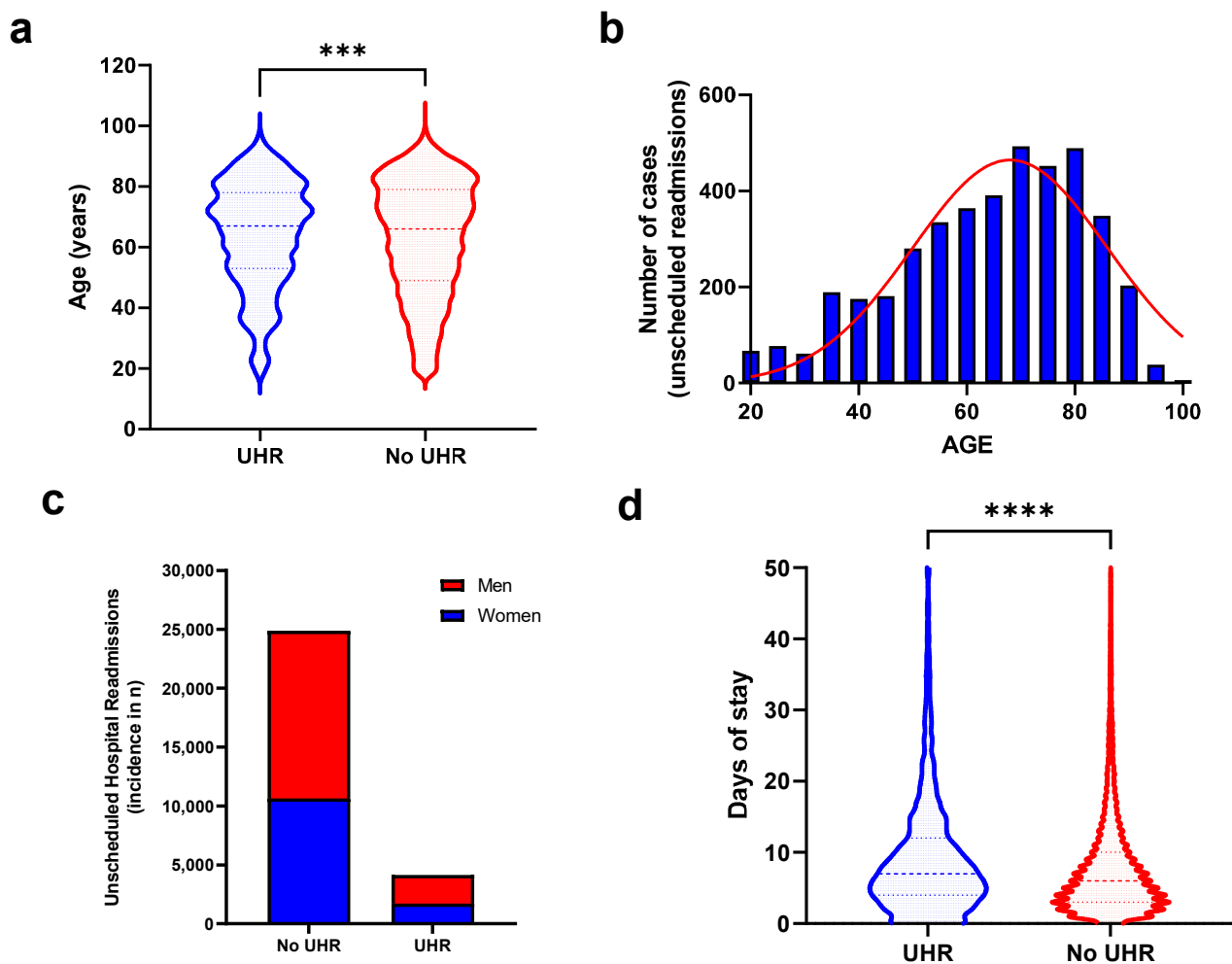


Figure 2. Characteristics of the population to the presence or not of unscheduled hospital readmission. (a) shows the distribution of ages depending on whether they were readmitted or not. The distribution of readmissions by age is shown in (b). The association with sex is also shown in (c). The distribution of patients' days of stay based on readmission is also shown in (d). Those patients that were readmitted were older (t -test $p < 0.001$) and stayed longer (t -test $p < 0.0001$).

Finally, we evaluated the influence of the patient's location on the probability of UHR. Interestingly, we observed several locations with UHR probabilities higher than 0.8, while in other regions, the probability was lower than 0.2, which indicates great disparities among regions in the same city (Figure 3).

3.2. Prediction Models of Unscheduled Hospital Readmission

To be able to predict the unscheduled hospital readmission, a logistic regression model was developed using unscheduled hospital readmission as dependent variable. First, a bivariate analysis was conducted to exclude variables without statistical significance regarding UHR (Supplementary Table S2). Consequently, we developed three logistic models to better identify those features involved in the prediction model (detailed information is shown in Table S3). To develop the final model, a stepwise multivariable logistic regression including only those variables with statistically significant effect on the previous models was performed. The final model (Table 1) revealed that chronic therapy was the feature with higher odds ratio (OR), in the sense that those patients undergoing a chronic therapy had a 3.75 increased risk of UHR. Those patients admitted from digestive surgery and

haematology also showed increased risk of UHR (OR: 2.99 and 2.27, respectively). The length of stay was also included in the model, since every day a patient spends in the hospital, their risk of being readmitted increases by 2%.

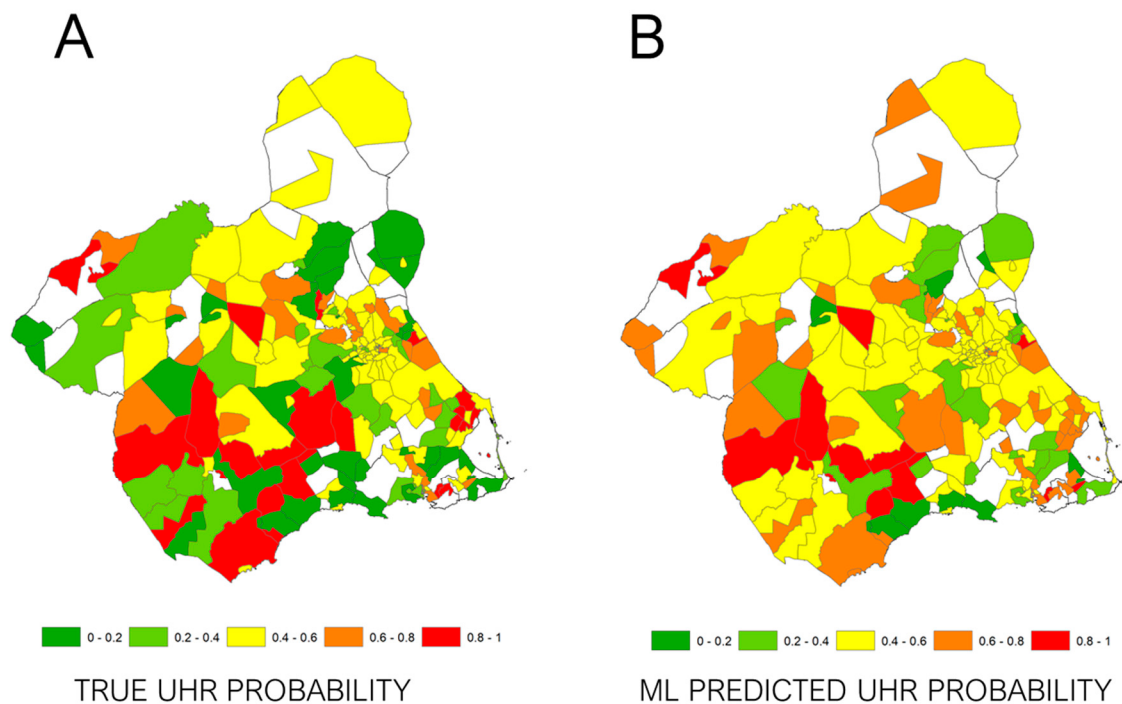


Figure 3. Observed (A) and machine learning (B) unscheduled hospital readmission (UHR) probability attending to the ZIP code. Colour scale represents probability of unscheduled hospital readmission, with green colour showing the lowest probability and red colours showing the highest.

Table 1. Multivariable logistic regression model to predict the unscheduled hospital readmission. The detailed procedure is described in Supplementary Table S3.

	B	S.E. (B)	p.	Odd Ratio	Lower CI95%	Upper CI95%
Chronic therapy	1.322	0.485	0.006	3.752	1.451	9.699
Diabetes mellitus	0.127	0.037	0.001	1.135	1.055	1.221
Days of stay	0.020	0.004	0.000	1.020	1.012	1.029
CCS-16	−1.457	0.530	0.006	0.233	0.082	0.658
CCS-17	−0.946	0.369	0.010	0.388	0.188	0.801
Adm:Cardiology	−0.697	0.238	0.003	0.498	0.312	0.793
Adm:Digestive Surgery	1.096	0.396	0.006	2.991	1.376	6.500
Adm:Haematology	0.819	0.412	0.047	2.267	1.011	5.083
Adm:Oncology	0.456	0.197	0.021	1.578	1.073	2.322
Adm:Internal Medicine	−0.257	0.127	0.042	0.773	0.603	0.991
Dis:Digestive Surgery	−0.777	0.377	0.039	0.460	0.220	0.963
Dis:Pneumology	−1.143	0.406	0.005	0.319	0.144	0.706

B: standard coefficient. S.E.: standard error. CI: confidence interval. Adm: admission service. Dis: discharge service. CCS: clinical classification software diagnosis: 16: injuries and poisoning. 17: symptoms; signs, and ill-defined conditions and factors influencing health status.

Regarding diagnostic performance, the data revealed that the model was able to properly diagnose 80.2% of UHR cases. The indicators showed high sensitivity (0.81), although the specificity was low (0.42). This situation motivated the use of a machine learning framework to integrate features into a predictive model of UHR.

To evaluate the validity of all the machine learning algorithms in the test set ($n = 23,205$), the receiver operating characteristic (ROC) curve and Area Under the Curve (AUC) of the models are shown in Supplementary Figure S4. The other evaluation metrics, including the accuracy, specificity, precision, and recall for the models, are presented in Figure S4. The findings demonstrated that the Random Forest (RF) algorithm performed best among the prediction models, with an Area Under the Curve of 0.986. The corresponding accuracy was 94% and the specificity was 91%. Therefore, the Random Forest model was selected as the best predictive model.

The features were ranked by their relative importance to UHR prediction according to the SHAP values of the model predictions. The days of stay were the most important feature, which was also included in the logistic model, clearly indicating that longer stays were associated with higher UHR probability. The next feature, in order of relevance was age; however, at this time, the sense of the association was not as clear as with the previous feature. In the ML model, history of allergies showed a monotonic relationship with UHR. Interestingly, history of adverse drug reactions and ZIP code were also important features to predict UHR. The relevance and sense of the other features are shown in Figure 4.

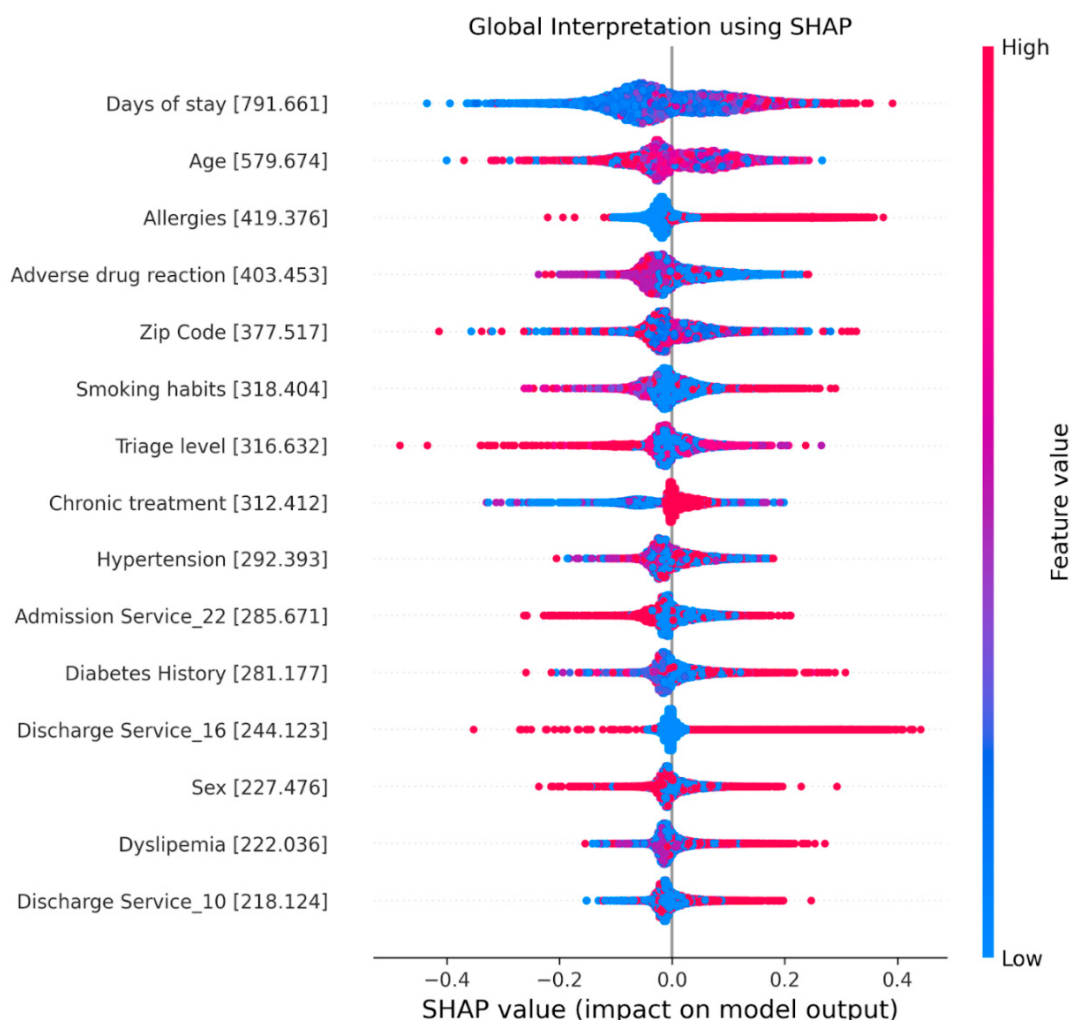


Figure 4. The impact of the input features on unscheduled hospital readmissions predictions. In the figure, each dot represents the effect of a feature on the prediction for one patient. The redder the colour of the dots is, the higher the value of the features is, and the bluer the colour of the dots is, the lower the value of the features is. Dots to the left x -axis represent patients with values of the features decreasing unexpected hospital readmission prediction, and dots to the right x -axis represent patients with values of the features increasing unexpected hospital readmission prediction.

Based on the higher accuracy and sensibility of the Random Forest model, we proposed a personalized risk factor analysis tool for explaining the UHR prediction for a specific individual. We showed the application of the personalized risk factor analysis method in one patient with UHR (Figure 5a) and one patient that was not readmitted (Figure 5b). In both cases, the probability was 0.999. In the individual prediction of UHR, history of allergies and dyslipidaemia were the most influential features, indicating an increasing impact on the probability of UHR. Days of stay and chronic treatment also were associated with UHR risk in this patient. However, in this individual, an age of 92 years had an inverse impact on the probability of UHR, which highlights the relevance of developing local interpretable explanations at individual-level. In the opposite direction (Figure 5b), not having chronic treatment and not having hypertension were determining a lower risk of UHR.

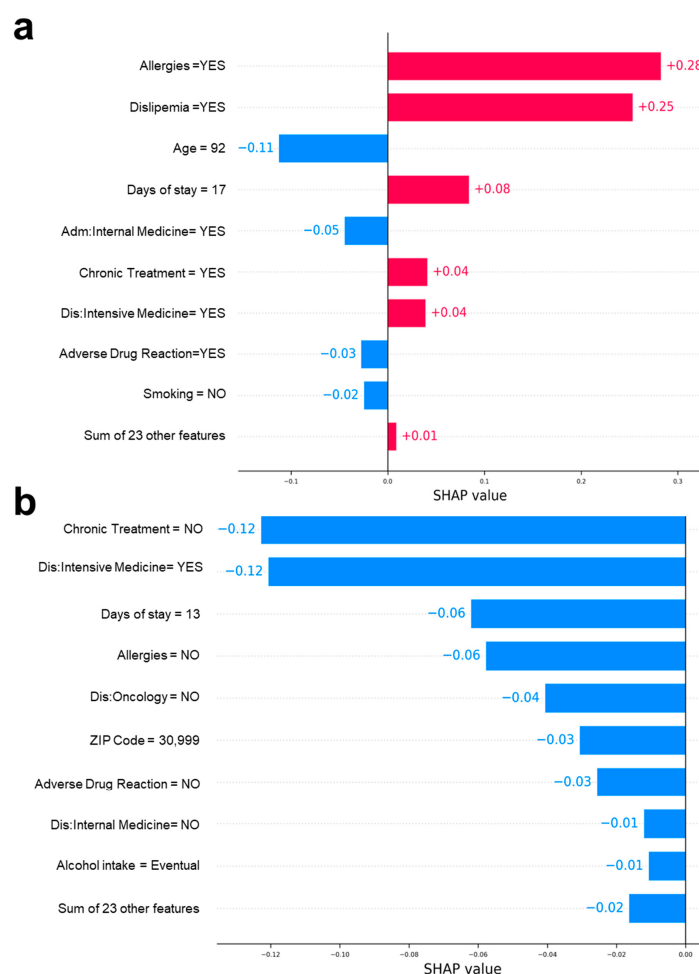


Figure 5. Examples of individualized unscheduled hospital readmissions (UHR) prediction. (a) is an example of personalized risk factor analysis for a patient in the test set identified as UHR. (b) is an example of personalized risk factor analysis for an individual identified as no-UHR.

4. Discussion

The implementation of machine learning (ML) models in healthcare settings has achieved substantial interest in recent years, offering promising advances for improving patient outcomes and optimizing resources utilization [8]. In the present study, we attempted to address the critical issue of hospital readmissions through the development and validation of an individualized interpretable ML model. This model aimed to predict the likelihood of readmission for 30 days after discharge. The present model will facilitate targeted interventions to reduce readmission rates.

A first evaluation of UHR data revealed an incidence of 14.3% in the studied population. This result was similar to that obtained in most studies in people over 65 years, with values between 10.7% and 24.8% [11–14]. Large population studies, like the work of Jencks et al., found that of the 2,961,460 beneficiaries who had been discharged from the hospital, 19.6% were readmitted within the first 30 days post-discharge, and 34% before 90 days [15].

Patients aged between 70 and 79 years suffered a higher incidence of unscheduled hospital readmissions than the rest of the study population. Age is classified as a risk factor for early readmissions, with a greater risk of readmission as age increases, mainly in those over 65 years of age. Several population-level studies carried out in adults and people ≥ 65 years have reinforced this association [16–18]. However, other studies have not found this relationship [19,20]. Even when assessing age within a predictive model of readmission risk, most researchers have not found an association strong enough to include it in the final predictive model [21]. Interestingly, in the present work, people over 80 were less likely to be readmitted. This may be due to higher mortality or greater hospital care in these people [22]. It is important to remember that the group of older people is highly heterogeneous; for example, the presence or absence of frailty could determine the prognosis of these patients. As Liperoti et al. suggested, the identification of frailty among older adults should be considered relevant to provide individualized strategies of care [23]. In our population, as in others in the literature, men entered more frequently [24,25]. However, sex has been found not to be relevant in other studies [17,26].

The service with the highest number of cases and, at the same time, with the highest number of UHR was internal medicine. There is little information about this regard, but a lower (5%) readmission rate was found in an earlier retrospective cohort study conducted by Fabbian et al. [27]. However, the mean age of this previous study was lower than that of our work. On the other hand, those with a worse medical history tended to be readmitted at a higher rate. This issue was also previously described, especially in chronic heart failure, chronic obstructive pulmonary disease (COPD), and diabetes mellitus [28]. Smoking and drinking habits were also related to UHR. Tobacco is the main cause of preventable death and has a special importance in the development of COPD. In fact, in a comparative analysis Europe and the United States readmission rates, COPD was shown to be a leading cause of hospital readmission rates [29]. The incidence of unscheduled hospital readmissions was quite similar throughout the different months, although during summer and December, an increase in the incidence of UHR was observed. Regarding time, the UHR incidence was higher at the morning. To the best of our knowledge, no earlier studies have been found that evaluated a relation between time of admission/discharge and UHR probability; therefore, these results leave an open line of research for future work.

As the main goal of this work was to develop a predictive model of UHR, as a first step, a multivariate logistic regression model was estimated. This predictive model deduced that patients with chronic therapy had a 3.75 higher risk of UHR. This result supports those of Morandi et al., who found that the consumption of ≥ 7 drugs increased the risk of early readmission in chronic patients in a post-acute rehabilitation unit [30]. On the other hand, this analysis also included length of stay in the model, since each additional day a patient spends in the hospital, his/her risk of readmission increases by 2%. This variable gained a lot of interest in the first works evaluating UHR, since it was thought that high-risk patients with an inappropriately “short” stay could have a higher rate of readmissions linked to low-quality hospital care [31]. However, most studies have described the opposite, with an increased risk of readmission in longer stays [32]. It cannot be ruled out that longer stays occur in those patients with pathologies that require certain hospital services with a greater probability of urgent readmission. In fact, those patients from the digestive surgery unit had a three-fold higher probability of UHR [33]. In contrast, those patients admitted in the cardiology unit showed a 50% lower probability of UHR, which may show a better hospital care in this service, or simply a lower probability of complications related to the patients’ conditions [34].

Considering the low performance in several metrics of the logistic model, a ML procedure was developed. The results showed that the Random Forest algorithm performed best among the prediction models selected for this study. Through this procedure, days of stay were identified as the most relevant factor. Again, this result is of great interest, since, in general, we can suppose that a short stay can be associated with inadequate care, or an imprecise diagnosis; however, a longer stay, in our case, was not associated with better patient care [35]. Interestingly, those patients with history of allergies were also at higher probability of unscheduled hospital readmissions [36].

A key advantage of ML-based predictions over traditional models lies in their ability to uncover and use variables that may not initially appear relevant, thereby enhancing the predictive accuracy. The present machine learning model has notably found several of such features as crucial for prediction, with ZIP code being a prime example. Traditional location studies, often conducted at the national level, have been instrumental in pinpointing regions with heightened disease incidence or prevalence. Nonetheless, our data facilitate more granular, regional analyses, revealing specific zones with an elevated probability of UHR, as in this study. This granular approach offers unprecedented insights, enabling the identification of areas where either post-hospital or primary care may be lacking, potentially contributing to higher readmission rates. Such findings underscore the superiority of machine learning methodologies in not only discerning subtle, yet impactful patterns within data but also in aiding the strategic enhancement of healthcare services to address and mitigate factors leading to increased unscheduled hospital readmissions rates [37]. Nevertheless, although all these parameters are important and should be considered to predict UHR, it is even more relevant to develop an individualized prediction of the UHR risk. Until recently, developing individual prediction models was very expensive regarding time and computing resources. Currently, we can identify the most important risk factors for each individual. For instance, in the example described previously, dyslipidaemia was identified as the most significant factor associated with UHR in a particular patient, while the ML or the logistic models did not identify this variable.

In this study, we have pioneered an individualized ML model that significantly enhances patient care by tailoring treatment plans to individual needs upon hospital admission. This model stands apart from traditional, non-interpretable models by offering a clear understanding of the factors influencing its predictions, thereby ensuring more precise and personalized care. The adoption of this innovative approach is pivotal in reducing unplanned hospital readmissions, a critical issue in clinical practice and the discharge process that imposes substantial economic and healthcare burdens. For instance, in the United States, annual costs for hospital readmission reach USD 41.3 billion, making it one of the costliest events to treat nationwide [38]. In our location, although the health system is complex since management is decentralized, it is estimated that the costs of hospital readmission reach EUR 500 million annually [39]. Moreover, UHR adversely affect the quality of life of patients and their families. By integrating interpretable and individualized models, we not only advance patient care but also contribute to a more sustainable healthcare system by mitigating the negative implications of UHR.

Several limitations should be considered at this time. The association of some factors with rehospitalization does not mean the identification of causal relationships, and therefore, we cannot affirm that the variables that define the predictive models are the causal factors of that readmission. Additionally, recall bias can affect future data collection. Another limitation derives from the origin of the data. The data obtained in this study are valid for the geographical area, so the performance of the present ML prediction model should be evaluated in other locations before its implementation. Although we have collected enough individuals in the final dataset, and even a much larger sample than in other works developing predictive models [37,40,41], it could be of interest to evaluate similar data from different regions to confirm the extrapolability of the current model. Finally, the presence of latent features that may be affecting hospital readmission and that have not been taken into account in this project, such as the exact values of certain biochemical or clinical tests,

could improve the machine learning model that we have obtained. In addition, it could be of great interest to include prognostic factors at discharge, like cognitive functioning, a situation that could be of specific relevance to older people and specific hospital services, as described by Fusco et al. [42].

However, we wanted to remember that, precisely, one of the objectives was to be able to determine, with basic variables and without added cost, the probability of readmission. Our model showed an AUC of 99%, demonstrating high data reliability.

5. Conclusions

Our analysis leveraged descriptive statistics to illuminate the demographic traits and the associated risk probabilities of emergency hospital admissions. Notably, we observed that most readmissions occurred before reaching 80 years, with factors such as the patient's locality and sex playing a pivotal role. To tackle the challenge of predicting UHR, we conducted a multivariate analysis focused on various demographic characteristics. The duration of the hospital stay emerged as the primary predictor for readmission, indicating a direct correlation between longer stays and increased readmission risks. Additional significant factors included a history of diabetes mellitus, allergies, or adverse reactions to drugs. The centrepiece of our study is the development of an individualized risk model, designed to offer personalized risk predictions for hospital readmission. While the current model showcases high accuracy and reliability, further enhancement is recommended through the inclusion of additional numerical variables, such as biochemical or clinical parameters. This would allow for a more comprehensive analysis and improve the model's effectiveness.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/make6030080/s1>, Figure S1: Incidence of unscheduled hospital readmission attending to the admission and discharge services and the classification systems (International Classification of Diseases) IDC-10 and Clinical Classifications Software (CCS); Figure S2: Incidence of unscheduled hospital readmission attending to the clinical attending to the date and time of admission (blue) and discharge (reds); Figure S3: Incidence of unscheduled hospital readmission attending to the clinical antecedents of participants; Figure S4: Machine learning performance parameters; Table S1: Number of patients who were readmitted or not, depending on the admission service (Table S1a) or the discharge service (Table S1b); Table S2: Contingency Tables to evaluate the association between clinical history and the presence of an unscheduled hospital readmission; Table S3: Logistic regression models grouped by different features. Model 1 was performed using admission and discharge services. Model 2 was performed using ICD-10 and CCS systems. Model 3 was performed using clinical history and other characteristics (age and sex).

Author Contributions: R.R.d.S.M.: conceptualization, methodology, formal analysis, data curation, writing—original draft. C.M.-H.: formal analysis, writing—original draft. C.B.: conceptualization, methodology, formal analysis, data curation, writing—original draft. C.M.-C.: data curation, software, formal analysis. A.J.B.-L.: data curation, software, formal analysis. F.J.S.-M.: data curation, software, formal analysis. H.P.-S.: data curation, software, formal analysis, project administration, funding acquisition. I.M.-M.: writing—original draft. J.J.H.-M.: investigation, resources, data curation, writing—original draft, writing—review and editing, visualization, supervision, project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by grants from the European Project Horizon 2020 SC1-BHC-02-2019 [REVERT, ID:848098]. Supercomputing resources in this work were supported by the Plataforma Andaluza de Bioinformática at the University of Málaga, the supercomputing infrastructure of the NLHPC (ECM-02, Powered@NLHPC), and the Extremadura Research Centre for Advanced Technologies (CETA-CIEMAT), funded by the European Regional Development Fund (ERDF). CETA-CIEMAT is part of CIEMAT and the Government of Spain.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Hospital Universitario “Virgen de la Arrixaca” (CSV verification code: CARM-665aef0c-6cb0-44ff-7120-0050569b6280; date of approval 30 April 2019).

Data Availability Statement: The code to perform all machine learning analysis is available at: <https://github.com/bio-hpc/sibila> (accessed on 28 June 2024). Clinical data can be requested from Dr. Juan José Hernández Morante (jjhernandez@ucam.edu) under reasonable circumstances.

Acknowledgments: The authors wish to express our gratitude to the Bioinformatics Service of the Hospital Virgen de la Arrixaca. We also wish to thank the Department of Biostatistics of the University of Murcia. Finally, we wish to thank Ascensión López (RN) for her assistance with data handling.

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

References

- Guntaka, S.M.; Tarazi, J.M.; Chen, Z.; Vakharia, R.; Mont, M.A.; Roche, M.W. Higher Patient Complexities Are Associated with Increased Length of Stay, Complications, and Readmissions After Total Hip Arthroplasty. *Surg. Technol. Int.* **2021**, *38*, 422–426. [CrossRef] [PubMed]
- Charlson, M.E.; Pompei, P.; Ales, K.L.; MacKenzie, C.R. A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *J. Chronic Dis.* **1987**, *40*, 373–383. [CrossRef]
- De Groot, V.; Beckerman, H.; Lankhorst, G.J.; Bouter, L.M. How to Measure Comorbidity: A Critical Review of Available Methods. *J. Clin. Epidemiol.* **2003**, *56*, 221–229. [CrossRef]
- Ministerio de Sanidad Ministerio de Sanidad—Sanidad En Datos—Registro de Altas de Los Hospitales Del Sistema Nacional de Salud. CMBD. Available online: <https://www.sanidad.gob.es/estadEstudios/estadisticas/cmbdhome.htm> (accessed on 28 June 2024).
- Pohjanpää, M.; Ojala, R.; Luukkaala, T.; Gissler, M.; Tammela, O. Association of Early Discharge with Increased Likelihood of Hospital Readmission in First Four Weeks for Vaginally Delivered Neonates. *Acta Paediatr. Int. J. Paediatr.* **2022**, *111*, 1144–1156. [CrossRef]
- Xue, H.T.; Stanley-Baker, M.; Kong, A.W.K.; Li, H.L.; Goh, W.W. Bin Data Considerations for Predictive Modeling Applied to the Discovery of Bioactive Natural Products. *Drug Discov. Today* **2022**, *27*, 2235–2243. [CrossRef] [PubMed]
- Sutter, T.; Roth, J.A.; Chin-Cheong, K.; Hug, B.L.; Vogt, J.E. A Comparison of General and Disease-Specific Machine Learning Models for the Prediction of Unplanned Hospital Readmissions. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 868–873. [CrossRef]
- Habehh, H.; Gohel, S. Machine Learning in Healthcare. *Curr. Genom.* **2021**, *22*, 291–300. [CrossRef] [PubMed]
- Zhou, Y.; Han, W.; Yao, X.; Xue, J.J.; Li, Z.; Li, Y. Developing a Machine Learning Model for Detecting Depression, Anxiety, and Apathy in Older Adults with Mild Cognitive Impairment Using Speech and Facial Expressions: A Cross-Sectional Observational Study. *Int. J. Nurs. Stud.* **2023**, *146*, 104562. [CrossRef] [PubMed]
- Sammur, S.J.; Crispin-Ortuzar, M.; Chin, S.F.; Provenzano, E.; Bardwell, H.A.; Ma, W.; Cope, W.; Dariush, A.; Dawson, S.J.; Abraham, J.E.; et al. Multi-Omic Machine Learning Predictor of Breast Cancer Therapy Response. *Nature* **2022**, *601*, 623–629. [CrossRef] [PubMed]
- Robinson, S.; Howie-Esquivel, J.; Vlahov, D. Readmission Risk Factors after Hospital Discharge Among the Elderly. *Popul. Health Manag.* **2012**, *15*, 338–351. [CrossRef]
- Laniée, I.; Couturier, P.; Dramé, M.; Gavazzi, G.; Lehman, S.; Jolly, D.; Voisin, T.; Lang, P.O.; Jovenin, N.; Gauvain, J.B.; et al. Incidence and Main Factors Associated with Early Unplanned Hospital Readmission among French Medical Inpatients Aged 75 and over Admitted through Emergency Units. *Age Ageing* **2008**, *37*, 416–422. [CrossRef] [PubMed]
- Donzé, J.; Lipsitz, S.; Bates, D.W.; Schnipper, J.L. Causes and Patterns of Readmissions in Patients with Common Comorbidities: Retrospective Cohort Study. *BMJ* **2013**, *347*, f7171. [CrossRef] [PubMed]
- Cornette, P.; D’Hoore, W.; Malhomme, B.; Van Pee, D.; Meert, P.; Swine, C. Differential Risk Factors for Early and Later Hospital Readmission of Older Patients. *Ageing Clin. Exp. Res.* **2005**, *17*, 322–328. [CrossRef] [PubMed]
- Jencks, S.F.; Williams, M.V.; Coleman, E.A. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. *N. Engl. J. Med.* **2009**, *360*, 1418–1428. [CrossRef] [PubMed]
- Billings, J.; Blunt, I.; Steventon, A.; Georgiou, T.; Lewis, G.; Bardsley, M. Development of a Predictive Model to Identify Inpatients at Risk of Re-Admission within 30 Days of Discharge (PARR-30). *BMJ Open* **2012**, *2*, e001667. [CrossRef] [PubMed]
- Allaudeen, N.; Vidyarthi, A.; Maselli, J.; Auerbach, A. Redefining Readmission Risk Factors for General Medicine Patients. *J. Hosp. Med.* **2011**, *6*, 54–60. [CrossRef] [PubMed]
- Coderch, J.; Sánchez-Pérez, I.; Ibern, P.; Carreras, M.; Pérez-Berrueto, X.; Inoriza, J.M. Predicción Del Riesgo Individual de Alto Coste Sanitario Para La Identificación de Pacientes Crónicos Complejos. *Gac. Sanit.* **2014**, *28*, 292–300. [CrossRef]
- Fleming, L.M.; Gavin, M.; Piatkowski, G.; Chang, J.D.; Mukamal, K.J. Derivation and Validation of a 30-Day Heart Failure Readmission Model. *Am. J. Cardiol.* **2014**, *114*, 1379–1382. [CrossRef]
- Dharmarajan, K.; Hsieh, A.F.; Lin, Z.; Bueno, H.; Ross, J.S.; Horwitz, L.I.; Barreto-Filho, J.A.; Kim, N.; Bernheim, S.M.; Suter, L.G.; et al. Diagnoses and Timing of 30-Day Readmissions after Hospitalization for Heart Failure, Acute Myocardial Infarction, or Pneumonia. *JAMA* **2013**, *309*, 355–363. [CrossRef]
- Kansagara, D.; Englander, H.; Salanitro, A.; Kagen, D.; Theobald, C.; Freeman, M.; Kripalani, S. Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA* **2011**, *306*, 1688–1698. [CrossRef]

22. Glans, M.; Kragh Ekstam, A.; Jakobsson, U.; Bondesson, Å.; Midlöv, P. Risk Factors for Hospital Readmission in Older Adults within 30 Days of Discharge—A Comparative Retrospective Study. *BMC Geriatr.* **2020**, *20*, 467. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Liperoti, R.; Vetrano, D.L.; Palmer, K.; Targowski, T.; Cipriani, M.C.; Lo Monaco, M.R.; Giovannini, S.; Acampora, N.; Villani, E.R.; Bernabei, R.; et al. Association between Frailty and Ischemic Heart Disease: A Systematic Review and Meta-Analysis. *BMC Geriatr.* **2021**, *21*, 357. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Silverstein, M.D.; Qin, H.; Mercer, S.Q.; Fong, J.; Haydar, Z. Risk Factors for 30-Day Hospital Readmission in Patients ≥ 65 Years of Age. *Bayl. Univ. Med. Cent. Proc.* **2008**, *21*, 363. [\[CrossRef\]](#)
25. Amarasingham, R.; Moore, B.J.; Tabak, Y.P.; Drazner, M.H.; Clark, C.A.; Zhang, S.; Reed, W.G.; Swanson, T.S.; Ma, Y.; Halm, E.A. An Automated Model to Identify Heart Failure Patients at Risk for 30-Day Readmission or Death Using Electronic Medical Record Data. *Med. Care* **2010**, *48*, 981–988. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Bisiani, M.A.; Jurgens, C.Y. Do Collaborative Case Management Models Decrease Hospital Readmission Rates among High-Risk Patients? *Prof. Case Manag.* **2015**, *20*, 188–196. [\[CrossRef\]](#)
27. Fabbian, F.; Boccafogli, A.; De Giorgi, A.; Pala, M.; Salmi, R.; Melandri, R.; Gallerani, M.; Gardini, A.; Rinaldi, G.; Manfredini, R. The Crucial Factor of Hospital Readmissions: A Retrospective Cohort Study of Patients Evaluated in the Emergency Department and Admitted to the Department of Medicine of a General Hospital in Italy. *Eur. J. Med. Res.* **2015**, *20*, 6. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Kroeze, E.D.; de Groot, A.J.; Smorenburg, S.M.; Mac Neil Vroomen, J.L.; van Vught, A.J.A.H.; Buurman, B.M. A Case Vignette Study to Refine the Target Group of an Intermediate Care Model: The Acute Geriatric Community Hospital. *Eur. Geriatr. Med.* **2024**. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Westert, G.P.; Lagoe, R.J.; Keskimäki, I.; Leyland, A.; Murphy, M. An International Study of Hospital Readmissions and Related Utilization in Europe and the USA. *Health Policy* **2002**, *61*, 269–278. [\[CrossRef\]](#)
30. Morandi, A.; Bellelli, G.; Vasilevskis, E.E.; Turco, R.; Guerini, F.; Torpilliesi, T.; Speciale, S.; Emiliani, V.; Gentile, S.; Schnelle, J.; et al. Predictors of Rehospitalization among Elderly Patients Admitted to a Rehabilitation Hospital: The Role of Polypharmacy, Functional Status and Length of Stay. *J. Am. Med. Dir. Assoc.* **2013**, *14*, 761. [\[CrossRef\]](#)
31. Grovu, R.; Huo, Y.; Nguyen, A.; Mourad, O.; Pan, Z.; El-Gharib, K.; Wei, C.; Mustafa, A.; Quan, T.; Slobodnick, A. Machine Learning: Predicting Hospital Length of Stay in Patients Admitted for Lupus Flares. *Lupus* **2023**, *32*, 1418–1429. [\[CrossRef\]](#)
32. Zanolchi, M.; Maero, B.; Martinelli, E.; Cerrato, F.; Corsinovi, L.; Gonella, M.; Ponte, E.; Luppino, A.; Margolici, A.; Molaschi, M. Early Re-Hospitalization of Elderly People Discharged from a Geriatric Ward. *Aging Clin. Exp. Res.* **2006**, *18*, 63–69. [\[CrossRef\]](#)
33. Im, K.M.; Kim, E.Y. Identification of ICU Patients with High Nutritional Risk after Abdominal Surgery Using Modified NUTRIC Score and the Association of Energy Adequacy with 90-Day Mortality. *Nutrients* **2022**, *14*, 946. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Albinali, H.H.; Singh, R.; Al Arabi, A.; Al Qahtani, A.; Asaad, N.; Al Suwaidi, J. Predictors of 30-Day Re-Admission in Cardiac Patients at Heart Hospital, Qatar. *Heart Views* **2023**, *24*, 125. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Mechelli, A.; Lin, A.; Wood, S.; McGorry, P.; Amminger, P.; Tognin, S.; McGuire, P.; Young, J.; Nelson, B.; Yung, A. Using Clinical Information to Make Individualized Prognostic Predictions in People at Ultra High Risk for Psychosis. *Schizophr. Res.* **2017**, *184*, 32–38. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Ortiz-Barrios, M.; Altamar-Maldonado, Z.; Martínez-Solano, C.; Petrillo, A.; De Felice, F.; Jiménez-Delgado, G.; García-Cuan, A.; Medina-Buelvas, A.M. Predicting 15-Day Unplanned Readmissions in Hospitalization Departments: An Application of Logistic Regression. *Ingeniare Rev. Chil. Ing.* **2021**, *29*, 378–398. [\[CrossRef\]](#)
37. Conilione, P.; Jessup, R.; Gust, A. Novel Machine Learning Model for Predicting Multiple Unplanned Hospitalisations. *BMJ Health Care Inform.* **2023**, *30*, e100682. [\[CrossRef\]](#)
38. Yhdego, H.H.; Nayebnazar, A.; Amrollahi, F.; Boussina, A.; Shashikumar, S.; Wardi, G.; Nemati, S. Prediction of Unplanned Hospital Readmission Using Clinical and Longitudinal Wearable Sensor Features. *medRxiv* **2023**. [\[CrossRef\]](#)
39. The Lancet Public Health. COVID-19 in Spain: A Predictable Storm? *Lancet Public Health* **2020**, *5*, e568. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Lo, Y.T.; Liao, J.C.H.; Chen, M.H.; Chang, C.M.; Li, C.T. Predictive Modeling for 14-Day Unplanned Hospital Readmission Risk by Using Machine Learning Algorithms. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 288. [\[CrossRef\]](#)
41. Yu, M.-Y.; Son, Y.-J. Machine Learning-Based 30-Day Readmission Prediction Models for Patients with Heart Failure: A Systematic Review. *Eur. J. Cardiovasc. Nurs.* **2024**, zvae031. [\[CrossRef\]](#)
42. Fusco, A.; Galluccio, C.; Castelli, L.; Pazzaglia, C.; Pastorino, R.; Pires Marafon, D.; Bernabei, R.; Giovannini, S.; Padua, L. Severe Acquired Brain Injury: Prognostic Factors of Discharge Outcome in Older Adults. *Brain Sci.* **2022**, *12*, 1232. [\[CrossRef\]](#)

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.