Predictive Application for Early *Delirium* Detection Subtypes Using GLM's

Alexandra Coelho[1(✉)] and Ana Cristina Braga[2]

[1] School of Engineering, University of Minho, 4710-057 Braga, Portugal
alexmcoelho11@gmail.com
[2] ALGORITMI Research Centre, LASI, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
acb@dps.uminho.pt

Abstract. *Delirium* is an acute neuropsychiatric dysfunction prevalent in patients admitted to inpatient and intensive care hospital settings. As a multifactorial manifestation, it is typically underdiagnosed and overlooked. *Delirium* can be categorized, based on motor activity profile, into hypoactive and hyperactive subtypes. Multinomial logistic regression predictive models are often implemented to identify the most influential variables, as they allow for modelling the relationship between predictors and a multinomial dependent variable. In this context, the goal of this paper arises, aiming to develop an application capable of predicting the occurrence of *delirium* and its subtypes using the GLMs methodology. Subsequently, variable selection was performed using various techniques, with the *Elastic Net* method having an *alpha* value of 0.1 showing the best performance. The model achieved combines the multinomial logistic regression algorithm with the *elastic net* method, which is included in the web application. For the hypoactive subtype, it allowed the selection of 27 variables, resulting in an AUC-ROC of 0.691. The most influential variables include the length of hospitalization in days, alcoholism, analgesics, cardiotonics, as well as the diagnostic group related to toxicity and drugs. Regarding the hyperactive subtype, the model identified 29 relevant variables, with an AUC-ROC of 0.531. The most impactful variables include PCR, age, pO_2, SIRS criteria, and the ER source, specifically UDC1 (Clinical Decision Unit identifies high-priority patients with assigned yellow wristbands). The application is available at https://alexandra-coelho.shinyapps.io/Delirium_detection_app/. While further enhancements are possible, this predictive model remains a valuable tool for healthcare professionals diagnosing *delirium* in emergency rooms.

Keywords: *delirium* · Generalized linear model · Multinomial logistic regression

This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024 Workshops, LNCS 14815, pp. 375–392, 2024.
https://doi.org/10.1007/978-3-031-65154-0_23

1 Introduction

In recent years, advancements in medical technology have played a crucial role in early disease detection and improved treatment outcomes, contributing to a noteworthy increase in life expectancy and overall well-being. However, the ageing demographic has led to a rise in neurocognitive decline, impacting both physical and cognitive functions.

The natural ageing process, or senescence, is influenced by a combination of biological, psychological, and social factors. While positive strides have been made, mental disorders continue to be a significant global health concern, affecting a quarter of the population. Consequently, there is a heightened medical focus on developing tools that can predict and detect mental illnesses early on, leveraging symptoms and associated risk factors for more effective intervention and management of mental health issues [33].

Currently, based on the DSM-V (*Diagnostic and Statistical Manual of Mental Disorders*) definition, *delirium* is a severe neuropsychiatric syndrome characterized primarily by disturbances in consciousness, incoherent thoughts, attention deficit, and cognitive dysfunction, sometimes accompanied by psychotic symptoms and mood alteration with fluctuations throughout the day [5–7,29].

Generally, it develops over a short period of hours or days [21], and there is evidence that the disturbance is a direct physiological consequence of a medical condition or multiple etiologies [5,6,28]. The symptomatology reflects deficits in three domains: the cognitive, which is related to orientation, attention, short and long-term memory, and visuospatial perception; the domain of higher-level thinking, which is involved in language impairment and thought processes; and the circadian, which includes the sleep-wake cycle and motor activity disturbance [15,18].

In general, the disorder is associated with adverse outcomes, reflecting increased mortality, prolonged hospitalization, functional and cognitive decline, as well as the exacerbation of a pre-existing clinical picture of dementia [14,36]. As the population ages, the prevalence of chronic diseases increases. In parallel to this fact, *delirium* is a common acute illness in elderly individuals [25], requiring urgent attention. It is often underdiagnosed and can lead to mortality rates exceeding 25%, accelerate the onset of dementia, decrease independence, and consequently, reduce the quality of life, with a significantly high associated hospitalization cost [20,26]. The prevalence of *delirium* varies from 3 to 52% depending on the studied population, their environment [2,19], the type of hospitalization, and the detection method applied [21]. The occurrence rate in the general community is less than 2%, but it increases with age [21].

According to Lipowski [24], *delirium* can be categorized into three clinically defined subgroups based on the motor activity profile. These subgroups differ in terms of pathophysiology, detection rates, and the duration of the episode [34]. Individuals with the hyperactive subtype exhibit a state of hyperalertness, often characterized by hypervigilance, agitation, hallucinations, verbal and physical confusion, aggression, hyperexcitability, and restlessness [4,12]. It is estimated

to occur in approximately 15 to 25% of all cases, mainly manifesting in *delirium tremens* and other syndromes resulting from substance withdrawal [8,10].

In contrast, the hypoactive subtype shows a decrease in alertness, lethargy, reduced awareness of the surrounding environment, apathy, and slow and inappropriate speech [32]. This form of *delirium* is the most common; however, it significantly presents a worse prognosis, resulting in a higher associated mortality rate [8,35]. Patients most prone to developing this phenotype are typically older individuals, anaemic, and more susceptible to injuries during hospitalization [32].

Currently, there are studies and clinical assessment tools, such as the Confusion Assessment Method, for the study and prediction of the disease based on a set of risk factors. However, in the context of a hospital emergency, the patient is subject to various risk factors that make them more susceptible to the progression of the disorder. In these circumstances, time is crucial, as well as the correct and early diagnosis to intervene as quickly as possible. Therefore, there is a pressing need for an effective tool capable of predicting the occurrence of *delirium* and its subtypes to minimize the detrimental consequences of this disease. In this regard, to achieve a more accurate diagnosis, we aim to develop an application based on the Generalized Linear Models (GLM) methodology, specifically multinomial logistic regression [1,17,27].

Predictive models of multinomial logistic regression are commonly implemented to determine the variables that have the greatest influence, as they allow modelling the relationship between multiple independent variables and a categorical dependent variable [9]. In the present case study, it will be used to estimate the probability of an individual presenting one of the subtypes of *delirium*.

After identifying the set of most influential independent variables to include in the model, the statistical significance level of the estimated parameters should be evaluated. To obtain robust results for this classification problem, validating the developed model is essential.

The main motivation for the development of this study stems from the increasing incidence of neuro and psychological disorders, coupled with the negligence associated with their diagnosis due to a lack of appropriate tools in a hospital emergency setting. In light of this, leveraging the data available upon a patient's admission to the emergency department (ED) of a portuguese hospital, the proposal is to develop a computer application that determines the probability of a patient presenting one of the subtypes of *delirium* in this context. The aim is to assist healthcare professionals in the diagnostic process and, consequently, restore the well-being of the patient.

This paper is divided into five sections, with the first part dedicated to contextualizing the subject. It offers a thorough examination of *delirium* and its prevalence, emphasizing the need to develop a predictive model for precisely determining its occurrence. The second section provides a comprehensive account of the methodology implemented, intricately explaining the modelling process, which encompasses strategies for feature engineering and variable selection.

In the third section, we present and discuss the results obtained from the implementation of the multinomial logistic regression models using different variable selection techniques. The aim is to identify the model that exhibits the most effective classificatory performance in predicting the occurrence of the disorder and its subtypes. An in-depth analysis of the most influential variables for each subtype is then presented. Subsequently, we introduce the developed application designed to aid in the diagnosis. The fourth section presents the deduced conclusions derived from the case study. Finally, the last section delves into prospects for improving the model to attain improved predictions, therefore enhancing the performance of the predictive model.

2 Material and Methods

2.1 Generalized Linear Models

Many data science models provide the option to specify the relationship used to describe the response variable. For instance, a generic family of models that can represent a wide variety of relationships between input and response variables is the Generalized Linear Model (GLM) [22]. GLMs follow the structure of a linear model but allow the response variable to have an arbitrary error distribution (rather than simply normal distributions). They enable the estimation of regression models for univariate response data that follow a very general distribution known as the exponential family, which includes Gaussian, Binomial, Poisson, and Gamma distributions. Each serves a different purpose, and depending on the distribution and the choice of the link function, it may be used for prediction or classification, resulting in various types of models [11].

In this case study, the focus will primarily be on multinomial logistic regression models, as they are frequently implemented to identify the most influential variables. These models capture the relationship between multiple independent variables and a multicategorical dependent variable to determine the most contributive variables for each subtype of delirium, thereby estimating the probability of an individual presenting one of the disorder's subclassifications.

2.2 Multinomial Logistic Regression

Given the exponential growth in data volume driven by technological advancements and the widespread use of digital devices, the application of machine learning techniques has become crucial for extracting knowledge from vast amounts of information. In the extensive field of machine learning, the goal is to predict an outcome based on available data, a task also known as a classification problem. When the outcome represents more than two different classes, it is referred to as multinomial. Regression methods have long been a fundamental part of any multivariate data analysis, essential for describing the relationship between an outcome variable and one or more explanatory variables.

Multinomial logistic regression (MLR) is an approach that allows the analysis of relationships between variables, both independent and dependent, when

the values of the dependent variable are categorical and have more than two categories. Typically, the dependent variable is nominal, and independent variables can be categorical and/or continuous. This method aims to identify distinctive characteristics among groups in the dependent variable based on the distribution of data from independent variables. Ultimately, it seeks to identify the factors or variables that most influence the clustering of data in both independent and dependent variables [3]. The impact of predictor variables is usually explained in terms of odds ratios (OR). Multinomial logistic regression is not the only estimation methodology available for situations involving unordered categorical dependent variables. However, it stands out for three significant advantages over other methods. Firstly, this technique is widely accessible as it is incorporated into almost all commonly used statistical packages. Additionally, computational efficiency allows for relatively quick calculations, particularly in scenarios with numerous categories. Finally, it is worth noting that the results obtained through this model are easily interpretable [23].

2.3 Modulation

As outlined in the preceding investigation [13], the initial dataset comprised 112 variables and 434 patients. Following thorough pre-processing, a standardized dataset was derived, featuring a reduced set of 29 independent variables. These variables encompass details pertaining to patient origin, admission category, patient characteristics, duration of stay in the emergency department, diagnostic group, blood test results, administered pharmacological groups, and alcohol dependence.

Feature Engineering. Recognizing the imbalanced distribution within the dataset, a stratified split methodology was adopted to maintain proportional class representation in both the training and testing sets. Post-split, the training set featured 238 instances characterized by the absence of delirium, 56 instances of the hypoactive subtype, and 10 instances of the hyperactive subtype. The residual examples were systematically assigned to the test dataset.

Nevertheless, to address the imbalance within the training dataset, methodologies were implemented to rectify this issue, with oversampling deemed the most appropriate approach in the context of limited data. Oversampling involves the generation of synthetic samples for the minority class. Consequently, two techniques were considered: SMOTE (Synthetic Minority Over-sampling Technique) and ADASYN (Adaptive Synthetic Sampling).

In contrast to SMOTE, ADASYN consistently achieved higher levels of precision, sensitivity, and discriminative capacity. This enhancement is perceptible in the superior values observed for the F_1 score, precision, AUC-ROC, and AUC-PR as achieved with ADASYN. These outcomes substantiate compelling evidence affirming the superiority of ADASYN in augmenting the performance of classification models within scenarios characterized by imbalanced data [13].

Variable Selection. The challenge posed by high-dimensional data is evident in machine learning methods. When confronted with a multitude of variables, models tend to overfit, resulting in a diminished performance. Variable selection is a widely employed approach to alleviate data dimensionality. Its objective is to identify the subset of variables that are most relevant and informative, leading to enhanced performance, reduced computational cost, and a better understanding of the model. Additionally, it results in more robust models with improved generalization capabilities when confronted with new data.

The process of variable selection holds paramount significance in the training phase of classification problems. Initially, this method discerns a subset comprising the most contributive variables, which are subsequently fed into the multinomial logistic regression model. The overarching objective of the optimal regression model is to fine-tune the relationship between the response variable and explanatory variables, striving to minimize the number of explanatory variables pivotal in distinguishing between different classes. Variable selection techniques are broadly classified into three categories: filtering, *wrapper*, and *embedded* methods.

Based on the characteristics of the data, filtering models evaluate variables without the need for implementing classification algorithms. Variables are classified according to specific statistical criteria, assigning a score to each one. However, due to its univariate nature, the individual calculation of scores for each variable does not consider their interactions, leading to a reduction in classification performance. This reduction is primarily attributed to the fact that the combination of numerous variables may result in the development of one of the subtypes of *delirium*. Consequently, predominantly due to this drawback, this method was not considered.

Wrapper methods are an approach involving the creation of variable subsets and the evaluation of their performance using a predictive model during the training phase. These methods include simple techniques such as sequential search. During training, multinomial logistic regression assesses the accuracy rate to determine the relevance of variables. Sequential search strategies are particularly computationally advantageous and robust against overfitting. This method includes commonly implemented techniques such as *forward*, *backward*, *bidirectional*, and Recursive Feature Elimination (RFE).

The *forward* strategy initiates the process with one predictive variable and iteratively incorporates a variable only if it leads to improved model performance. When the potential inclusion of a new variable does not enhance performance, the procedure concludes. The method continues until none of the remaining variables provides a significant improvement or until specific stopping criteria are met.

In the *backward* method, the model initially includes all variables, and in each iteration, it progressively eliminates the least significant variable until reaching a reduced model that still adequately describes the data and improves model performance. The variable elimination procedure occurs sequentially until a predefined stopping rule is satisfied. A common example of a stopping rule is when

all remaining variables in the model are statistically significant at a predefined significance level, meaning that each remaining variable is statistically relevant to the model.

The bidirectional method, in turn, combines the two aforementioned methods. At each step of the process, it chooses the best predictor to add to the model and removes the least relevant variable from the remaining ones. This method reevaluates both the addition and removal of attributes as the procedure advances. To implement the first three sequential search methods for variable selection in multinomial logistic regression, it is necessary to use the `MASS` package in combination with the `multinom` function from the `nnet` package [30,31].

The RFE is a technique aimed at selecting variables iteratively. Initially, the model is trained with all variables, and the importance of each is assessed through a specific attribute. Then, the less relevant attributes are removed from the current set of predictors. This process is repeated recursively until the desired number of variables is reached.

The *embedded* method combines variable selection with the construction of the learning algorithm, similar to *wrapper*. Moreover, it is seen as the merger of the advantages of *wrapper* models, as it involves interaction with the classification model, and filtering models, which assume lower computational power. These models implement regularization techniques, seeking to minimize model fitting errors while continuously forcing the coefficients to be small or even equal to zero, penalizing them if they are too far from this value. In this way, it reduces the complexity of the model, leading to the elimination of variables whose coefficients approach zero. Some examples of predictive methods that perform *embedded* variable selection include *Least Absolute Shrinkage and Selection Operator* (LASSO) regularization, *Ridge* regularization, and *Elastic Net*.

The LASSO regularization is based on the use of $L1$ penalty to assess coefficients. This technique imposes penalties on the non-zero coefficients of predictors, penalizing the sum of their absolute values.

On the other hand, *Ridge* regularization penalizes the sum of squared coefficients, which corresponds to the $L2$ penalty. This method adds a penalty associated with the coefficients to the logistic regression cost function, aiming to minimize multicollinearity issues and overfitting to the data. Thus, it is possible to reduce the complexity of the model while including all variables. As a consequence, in cases of significantly high values of *lambda*, several coefficients exhibit precisely zero in the LASSO method, a situation that never occurs in *Ridge* regularization.

It is common for some variables to be highly correlated. In these situations, the LASSO method tends to choose only one of the correlated attributes. Thus, *Elastic Net* arises as the combination of *Ridge* and LASSO, which applies penalties to the coefficients of the variables using two hyperparameters, one related to the $L1$ penalty and the other to the $L2$ penalty. This results in the fusion of the advantages of the aforementioned techniques, i.e., automatic variable selection ($L1$) and control of multicollinearity ($L2$). The values of *alpha* control the type of regularization. When *alpha* is set to 0, it corresponds to *Ridge* regularization,

1 refers to LASSO, and values between 0 and 1 represent *Elastic Net*. It is noteworthy that the *alpha* parameter in the *Elastic Net* algorithm is responsible for balancing variable selection and overfitting by combining $L1$ and $L2$ penalties.

3 Results

This section discusses the results derived from the statistical analysis of different models, as per the methodology detailed in Sect. 2. The model selected for implementation in a web application, to predict the occurrence of *delirium* and its subtypes, is the one exhibiting optimal results based on performance evaluation metrics.

The primary outcome of this investigation is the formulation of a comprehensive pipeline for multinomial logistic regression, which elucidates essential procedural steps and outcomes (see Fig. 1) [13].

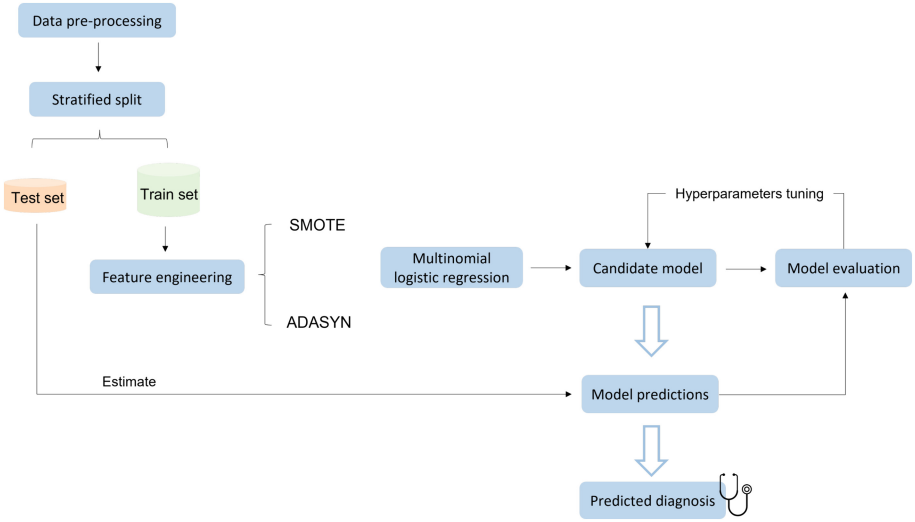

Fig. 1. Main steps of the pipeline for deploying multinomial logistic regression to forecast *delirium* outcome.

The implementation of various variable selection techniques was carried out, aiming to address the challenge of high-dimensional features by excluding less contributive attributes for the model's performance. In the context of the training dataset, it is crucial to identify which attributes will be incorporated into the model based on their impact on the correct and accurate classification of individuals with the absence or presence of one of the subtypes (hypoactive or hyperactive) of *delirium*.

Thus, after an analysis of the respective strategies as described in Subsect. 2.3, the implementation of *wrapper* and *embedded* selection methods took

place, which will be analysed and compared. The preference for these methods over filtering methods is due to the fact that these approaches take into account the existence of interactions between variables.

To seek the best subset of variables for the classifier (multinomial logistic regression), for the *wrapper* method, strategies such as *forward*, *backward*, *bidirectional*, and RFE were employed.

The implementation of cross-validation, specifically stratified repeated *k-fold*, for the first three techniques did not show improvements in the model's performance. Therefore, they were not considered for the comparison of selection techniques, and it was decided to opt for methods without cross-validation. This may be attributed to variations in the training and testing sets, the dataset dimensions, the model configuration, and the inherent randomness in the data. However, stratified repeated cross-validation was used for the RFE algorithm.

On the other hand, for the *embedded* method, regularization strategies such as *Ridge*, LASSO, and *Elastic Net* were implemented, coupled with stratified repeated k-fold cross-validation with 10 validation folds and 3 repetitions.

Specifically, for the *Elastic Net* regularization strategy, different values were considered for the *alpha* parameter, which can vary between 0 and 1, to obtain a more viable comparison with other methods. This parameter is crucial in allowing control over the balance between variable selection and preventing overfitting of the model to the training data. Values of *alpha* close to 0 result in a more generalizable model with more features, while values closer to 1 emphasize a more stringent selection of more contributive predictors. Note that the choice of the ideal *lambda* value is determined by the cross-validation approach and depends on numerous factors, such as the dataset and the nature of the problem.

To compare the various models of multinomial logistic regression (MLR) with different subsets of variables resulting from feature selection, it is necessary to construct a reference model, denoted as Model 0, without any selection technique. This model includes all 39 dummy variables, and the prediction results obtained through evaluation metrics have been recorded.

Thus, the aforementioned selection methods, both sequential search and regularization methods, use coefficients to decide which variables to include or exclude in the final model. Predictors with statistically significant coefficients are retained, while non-significant ones may be removed to simplify the model. In addition to indicating the direction (positive or negative) of the relationship between each independent variable and the classes of the dependent variable, they also reveal the magnitude of this relationship.

In multinomial logistic regression, one category of the response variable is selected as the reference, while the other classes are compared to this reference. The choice of the reference category is arbitrary and does not affect the predictive ability of the model. The coefficients for the reference category are set to zero, while the coefficients for the remaining classes represent the logarithm of the odds ratio of belonging to each category compared to the reference category.

However, it is important to note that in the case of an imbalanced dataset, with a greater emphasis on the test set, which has only 18 observations for the

Hypoactive class and 8 for the Hyperactive subtype, the metric related to the AUC-ROC (Area Under the Receiver Operating Characteristic curve) should be given greater consideration. This specific metric is more robust, as it is particularly useful in discriminating between different classes during classification (higher values indicating better distinction).

As observed from the analysis of Table 1, the models related to sequential search contribute little to the improvement of predictive performance compared to the reference model. Specifically, the forward method does not show any improvement, yielding the same estimates as the model without variable selection, suggesting that selection based on this method did not bring discernible benefits to the model.

Similarly, the *backward* and bidirectional methods reduced the number of attributes used, providing a slight improvement in predictive capability. Results for the *backward* method were similar or even superior for most evaluation metrics. On the other hand, for the model implementing the RFE (Recursive Feature Elimination) technique with stratified cross-validation and repetitions, the number of predictors significantly decreased to 27, and the results approached those of the reference model.

As mentioned earlier, the *Ridge* regularization ($alpha = 0$) is designed to prevent overfitting; consequently, the resulting model is more generalizable. This characteristic was observed as the model exhibits a high number of features (36) and, therefore, a lower predictive capacity as expected. On the other hand, the LASSO regularization ($alpha = 1$) aims to implement a more rigorous selection, reflected in a substantial reduction of variables to only 12. The estimates are similar to the reference model; however, it is still not sufficient to surpass the performance of the model without variable selection.

Finally, several models were implemented for *Elastic Net* regularization with different values of *alpha* (models 7 to 11). The model with $alpha = 0.1$ emerged as the one showing the best results compared to others, including the reference model, with superior and noticeable values, particularly for the AUC-ROC metric and a slight increase for AUC-PR and F_1. These results align with expectations, as lower values for the *alpha* parameter tend to retain more variables in the model, while higher values perform a more meticulous and selective variable selection, more aggressively reducing multicollinearity.

Although models 8, 10, and 11 show results very similar to model 7, with a smaller set of variables and reasonable values for AUC-ROC, model 7 was chosen as it placed more emphasis on the higher value achieved by the evaluation metrics. This was a decisive criterion, considering the problem involves a relatively low number of observations.

It is noteworthy that when comparing models 2, 3, and 7, although the first two implement wrapper methods and the last one employs an *embedded* method, all reflect similar performances. Therefore, the model that showed a greater reduction in dimensionality and different predictive variables for each category, corresponding to model 7, was selected.

Table 1. The results obtained by the classifier algorithm, multinomial logistic regression, for variable selection.

Multinomial logistic regression							
Model	Variables	Accuracy	Sensibility	Precision	F1	AUC-PR	AUC-ROC
0	**Reference model (without selection methods)**						
Hypoactive	39	0.662	0.242	0.444	0.320	0.227	0.638
Hyperactive	39	0.662	0.195	0.333	0.241	0.086	0.493
1	**Forward**						
Hypoactive	39	0.662	0.242	0.444	0.320	0.227	0.638
Hyperactive	39	0.662	0.195	0.333	0.241	0.086	0.493
2	**Backward**						
Hypoactive	34	0.654	0.333	0.500	0.301	0.232	0.620
Hyperactive	34	0.654	0.226	0.389	0.186	0.070	0.513
3	**Bidirecional**						
Hypoactive	34	0.652	0.367	0.500	0.222	0.287	0.669
Hyperactive	34	0.652	0.154	0.167	0.205	0.070	0.585
4	**RFE**						
Hypoactive	27	0.659	0.284	0.456	0.243	0.249	0.641
Hyperactive	27	0.659	0.106	0.158	0.176	0.066	0.483
5	**Regularization *Ridge***						
Hypoactive	36	0.631	0.253	0.333	0.286	0.254	0.646
Hyperactive	36	0.631	0.158	0.125	0.071	0.056	0.457
6	**Regularization LASSO**						
Hypoactive	12	0.569	0.242	0.444	0.314	0.265	0.643
Hyperactive	12	0.569	0.145	0.258	0.105	0.072	0.595
7	**Regularization *Elastic Net* (*alpha* = 0.1)**						
Hypoactive	27	0.638	0.320	0.500	0.367	0.307	0.691
Hyperactive	29	0.638	0.189	0.251	0.121	0.074	0.531
8	**Regularization *Elastic Net* (*alpha* = 0.3)**						
Hypoactive	25	0.600	0.290	0.444	0.372	0.270	0.669
Hyperactive	27	0.600	0.145	0.125	0.067	0.056	0.467
9	**Regularization *Elastic Net* (*alpha* = 0.5)**						
Hypoactive	15	0.562	0.313	0.556	0.400	0.239	0.646
Hyperactive	20	0.562	0.138	0.168	0.059	0.058	0.492
10	**Regularization *Elastic Net* (*alpha* = 0.7)**						
Hypoactive	11	0.592	0.333	0.667	0.444	0.246	0.660
Hyperactive	12	0.592	0.137	0.125	0.057	0.054	0.513
11	**Regularization *Elastic Net* (*alpha* = 0.9)**						
Hypoactive	7	0.569	0.265	0.450	0.346	0.255	0.652
Hyperactive	8	0.569	0.165	0.217	0.103	0.061	0.594

Furthermore, it is worth emphasizing that model 7 exhibits a relatively reasonable AUC-ROC value, with 0.691 and 0.531 for the hypoactive and hyperactive classes of the disorder, respectively. A value of AUC above 0.5 indicates that the classifier is capable of classifying the different categories. However, concerning the AUC-PR metric, only values of 0.307 and 0.074 were obtained for the hypoactive and hyperactive categories, respectively. This highlights that the predictive model faced significant difficulty in classifying hypoactive cases and, more specifically, hyperactive cases with the test data, where it showed very low values, indicating its limited classification ability for this subtype. This suggests that the model is less sensitive and more specific, capturing more information related to the absence of the syndrome.

These results are explained by the low representation of these classes in the data, as out of 438 records, only 12.90% of cases pertain to the hypoactive subtype, and only 2.30% to the hyperactive subtype. In the context of limited data, there may also be information gaps regarding the combinations between different classes.

It is important to note that when fitting the multinomial logistic regression model with the `nnet` package, it is common to observe that the same independent variables are used for all categories of the response variable. The artificial neural network package models the relationships between the independent variables and the various categories of the response variable. During training, all predictors are considered relative to all categories of the dependent variable, resulting in the use of the same independent variables for all classes.

However, the regularization approach of the `glmnet` package identifies only some dummy variables as relevant for certain categories [16]. This reflects that the categorical variable as a whole has some impact on the outcome, but not all categories have the same effect. The discrepancy arises from the methodology adopted by the packages.

Although there was not a substantial decrease in the number of initial variables, the model's dimensionality was reduced by only two predictors. The variables excluded for not demonstrating significance in this dataset, according to the *Elastic Net* method of model 7, correspond to pCO_2 and the pharmacological group of antihypertensive drugs.

3.1 Most Contributing Variables

After selecting the best model, there is a pressing need to interpret the coefficient values. The results returned by the model are in logarithmic form, adding complexity to their comprehension. To simplify the interpretation of the coefficients concerning the dependent variable "Outcome" related to the diagnosis of *delirium*, a transformation was performed by exponentiating the model's coefficients by 7, resulting in odds ratios (OR). This conversion, as represented in Table 2, aims to interpret the variable effects more accessible and understandable within the model context.

For the Hypoactive category compared to the reference category (Absent), the result indicates that for a one-unit change in blood pH, the probability of a

patient developing the hypoactive subtype of *delirium* increases by 1.351 times. However, it is noteworthy that an increase of one day in the attribute related to hospitalization days shows a 9.044 times increase in the occurrence of this subtype. Regarding categorical predictors such as alcoholism, analgesic drugs, and cardiotonics, individuals exhibiting these habits and consuming these medications are shown to have 2.203 times, 2.018 times, and 1.806 times, respectively, greater probability of exhibiting this phenotype compared to those without the habit or not taking these pharmacological groups.

On the other hand, for male individuals, the probability of developing this subtype decreases by about 48.8% (0.512 times), meaning that the reference class, the absence of the disorder, is more likely. If there is a one-unit change in pO_2, the probability of the hypoactive subtype occurring decreases by 72.2%. In other words, the likelihood of an individual developing this subtype is 0.278 times lower when pO_2 increases by one unit, while the other attributes remain constant.

Similarly, for the Hyperactive category, in contrast to the Absent class, the result shows that a one-unit increase in PCR increases the probability of an individual exhibiting the hyperactive subtype by 14.856 times. A one-unit change in age implies that the probability of this subtype occurring increases by 9.650 times compared to the absence of the disorder. However, a one-unit change in glucose and sodium values reveals a decrease of 84% and 79.2%, respectively, in the development of the hyperactive phenotype.

In summary, it can be concluded that the best predictive model implemented for multinomial logistic regression leaned towards the *Elastic Net* technique with an *alpha* of 0.1. This model will lead to the achievement of the main goal of this dissertation, which is a web application capable of predicting *delirium* diagnosis based on 26 independent predictors (which includes 27 dummy variables for hypoactive and 29 for hyperactive) and the utilized hyperparameters, corresponding to a *lambda* value of 0.1.

3.2 Web Application

As previously verified, the multinomial logistic regression model that demonstrated superior predictive capability was Model 7, which implemented an *embedded* method, *Elastic Net* with an *alpha* of 0.1. Therefore, considering the goal of this study, we proceeded to build a web application using the `shiny` package, aiming to facilitate the use of this predictive model to estimate the diagnosis of *delirium* in the emergency department.

This tool allows real-time access from any device to the prediction estimates. The estimates are generated by filling out the fields in the application form, which correspond to the 26 independent variables selected previously, as mentioned. In this way, it aims to efficiently calculate the probability of the occurrence of one of the subtypes of *delirium* in individuals admitted to the emergency department, assisting healthcare professionals in the diagnostic process.

Figure 2 represents the form developed for data collection to be filled out by the user, which is included in the "Form" section. To obtain the prediction of the

Table 2. Interpretation of the estimated coefficients of the independent variables in the selected model (Model 7) for each category of the outcome variable, according to the odds ratios.

Logit	Variables	Coef.	OR	Logit	Variables	Coef.	OR
0	**(The reference category corresponds to Absent.)**						
1 Hypoactive	Intercept	0.146	1.157	**2 Hyperactive**	Intercept	−0.942	0.390
	OriginHome	0.007	1.007		OriginHome	−0.606	0.546
	OriginInter-Hospitalar	−0.026	0.974		OriginIntra-Hospitalar	0.288	1.334
	ConditionUDC1	−0.217	0.805		OriginLar	0.302	1.353
	Age	0.087	1.091		ConditionUCI	−0.314	0.731
	GenderMale	−0.669	0.512		ConditionUDC1	0.591	1.806
	Hospitalization_Days	2.021	9.044		ConditionUDC2	−0.004	0.996
	Diagn_GroupGastrointestinal	−0.045	0.956		Age	2.267	9.650
	Diagn_GroupHemato-oncologic	−0.044	0.957		GenderMale	0.404	1.498
	Diagn_GroupNeurologic	−0.276	0.759		Diagn_GroupHemato-oncologic	−0.143	0.867
	Diagn_GroupOther	0.087	1.091		Diagn_GroupGeniturinary	−0.009	0.991
	Diagn_GroupRespiratory	0.075	1.078		Diagn_GroupNeurologic	0.058	1.060
	Diagn_GroupToxi	0.544	1.723		Diagn_GroupOther	−0.348	0.706
	SIRS	−0.493	0.611		SIRS	0.545	1.725
	Glicose	0.325	1.384		Glicose	−1.833	0.160
	Creatinine	−0.099	0.906		Sodium	−1.572	0.208
	PCR	−0.764	0.466		Urea	0.511	1.668
	pH	0.301	1.351		Creatinine	0.451	1.570
	Ionized_calcium	0.046	1.047		PCR	2.698	14.856
	pO2	−1.279	0.278		pH	−0.041	0.960
	Antidyslipidemics	0.481	1.618		Ionized_calcium	−0.608	0.544
	Anxiolytics	0.405	1.499		pO2	1.346	3.844
	Analgesics	0.702	2.018		HCO3	−0.310	0.733
	Antipsicotics	0.183	1.201		Antidyslipidemics	−0.221	0.802
	Anticoagulants	0.327	1.387		Anxiolytics	−0.327	0.721
	Cardiotonics	0.591	1.806		Analgesics	−0.578	0.561
	Other_Med	0.400	1.492		Antidepressants	0.164	1.178
	AlcoolicYes	0.790	2.203		Corticosteroids	0.269	1.309
					Cardiotonics	−0.480	0.619
					AlcoolicYes	−0.332	0.718

occurrence of the disorder subtypes, it was essential to include the same variables as those in the training model, corresponding to 26 independent variables. Note that the application displays default values that need to be modified by the user to make the prediction. Each variable is restricted to a certain range of values, similar to those in the training dataset. Thus, the user cannot input values that exceed the limits otherwise, it would hinder the use of the application based on the multinomial logistic regression model for prediction.

A tabular section was created considering how crucial it is to verify the data entered in the form by the user, as a small change can significantly influence the obtained results, aiming to mitigate errors. The prediction result will only be generated when the "Estimate" button is pressed. The prediction result obtained through the implementation of the multinomial logistic regression model is returned as 0, 1, or 2 according to the corresponding category of the response variable (diagnosis outcome of *delirium*). To enhance the readability and intuitiveness of the result interpretation, values have been categorized to prevent confusion for the user.

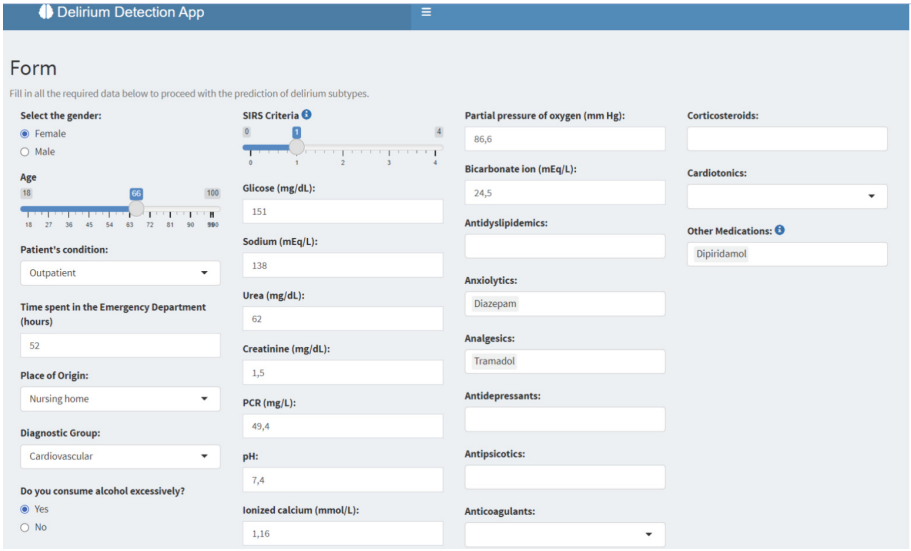

Fig. 2. The web application form for collecting data from individuals in the emergency department.

4 Conclusion

This study demonstrated the importance of data preparation along with data modulation to predict a previous diagnosis outcome. During the modelling process, the variable selection was performed to reduce the model's complexity and facilitate data understanding. Among the various techniques implemented for both the wrapper and *embedded* methods, the technique that produced the best performance for the predictive model was *Elastic Net* with an *alpha* value of 0.1. The model removed the variables pCO_2 and the group of antihypertensive drugs, as they did not prove significant in this dataset. The exclusion of this pharmacological group raised some perplexity; however, it is concluded that it may result from its low representativeness in the population sample.

Specifically, this *embedded* method allowed the selection of 27 dummy variables for the hypoactive subtype, achieving an AUC-PR of 0.307 and an AUC-ROC value of 0.691. The most contributive variables for its development include the length of hospitalization days in the emergency department (ED), alcoholism, drugs from the analgesic and cardiotonic groups, as well as the diagnostic group related to toxicity and drugs. The analgesic group consists of codeine, morphine, and tramadol, while the cardiotonic group includes digoxin. The fact that these pharmacological groups prove to be the most influential for predicting the hypoactive subtype is consistent with expectations since analgesics showed 14 records, of which 6 exhibited this phenotype (42.86%). In turn, from the population sample of 28 individuals regularly taking cardiotonic glycosides, 35.71% (10 people) exhibit hypoactive *delirium*. Due to these statistics, it is reasonable

to conclude that these variables have a significant impact on predicting this subtype, as the probability of such cases occurring is substantially higher when one of these drugs is present.

Regarding the hyperactive subtype, the selected model determined 29 dummy variables as the most relevant, obtaining an AUC-PR value of 0.074 and an AUC-ROC value of 0.531. It is speculated that the low values associated primarily with the hyperactive subtype are due to its low representation in the data, making the model unable to accurately predict this class. Among the variables that have the most impact on predicting this subtype are PCR, age, pO_2, SIRS criteria, and the place of origin in the ED, specifically UDC1 (patients assigned the yellow bracelet).

Although the predictive capacity is reasonable, it is assumed to stem from the small number of observations and a high number of predictors. Consequently, the classifier faces increased difficulty in detecting patterns in the data due to gaps in combinations between different categories of the response variable. If the number of records were substantially higher, it is postulated that the prediction process would result in a higher accuracy rate.

After the process of selecting the best predictive model and adjusting the hyperparameters accordingly, the development of the web application using the `shiny` package was undertaken. This application implements the selected model and is built on a total of 26 independent variables. While there is potential for enhancement, this predictive model could prove valuable to assist healthcare professionals in diagnosing *delirium* and its respective subtypes in the context of the emergency department through an intuitive and easily interpretable interface.

5 Future Work

In future work, exploring alternative machine learning algorithms and implementing different variable selection methods is suggested to enhance the predictive capability of the model for the imbalanced and multicategorical variable. Integrating the predictive model into the hospital system would automate the process, offering healthcare professionals timely information about an individual's probability of developing *delirium* subtypes upon admission to the emergency department, potentially reducing underdiagnosed cases and improving patient outcomes. However, to substantially augment the predictive capacity of the model, it is imperative to augment the volume of records in the training set, focusing particularly on the hypoactive and hyperactive subtypes of *delirium*.

References

1. Field, A.: Discovering Statistics Using R. SAGE, Thousands Oaks (2012)
2. Adamis, D., McCarthy, G., O'Mahony, E., Meagher, D.: Motor disturbances in elderly medical inpatients and their relationship to delirium. J. Geriatr. Psychiatry Neurol. **30**(4), 214–219 (2017)

3. Ai, M.T., Sumiati, S., Rosalina, V.: Predictive model for heart disease diagnosis based on multinomial logistic regression. Inf. Technol. Control **50**(2), 308–318 (2021)
4. Ali, S., et al.: Insight into delirium. Innov. Clin. Neurosci. **8**(10), 25–34 (2011)
5. Asken, M.J., Grossman, D., Christensen, L.W.: American psychiatric association. diagnostic and statistical manual of mental disorders. American psychiatric publishing, Arlington (2013)
6. Archibald, H.C., Tuddenham, R.D.: persistent stress reaction after combat: a 20-year follow-up. Arch. Gener. Psy. Therapy **45**(10), 2317–2325 (2007)
7. Assefa, M.T., Chekol, W.B., Melesse, D.Y., Nigatu, Y.A., Bizuneh, Y.B.: Incidence and risk factors of emergence delirium in elderly patients after general or spinal anesthesia for both elective and emergency surgery. Ann. Med. Surg. **84**, 104959 (2022)
8. Bellelli, G., Speciale, S., Barisione, E., Trabucchi, M.: Delirium subtypes and 1-year mortality among elderly patients discharged from a post-acute rehabilitation facility. J. Gerontol. Series A **62**, 1182–1183 (2007)
9. Boateng, E.Y., Abaye, D.A., Boateng, E.Y., Abaye, D.A.: A review of the logistic regression model with emphasis on medical research. J. Data Anal. Inf. Process. **7**, 190–207 (2019)
10. Boettger, S., Breitbart, W.: Phenomenology of the subtypes of delirium: phenomenological differences between hyperactive and hypoactive delirium. Palliat. Support. Care **9**, 129–135 (2011)
11. Breslow, N.E.: Generalized linear models: checking assumptions and strengthening conclusions. Statistica Applicata **8**(1), 23–41 (1996)
12. Bulic, D., et al.: Cognitive and psychosocial outcomes of mechanically ventilated intensive care patients with and without delirium. Ann. Intensive Care **10**, 1–10 (2020)
13. Coelho, A., Braga, A.C., Mariz, J.: GLM's in data science as a tool in the prediction of delirium. In: Optimization. Learning Algorithms and Applications, pp. 591–606. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-53025-8_40
14. Davis, D.H., et al.: Delirium is a strong risk factor for dementia in the oldest-old: a population-based cohort study. Brain **135**(9), 2809–2816 (2012)
15. Franco, J., et al.: Three core domains of delirium validated using exploratory and confirmatory factor analyses. Psychosomatics **54**(3), 227–238 (2013)
16. Friedman, J., et al.: Package 'glmnet'. CRAN R Reposit. **595** (2021)
17. Garson, G.D.: Logistic Regression: Binary and Multinomial, 2014th edn. Asheboro (2014)
18. Harrington, C.J., Vardi, K.: Delirium: presentation, epidemiology, and diagnostic evaluation (part 1). RI Med. J. **97**(6), 18–23 (2014)
19. Hosie, A., Davidson, P.M., Agar, M., Sanderson, C.R., Phillips, J.: Delirium prevalence, incidence, and implications for screening in specialist palliative care inpatient settings: a systematic review. Palliat. Med. **27**(6), 486–498 (2013)
20. Huson, K., Stolee, P., Pearce, N., Bradfield, C., Heckman, G.A.: Examining the hospital elder life program in a rehabilitation setting: a pilot feasibility study. BMC Geriatr. **16**(1), 1–12 (2016)
21. Inouye, S.K., Westendorp, R.G., Saczynski, J.S.: Delirium in elderly people. Lancet **383**(9920), 911–922 (2014)
22. Karpatne, A., et al.: Theory-guided data science: a new paradigm for scientific discovery from data. IEEE Trans. Knowl. Data Eng. **29**(10), 2318–2331 (2017)
23. Kwak, C., Clayton-Matthews, A.: Multinomial logistic regression. Nurs. Res. **51**(6), 404–410 (2002)

24. Lipowski, Z.J.: Delirium: how its concept has developed. Int. Psychogeriatr. **3**(2), 115–120 (1991)
25. Maldonado, J.R.: Delirium pathophysiology: an updated hypothesis of the etiology of acute brain failure. Int. J. Geriatr. Psychiatry **33**(11), 1428–1457 (2018)
26. Marcantonio, E.R., et al.: Outcomes of older people admitted to postacute facilities with delirium. J. Am. Geriatr. Soc. **53**(6), 963–969 (2005)
27. Milewska, A.J., Jankowska, D., Więsak, T., Acacio, B., Milewski, R.: The application of multinomial logistic regression models for the assessment of parameters of oocytes and embryos quality in predicting pregnancy and miscarriage. Stud. Logic Gram. Rhetoric **51**, 7–18 (2017)
28. Morandi, A., et al.: Clinical features associated with delirium motor subtypes in older inpatients: results of a multicenter study. Am. J. Geriatr. Psychiatry **25**(10), 1064–1071 (2017)
29. Rabinowitz, T.: Delirium: an important (but often unrecognized) clinical syndrome. Curr. Psychiatry Rep. **4**(3), 202–208 (2002)
30. Ripley, B., et al.: Package 'mass'. Cran R **538**, 113–120 (2013)
31. Ripley, B., Venables, W.: Package 'nnet': feed-forward neural networks and multinomial log-linear models (2023)
32. Robinson, T.N., Raeburn, C.D., Tran, Z.V., Brenner, L.A., Moss, M.: Motor subtypes of postoperative delirium in older adults. Arch. Surg. **146**, 295–300 (2011)
33. Ruano, L., et al.: Prevalence and causes of cognitive impairment and dementia in a population-based cohort from Northern Portugal. Am. J. Alzheimer's Dis. Other Dementias® **34**(1), 49–56 (2019
34. Slor, C.J., et al.: Delirium motor subtypes in elderly hip fracture patients: risk factors, outcomes and longitudinal stability. J. Psychosomatic Res. **74**, 444–449 (2013)
35. Yang, F.M., et al.: Phenomenological subtypes of delirium in older persons: patterns, prevalence, and prognosis. Psychosomatics **50**, 248–254 (2009)
36. Zamoscik, K., Godbold, R., Freeman, P.: Intensive care nurses' experiences and perceptions of delirium and delirium care. Intensive Crit. Care Nurs. **40**, 94–100 (2017)