# Applying Prediction Models Based on Ensemble Machine Learning Algorithms to Estimate Resource Requirements at Healthcare Centers

**Carlos Hernández and Paola Leal**
Departamento de Procesos Industriales
Universidad Católica de Temuco
Temuco, Chile
carlos.hernandez.zavala@uct.cl, pleal@uct.cl

## Abstract

Either a high complexity hospital or a smaller clinic, healthcare centers have to withstand the constant pressure of the incoming flow of new patients. While some patients require a simple medical procedure, others will need further examination and probably have remain in observation for some time. This situation is particularly complicated in times of sanitary crisis. Since the infrastructure, supplies, and medical staff are limited resources, there is a real need for utilizing them efficiently. This research is focused on the use of ensemble machine learning algorithms to develop models for predicting the destination of patients who are discharge after a stay at an intensive care unit (ICU). The investigation was carried out following a 4-phase methodology: analysis, design, development, and validation. During the analysis, an extensive review and preprocessing of patient records collected from a public hospital was carried out. Then, during the design several ensemble machine learning algorithms were compared and selected for the investigation. To name a few: Linear Regression, Decision Tree, Stacking, Bagging, and Random Forest. The following phases, development and validation were completed using data processing software. In all models proposed here, instead of a simple hold-out, a 10-fold cross-validation scheme was applied. For the purposes of this research, twenty thousand patient records collected in 2020 were considered. The complete dataset was split in two subsets. One subset for training and test with 80% of the data and another dataset for validation with the remaining 20%. During the development of the models, only data for training and for test were used. The validation data were used only to measure the models performance with unseen data. Results revealed that regardless the size of the training and test dataset, there was a notorious consistency in the correct prediction rates. The proposed ensemble scheme made of three base learner plus a meta algorithm, systematically leaded to correct prediction rates close to 82%. In conclusion, the proposed models proved that, with based on the existing data, high rates of correct prediction can be achieved when an ensemble scheme is used. In this case, with a reasonable certainty, it was possible to predict whether a patient was going to be referred to another unit or sent home after his or her stay at ICU.

## Keywords
Machine Learning, Ensemble Algorithms, Artificial Neural Network, Support Vector Machine, Healthcare Center Infrastructure.

## 1. Introduction
Delivering high quality medical care and services requires collecting, preprocessing and analyzing large amounts of heterogeneous data to extract valuable information that must presented in intelligible and user-friendly format. The availability of new technologies and the adoption of concepts such as internet of things (IoT), intelligent environments (SE), and smart devices are becoming an enormous contribution to different medical areas (Ivanovic, 2023).

Artificial intelligence (AI) is contributing decisively to consolidate the so called P4 Medicine (Predictive, Personalized, Preventive and Participatory) by means of adding new data processing capabilities in diverse areas such as DNA sequencing, electronic medical records, and the environmental variables to which people have been exposed which can be understood by analyzing computer tomography images, electroencephalograms, text in electronic medical records, pharmacological data, etc. (Ruiz and Velasquez, 2023).

In recent years, new technologies have been developed and have been used in a rage of medical applications. A good example is the prevention of diabetes using machine learning algorithms (ML) such as K-nearest neighbor (KNN), support vector machine (SVM), decision trees (DT), Naive Bayes (NB), and logistic regression (LR) (Goyal et al., 2023). Another good example is the prediction heart related diseases by means of analyzing data with deep learning techniques by means of comparing the results produced by different algorithms such as SVM, Naive Bayes, KNN algorithm (Harlapur and Handur, 2023). The detection and identification of patterns in medical data sources allow medical doctors and data scientists make timely decisions with based in non-trivial predictions that can be crucial in cases of cancer or heart diseases. In these cases, the combined result of several algorithms can help improve the performance and accuracy (Al-Ahdal and Chawla, 2023). Recently, COVID-19 sanitary crisis catalyzed the used of machine learning algorithms and other forms of artificial intelligence to predict, to diagnose, and to detect positive patients (Kejriwal and Rajagopalan, 2023).

It also possible to find applications for managing clinical processes and for supporting decision makers in the improvement of public policies. An example is the relation between the exposure to heavy metals and coronary heart diseases (CHD) using data from the National Survey of the US Health and Nutrition, in which five ML models were used to identify CHD derived from the exposure to heavy metal and 11 discrimination characteristics were used to test the strength of the models (Li et al., 2023). Another example is the prediction of patient non-attendance to medical appointments by means of an indicator for the non-attendance risk that can be estimated using patients' medical records (Valero-Bover et al., 2022). The improvement in the management of care services for cancer survivors with the help of digital health technologies (DHT) supported by artificial intelligence (Pan et al., 2022), the use of clinical artificial intelligence (cAI) models to predict the stay of patients at intensive care units (ICU), and the ability to predict mortality among discharged patients even a year after leaving ICU up to more than a year, are important tools that can be used to improve the availability of hospital beds at ICUs (Ishii et al., 2023).

Consequently, the analysis and use of large amounts of data to extract valuable information can help healthcare organizations improve their performance by means of reducing the cost of data processing during the process of generating prospective results (Srivastava et al., 2023).

This investigation proposed models based on ensemble machine learning algorithms to help organizations manage critical infrastructure such as medical staff, supplies and beds at an ICU by means of predicting whether patients are sent home or referred to another hospital unit after being discharged from ICU.

### 1.1 Objective

To develop models based on ensemble machine learning algorithms to estimate resource requirements at healthcare centers by means of predicting the destination of discharged ICU patients.

## 2. Literature Review

### 2.1 Machine learning

Machine learning is usually referred as the branch of artificial intelligence (AI) that uses algorithms to find patterns and to learn from datasets through experience. There several types of machine learning algorithms: supervised, unsupervised, and reinforcement algorithms. In supervised learning, the training is carried out using labelled datasets. This means that the class or the value to be predicted is included in the dataset so it can be used for training. In the case of unsupervised learning, instead, the desired class is not known.

### 2.2 Classification

In machine learning there several important task: classification, regression, and forecasting. Classification can be understood as the determination of the class, a nominal value, in an unseen dataset using a previously trained model. In a regression problem, instead, the objective is the estimation of a numeric value using independent variable. On the other hand, in a forecasting problem time series are used to predict future values.

### 2.3 Hold out and cross-validation

Holding out implies the splitting up of a dataset into a set for training and another for testing. The test dataset is employed to assess the performance of the classification model with unseen data. Usually the split up proportion is 80% for training and 20% for testing.

On the other hand, cross-validation is the random split up of a dataset into k folds. During the model building, k-1 folds are employed for training while the left one is used to test model's performance. Training and testing are repeated iteratively k times until all folds have been used for testing (Figure 1). The goal is to minimize the risk of overfitting that can happen when holding out. In the case of cross-validation, each iteration produces different results because the folds for training and for testing have been interchanged. These k results are finally averaged.

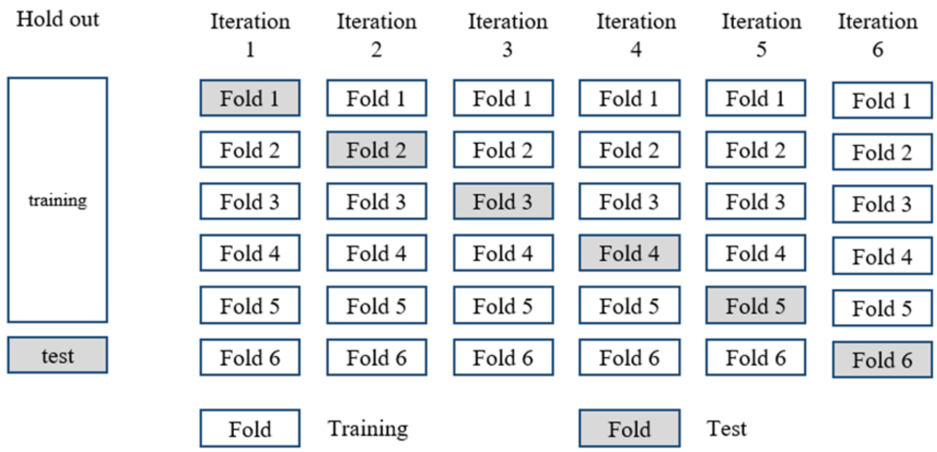


Figure 1. Hold out and cross-validation (k=6)

## 2.4 Overfitting and generalization

Overfitting occurs when a model learns well from the training dataset but it does not have a good performance when tested with an unseen dataset. In such situation, it is said that the model cannot generalize. This might happen due to the incorporation of many details from the training data that will not be easily found in new data (Figure 2).

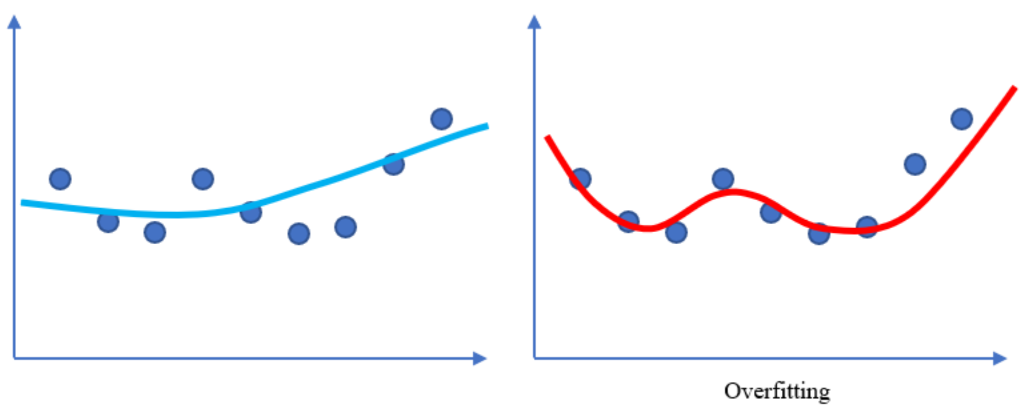


Figure 2. Overfitting

## 2.5 Replication

Replication is repetition of an experiment under similar conditions to estimate the variability of the results. When using cross-validation, the dataset partitioning into k folds depends on a specific seed number (Figure 3). Since different seed numbers produce different folds, the results of the training and test are different too. By means of replicating the experiments with random seeds each time, it would possible to obtain several test results from which the mean and the standard deviation can be estimated and analyzed afterwards. Thus minimizing the effect of an unfortunate partitioning.

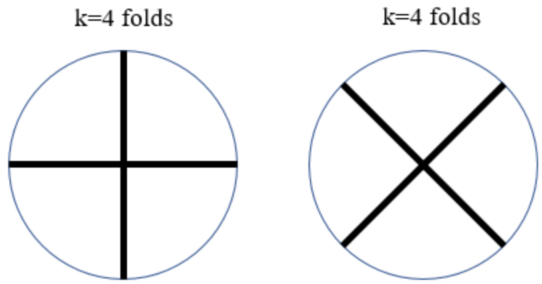


Figure 3. Different folds in cross-validation (k=4)

## 2.6 Meta-learning

Meta-learning, or learning to learn, is the use of learning algorithms to learn from the prediction of other learning algorithms. The underlying idea is to combine the predictions of several machine learning algorithms to make new predictions.

## 2.7 Ensemble algorithms

Ensemble machine learning algorithms are multi-level structures to carry out learning tasks. In the simplest configuration there is a meta-algorithm (level 1) that leans from the predictions made by the base algorithms (level 0). The ensemble usually can predict better than any of its single algorithms. The stacking is one of the most widely used ensemble schemes (Figure 4).

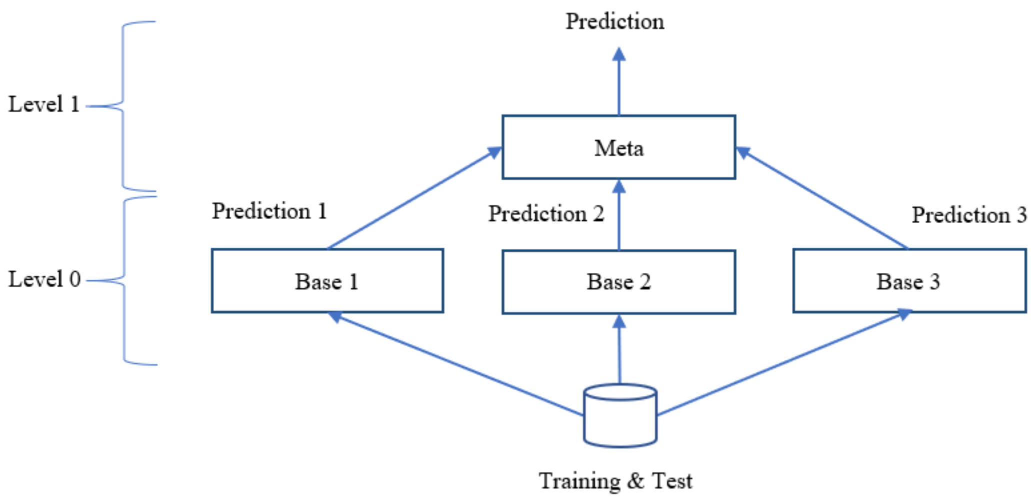


Figure 4. Ensemble algorithm

# 3. Methods

This investigation is carried out following a classic 4-phase methodology: analysis, design, construction, and validation (Figure 5).

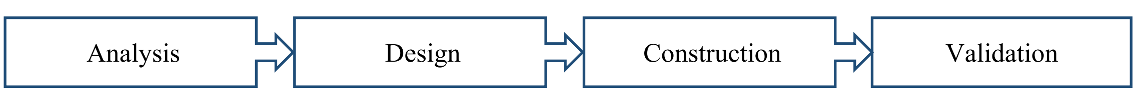


Figure 5. Four-phase methodology

### 3.1 Analysis

During the phase of analysis, a full review and preprocessing of ICU records was completed. Even though counting with more than 70,000 records collected during 2020, for the purposes of this investigation only 20,000 were considered. Each record contains data of a patient who received medical care at ICU. Every time a patient arrived at ICU a new record was appended to the database. The data contained in a single record contained several fields, from the arrival time to the health insurance type. The fields considered in this work are the following: age, gender, reason for visit, arrival mode, source of admission, presence of trauma, alcohol test result, discharge status, medical specialist, typo of urgency, and post-discharge destination. Being the latter, the target class to be predicted (Table 1).

Table 1. Patient record selected fields

| Field | Description |
|---|---|
| Age | The age of patient at date of admission. |
| Gender | The gender of a patient |
| Reason for visit | The reason for the patient requiring involvement with ICU medical staff. |
| Arrival mode | The principal means by which a patient arrives at ICU. |
| Source of admission | Place from where patients came. |
| Trauma | Presence of trauma. |
| Alcohol test | Test to detect the presence of alcohol in the patient. |
| Discharge status | Identifies the condition of the patient at the conclusion of a health care process. |
| Medical specialist | Doctor focused on a defined group of patients, diseases, skills, or philosophy. |
| Urgency type | Indicates the classification of the patient healthcare. |
| Destination | The destination of the patient after leaving ICU. |

The preprocessing of data revealed that most of the ICU patients were sent home after being discharged (70%) and that only 30% were referred to another hospital unit for further treatment (Table 2). A summary of the number of ICU post-discharge destinations according to medical specialties Table 3.

Table 2. ICU post-discharge destination per medical specialty

| Medical specialties | Hospitalization | Home | Total records |
|---|---|---|---|
| General surgery | 9,770 | 23,203 | 32,973 |
| Midwifery | 2,938 | 804 | 3,742 |
| Obstetrics and gynecology | 2,079 | 8,585 | 10,664 |
| Pediatrics | 1,518 | 7,711 | 9,229 |
| Neurologist | 1,296 | 1,109 | 2,405 |
| Emergency medicine | 923 | 2,193 | 3,116 |
| General medicine | 847 | 1,996 | 2,843 |
| Adult Trauma | 528 | 3,059 | 3,587 |
| Neurosurgery | 294 | 428 | 722 |
| Internal medicine | 246 | 286 | 532 |
| Psychiatry | 142 | 43 | 185 |
| Pediatric surgery | 128 | 334 | 462 |
| Odontology | 37 | 144 | 181 |
| Urology | 15 | 31 | 46 |
| Gynecology | 1 | 10 | 11 |
| **Total records** | **20,762** | **49,936** | **70,698** |

It is always interesting to verify one of the recurrent proportion in engineering, the principle of Pareto. In this case, Table 3 and Figure 6 show the number of hospitalizations associated with the available medical specialties. It can be seen that roughly 70% of all hospitalizations corresponds to roughly 20% of all specialties.

Table 3. Specialties (%) v/s Hospitalizations (%)

| Medical specialties | Specialties | Hospitalizations |
|---|---|---|
| General surgery | 7 % | 47 % |
| Midwifery | 13 % | 61 % |
| Obstetrics and gynecology | 20 % | 71 % |
| Pediatrics | 27 % | 79 % |
| Neurologist | 33 % | 85 % |
| Emergency medicine | 40 % | 89 % |
| General medicine | 47 % | 93 % |
| Adult Trauma | 53 % | 96 % |
| Neurosurgery | 60 % | 97 % |
| Internal medicine | 67 % | 98 % |
| Psychiatry | 73 % | 99 % |
| Pediatric surgery | 80 % | 100 % |
| Odontology | 87 % | 100 % |
| Urology | 93 % | 100 % |
| Gynecology | 100 % | 100 % |

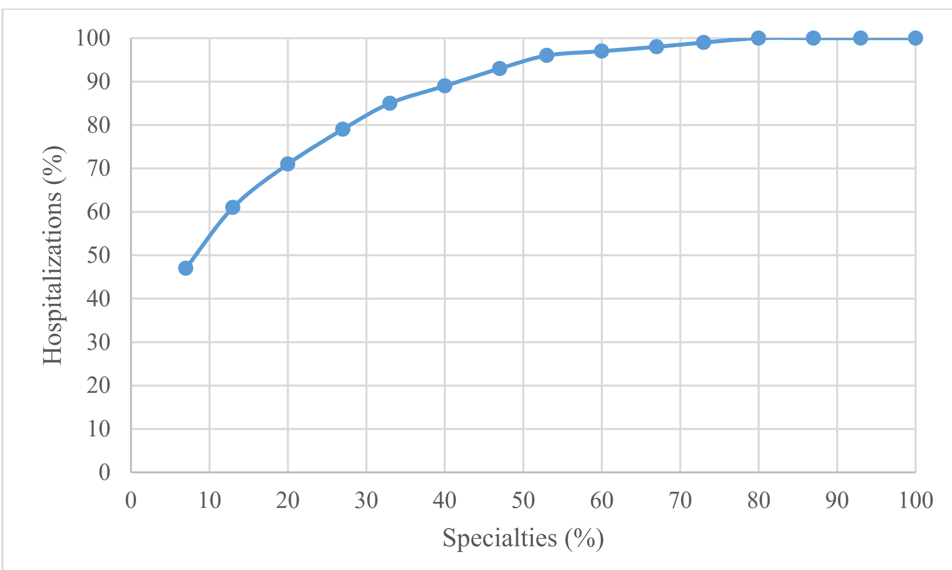


Figure 6. Medical specialties (%) v/s hospitalizations (%)

### 3.2 Design

ICU patient record database can be understood as a large matrix whose rows represent single records and columns represent fields or attributes. The challenge is, by using a limited number of attributes, to classify each record into a certain class value. In this case, the target class *post-discharge destination* had only two possible values: home and hospitalization.

During the design, four datasets of different sizes were prepared to quantify influence of the number of records on the performance of the prediction models (Table 4).

Table 4. Dataset creation

| Dataset | Number of fields | Number of records |
|---|---|---|
| DS-05 | 11 | 5,000 |
| DS-10 | 11 | 10,000 |
| DS-15 | 11 | 15,000 |
| DS-20 | 11 | 20,000 |

Each dataset was divided to create two subsets in a proportion of 80% and 20% respectively. The first dataset contained records for training and test with 80% of data, whereas the second dataset contained records for validation only (Figure 7 and Table 5).

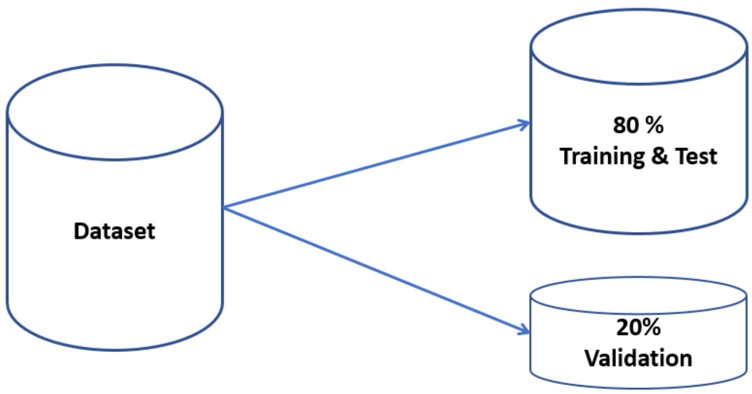


Figure 7. Dataset split up

Table 5. Datasets for training and test, and for validation

| Dataset | Records for training and test | Records for validation |
|---|---|---|
| DS-05 | 4,000 | 1,000 |
| DS-10 | 8,000 | 2,000 |
| DS-15 | 12,000 | 3,000 |
| DS-20 | 16,000 | 4,000 |

The validation dataset is used to confirm whether the proposed prediction model, prepared with the dataset for training and test, is able to generalized properly when predicting with unknown validation data.
In this work, an ensemble scheme with three base learners plus a meta learner was implemented. Based on preliminary trial and error experiments, four machine learning algorithms were selected. Namely, support vector machine, logistic regression, decision tree, and nearest neighbor k=1 (Table 6). The ensemble scheme selected is the well-known stacking, which has been matter of study in previous investigations (Divina et al., 2018).

Table 6. Ensemble algorithm's configuration

| Ensemble | Meta learner (level 1) | Base learner (level 0) |
|---|---|---|
| Stacking | Support vector machine | Logistic regression |
| | | Decision tree |
| | | Nearest neighbor k=1 |

The resulting models are compared by means of the corresponding correlation coefficient, MAE (mean absolute error), and RMSE (root mean squared error). These are common measures used for comparing models (Weijie and Yanmin, 2018).

When dealing with classification problems it is important to keep in mind the class balance. In the case of heavily imbalanced datasets, some authors recommend the inclusion of additional performance metrics such as Precision-Recall curves along with ROC AUC. Saito and Rehmsmeier offered a deeper analysis of these curves (Saito and Rehmsmeier, 2016).

The resulting classification models wear compared by means of the percentage of correct predictions with the validation dataset. Additionally, curves Precision-Recall and the area under the curve ROC (ROC AUC) were considered during the comparison too. Davis and Goadrich presented an interesting analysis on this subject (Davis and Goadrich, 2006).

### 3.3 Construction

The objective was to build models capable of predicting whether a ICU discharged patient was sent home or referred to another hospital unit by means of applying an ensemble machine learning algorithm. In principle, 11 fields for record were considered.

To help overcome ambiguities caused by the low number of fields (attributes), the field *Reason for visit* was converted from a string of character to a vector of words. Generating in this way a large number of new attributes to provided additional pieces of information.

All the models were developed using WEKA 3.9.6 (Witten et al., 2017). Initially, all models were trained and tested applying a cross-validation scheme of k=10 folds. (Table 7)

Table 7. Predictions with training and test dataset and cross-validation k=10

| | | | Weighted average | | |
|---|---|---|---|---|---|
| **Model** | **Records** | **Correct prediction (%)** | **Precision** | **Recall** | **ROC AUC** |
| M-04 | 4,000 | 82.4 % | 0.814 | 0.824 | 0.819 |
| M-08 | 8,000 | 82.4 % | 0.814 | 0.824 | 0.803 |
| M-12 | 12,000 | 82.8 % | 0.818 | 0.828 | 0.809 |
| M-16 | 16,000 | 82.8 % | 0.818 | 0.828 | 0.807 |

Although cross-validating helped reduce the risk of overfitting, the effect of the fold partitioning remained. This fact was already analyzed by Powers and Atyabi (Powers and Atyabi, 2012). Replicating experiments could mitigate this issue by means of using different folds in each iteration. For the purposes of this research 10 replications were run, which means that each model was trained and tested 100 times. The results of such strategy showed can be better understand with the help of the standard deviation of the prediction rate during the replications (Table 8).

Table 8. Predictions with training and test dataset, cross-validation k=10, and 10 replications

| Model | Average correct predictions | Standard deviation (10 replications) |
|---|---|---|
| M-04 | 82.7 % | 1.67 |
| M-08 | 82.3 % | 1.20 |
| M-12 | 82.8 % | 0.87 |
| M-16 | 82.8 % | 0.75 |

### 3.4 Validation

The validation of the models was carried out with unknown records from the validation dataset held out during the phase of analysis. This dataset contains unseen records which corresponds to 20% of the data allocated for each model (Table 9).

Table 9. Correct predictions rates with unknown validation data

| | | | Weighted average | | |
|---|---|---|---|---|---|
| **Model** | **Unknown records** | **Correct predictions** | **Precision** | **Recall** | **ROC AUC** |
| M-04 | 1,000 | 78.5 % | 0.771 | 0.785 | 0.766 |
| M-08 | 2,000 | 83.8 % | 0.828 | 0.838 | 0.816 |
| M-12 | 3,000 | 82.4 % | 0.816 | 0.824 | 0.824 |
| M-16 | 4,000 | 81.1 % | 0.799 | 0.811 | 0.768 |

## 4. Data Collection
Validation results showed consistency between the results obtained with the training and test dataset, and the result with the validation dataset (Table 10).

Table 10. Prediction model comparison

| | Train and test dataset | | Validation dataset | |
|---|---|---|---|---|
| **Model** | **Records** | **Correct predictions** | **Records** | **Correct predictions** |
| M-04 | 4,000 | 82.4 % | 1,000 | 78.5 % |
| M-08 | 8,000 | 82.4 % | 2,000 | 83.8 % |
| M-12 | 12,000 | 82.8 % | 3,000 | 82.4 % |
| M-16 | 16,000 | 82.8 % | 4,000 | 81.1 % |

## 5. Results and Discussion
Contrary to what it might be expected, larger training and test dataset do not lead to higher rates of correct predictions.

### 5.1 Numerical Results
A confusion matrix summarizes the prediction results. While the diagonal contains the number of instances correctly classified, the other cells present incorrect classifications (Table 11).

Table 11. Confusion matrix

| | **Class 1** | **Class 2** |
|---|---|---|
| **Class 1** | Record of class 1 correctly classified | Instance of class 1 misclassified as class 2 |
| **Class 2** | Record of class 2 misclassified as class 1 | Record of class 2 correctly classified |

One common manner to improve the ratio of correct predictions is by means of penalizing the misclassification of instances. This technique was not included in this investigation.
The possibility of misclassification is always present. Data processing techniques like the transformation of the field *Reason for visit* from a string into a vector of words added new pieces of information. However, it still possible that patients with similar records had ended in a different post-discharge destination. All the confusion matrices are presented in Table 12, Table 13, Table 14, and Table 15.

Table 12. Confusion matrix of M-04

| Training and test, cross-validation, and 4,000 records | Unknown validation dataset with 1,000 records |
|---|---|

| | Hospitalization | Home |
|---|---|---|
| Hospitalization | 485 | 513 |
| Home | 191 | 2,811 |

| | Hospitalization | Home |
|---|---|---|
| Hospitalization | 118 | 158 |
| Home | 57 | 667 |

Table 13. Confusion matrix of M-08

| Training and test, cross-validation, and 8,000 records | | |
|---|---|---|
| | **Hospitalization** | **Home** |
| Hospitalization | 999 | 1,021 |
| Home | 388 | 5,592 |

| Unknown validation dataset with 2,000 records | | |
|---|---|---|
| | **Hospitalization** | **Home** |
| Hospitalization | 231 | 236 |
| Home | 88 | 1,445 |

Table 14. Confusion matrix of M-12

| Training and test, cross-validation, and 12,000 records | | |
|---|---|---|
| | **Hospitalization** | **Home** |
| Hospitalization | 1,459 | 1,497 |
| Home | 563 | 8,481 |

| Unknown validation dataset with 3,000 records | | |
|---|---|---|
| | Hospitalization | Home |
| Hospitalization | 392 | 388 |
| Home | 139 | 2,081 |

Table 15. Confusion matrix of M-16

| Training and test, cross-validation, and 16,000 records | | |
|---|---|---|
| | **Hospitalization** | **Home** |
| Hospitalization | 2029 | 1,964 |
| Home | 795 | 11,212 |

| Unknown validation dataset with 4,000 records | | |
|---|---|---|
| | **Hospitalization** | **Home** |
| Hospitalization | 237 | 277 |
| Home | 101 | 1,385 |

## 6. Conclusion

The extensive ICU database allowed researchers to gain a good understanding of a critical hospital unit. The application of adequate tools and innovative techniques can certainly be of great help during the decision making process by means of revealing more improvement opportunities.

It always interesting to see how once and again the principle of Pareto is present. During the preprocessing of data, it was verified that almost 70% of all hospitalization were associated to only 20% of all medical specialties.

Decision making support tools based on techniques and software packages to extract valuable information from large repositories of data, undoubtedly are an alternative that deserve proper attention. Although counting with a rather large database, the proposed models minimized the data requirements and yet delivered promising results. In total, eleven fields were considered to develop the prediction models. The transformation of the field *Reason for visit* from a string to a vector of words provided additional pieces of information that helped develop more complex models.

In general, the use of an ensemble scheme to combine the features of different learning algorithms produces better results than the use single algorithms predicting alone. When working with machine learning schemes, it is advisable to apply a cross-validation to minimize the influence of the dataset split up instead of simply holding out part of the data. Averaging several results will be better than having only one result. Furthermore, running replications help reduce the bias caused by the fold partitioning. Additionally, having a set of unknown data only for validation help confirming whether the models can generalize properly or not.

The experimental result showed that having larger datasets does not necessarily lead to better results in terms of the rate of correct predictions. Either with four or sixteen thousand records, the correct prediction rate fluctuated barely around 82%.

The consistency found in the prediction rates obtained with a cross-validation scheme on training and test data, with a set of unknown validation data, and with set with unknown data proved that the proposed models were able to generalize properly.

Finally, prediction ratios, precision and recall curves, ROC area, and the corresponding confusion matrices suggest that the proposed prediction models based on ensemble scheme can predict the destination of a ICU-discharged patient and, therefore, they could be used to estimate resource requirements at hospitals and clinics.

## References

Al-Ahdal, A. and Chawla, P., A review on machine learning-based patient scanning, visualization, and monitoring, *Lecture Notes in Networks and Systems*, vol. 473, pp. 487–497, 2023.

Davis, J., Goadrich, M., The Relationship Between Precision-Recall and ROC Curves, *Proceedings of the 23rd International Conference on Machine Learning*, 2006.

Divina, F., Gilson, A., Gómez-Vela, F., Garcia, M., and Torres, J., Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting, *Energies 2018*, vol. 11, no. 4, p. 949, 2018.

Goyal, S., Batra, N., and Chhabra, K., Diabetes disease diagnosis using machine learning approach, *Lecture Notes in Networks and Systems*, vol. 473, pp. 229–237, 2023.

Harlapur, N. H., and Handur, V., Text-Based Prediction of Heart Disease Doctor Chatbot Using Machine Learning, *Smart Innovation, Systems and Technologies,* vol. 311, pp. 231–245, 2023.

Ishii, E., Nawa, N., Hashimoto, S., Shigemitsu, H., and Fujiwara, T., Development, validation, and feature extraction of a deep learning model predicting in-hospital mortality using Japan's largest national ICU database: a validation framework for transparent clinical Artificial Intelligence (cAI) development, *Anaesthesia Critical Care and Pain Medicine*, vol. 42, no. 2, 2023.

Ivanovic, M., How artificial intelligent approaches support medical decisions and patients' wellbeing, *Lecture Notes in Networks and Systems*, vol. 556, pp. 181–189, 2023.

Kejriwal, S., and Rajagopalan, N., A technical review on machine learning-based prediction on COVID-19 diagnosis, *Smart Innovation, Systems and Technologies*, vol. 311, 2023.

Li, X., Zhao, Y., Zhang, D., Kuang, L., Huang, H., Chen, W., Fu, X., Wu, Y., Li, T., Zhang, J., Sun, X., and Hu, D., Development of an interpretable machine learning model associated with heavy metals' exposure to identify coronary heart disease among US adults via SHAP: Findings of the US NHANES from 2003 to 2018, *Chemosphere*, vol. 311, 2023.

Pan, L.-C., Wu, X.-R., Lu, Y., Zhang, H.-Q., Zhou, Y.-L., Liu, X., Liu, S.-L., and Yan, Q.-Y., Artificial intelligence empowered digital health technologies in cancer survivorship care: A scoping review, *Asia-Pacific Journal of Oncology Nursing*, vol. 9, no. 12, 2022.

Powers, D., Atyabi, A., The Problem of Cross-Validation: Averaging and Bias, Repetition and Significance. *2012 Spring World Congress on Engineering and Technology*, SCET 2012 - Proceedings. 1-5, 2012.

Ruiz, R. and Velásquez, J., Artificial intelligence for the future of medicine, *Intelligent Systems Reference Library*, vol. 229, 2023.

Saito, T., Rehmsmeier, M., The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets, *PLoS ONE 10(3): e0118432*, 2015.

Srivastava, D., Pandey, H., and Agarwal, A., Complex predictive analysis for health care: a comprehensive review, *Bulletin of Electrical Engineering and Informatics*. vol. 12(1), 521–531, 2023.

Valero-Bover, D., González, P., Carot-Sans, G., Cano, I., Saura, P., Otermin, P., Garcia, C., Gálvez, M., Lupiáñez-Villanueva, F. and Piera-Jiménez, J., Reducing non-attendance in outpatient appointments: predictive model development, validation, and clinical assessment, *BMC Health Services Research*, vol. 22, no. 1, p. 451, 2022.

Weijie, W., Yanmin, L., Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model, *IOP Conference Series: Materials Science and Engineering*, 2018.

Witten, I., Frank, E., Hall, M., and Pal, C., *Data Mining: Practical Machine Learning Tools and Techniques*, 4th Edition, Morgan Kaufmann, Cambridge, 2017.

## Biographies

**Carlos Hernández** is an industrial engineer, consultant, and university professor. He earned Master of Sciences in Engineering and Doctor of Engineering from Technische Universität Braunschweig, Brunswick, Germany. He is the author of several scientific and engineering articles. He has taught lectures in Discrete Event Simulation, Engineering Economics, Corporate Finances, Data Mining and Machine Learning for engineering students. He has developed a professional career working for large multinational companies (PricewaterhouseCoopers, BHP Billiton, and Merck Sharp & Dohme). He also worked as a scientific researcher in the Institut für Produktionsmesstechnick at

TU Braunschweig, Germany. His research interests include manufacturing process simulation, supply chain design and simulation, and machine learning for finances. He is a member of IEOM.

**Paola Leal** is an industrial engineer and university professor. She earned Licentiate Degree in Engineering from Universidad de La Frontera, and Master of Sciences from Universidad Mayor, Temuco, Chile. She has taught lectures in Operations Research, Logistics, and Supply Chain Management for engineering students. During her academic tenure she has been appointed in different management positions and has mentored over a sixty students. Her research interests include logistics, optimization, and supply chain management.