Developing Models Based on Machine Learning Algorithms to Estimate Medical Staffing Needs at Intensive Care Units

Carlos Hernández and Dafne Lagos
Departamento de Procesos Industriales
Universidad Católica de Temuco
Temuco, Chile
carlos.hernandez.zavala@uct.cl, dlagos@uct.cl

Abstract

Generally, hospitals and health centers have a specialized area called Intensive Care Unit (ICU) aimed to provide treatment for people who are extremely ill. An ICU is staffed with highly trained professionals who operate advance equipment. Patients are referred to ICUs when they have a condition that cannot be treated in a regular hospital room. For example: after being in a serious car accident, after having a heart attack or stroke, when suffering a serious infection such as pneumonia, or after having major surgery. The present research focusses in the development of prediction models based on machine learning algorithms to help planning and allocating medical staff at an ICU. The research was carried out following a classic 4-phase methodology: analysis, design, development, and validation. In the analysis, independent variables such as patient's age, gender, reason for visit, and arrival mode are identified. Data are then pre-processed and prepared for the following phases. Specialize software packages for data mining and machine learning were selected for the investigation. Most of the investigation's activities were planned at this point. During the design, different subsets of the independent variables were selected to be build prediction models. The size of the training and test set as well as the size of the validation set were also defined here. The phases of development and validation were carried out entirely using the software WEKA 3.9.6. By means of executing a battery of experiment and "trial and error" adjustments the prediction models were completed. To accomplish the goals of this research, thousands of records collected from a public hospital were considered. The dataset was divided in a part for training and test (80%) and another part for validation (20%). The approach to predict the medical specialty require for the arriving patients considered the development of models based on artificial neural networks and decision trees. In all cases, instead of a simple hold-out, a 10-fold cross-validation scheme was applied. Results showed that, in general, prediction models with fewer classes generate higher rates of correct predictions and, therefore, they can be more useful when planning and allocating medical staff at an ICU. Even though correct prediction rates fluctuate significantly from one model to another, between 80% and 83%, some of the proposed models are reliable. In conclusion, results showed that prediction models based on machine learning algorithms can help planning and managing the need for professionals and support staff at ICU. The research might help understand the benefits of using machine learning in hospitals and health centers.

Keywords
Prediction Model, Machine Learning, Artificial Neural Network, Medical Staffing Need, Intensive Care Unit.

1. Introduction
According to human resources management theories, employees are commonly referred as the main asset of an organization. Based on that premise, it is essential to have an efficient, friendly and sustainable management of such valuable asset to allow organizations grow up steadily and to endure through time. In healthcare organizations such as a hospital and clinics, a timely and appropriate management of medical staff, medical supplies, and infrastructure is essential. Patients' life depends on that. In particular, since medical specialist, nurses, paramedics, and assistants are in contact with patients, they are responsible for delivering a proper treatment, medicines and usually for taking care during the whole recovery process. However, human resources are also limited. This leads to the need for establishing methods to reach the highest healthcare standards within certain constrains.

The management of knowledge related to organizational processes and its role in the generation of new knowledge that can be used to identify key resources, to improve processes, to implement more sophisticated planning tools, and to improve customer satisfaction (Hamidi et al., 2016), emphasizes the importance of counting with enough information before making decisions. That is the reason why problems related to scheduling, planning, and allocating medical staff have received much attention from researchers in recent years (Chen, P.-S. et al., 2016). The situation is particularly interesting when the available data vires significantly. For instance, the arrival rate of patients at an ICU depends highly on unpredictable situations such as accidents and natural disasters. Specially in densely populated urban centers. An unexpected peak on the number of arrivals, besides of being hard to predict, can certainly have negative consequences (Shavarani et al., 2020).

In Chile, almost 53% of the annual health expenditure is allocated for financing public hospitals. While this figure has been increased steadily through the years, the productivity in the health institutions has not shown the same dynamic during the last decade. Furthermore, if hospitals were classified in groups considering only by their level of complexity, instead of using a budget-based classification, the efficiency breach could reach up to 2.5 times between the most and the least efficient hospital within the same complexity group (Hernríquez, 2020). According to Hernríquez, the potential annual savings of the least efficient hospital can fluctuate from USD 231 million, if the spent were equivalent to the median of the corresponding complexity group, up to USD 846 million, if the hospital had costs equivalent to the most efficient hospital of each complexity group. In other words, these figures represent potential savings that fluctuates between 8% and 29% of the annual healthcare expenditure of 2019. These numbers justify the need for finding opportunities and for developing methods to increase the efficiency with the existing resources.

In recent years, important and challenging research projects were carried out in universities, governmental organizations, and health centers to develop methods and techniques to improve the efficiency of the healthcare organizations. Among the results is the adoption of new tools and methodologies to make better use of data collected and stored in the databases of the public healthcare system and to support the decision making process when planning and allocating medical staff, supplies and infrastructure. The following are some examples:

- a study on the comprehensive application of PDCA and FMEA in healthcare centers' management, whose findings revealed significant improvements in efficiency, in teamwork, and in the moral of the medical staff (Chen, H. et al., 2022).
- an investigation on the allocation of medical staff and the corresponding occupation levels carried out by personnel of the Community Child Healthcare Medical Staff (CCHCMS) of Guangzhou in China revealed understaffing problems and high risk of staff turnover. Furthermore, factors such as low wages, contract type, education level, and work environment would increase the probability of staff turnover (Hu et al., 2020).
- the development of a discrete event simulation model using an algorithm to prioritize patients according to a triage-based criterion. Triage is a method of selection and classification of patients used in nursing and emergency and disaster medicine. In this research, mayor challenges came from the stochastic nature of the variables and from time constrains to compute solutions. Nonetheless, numerical experiments showed that in comparison to the current staff allocation the new proposal could improve the overall performance by 38.28% (Chen, W. et al., 2022).
- the development of a two-stage strategy for allocating medical staff in healthcare center networks and for defining their monthly schedules. The proposal helped decision makers minimize the medical staff required at each center (Chen, P.-S. et al., 2020).
- an information system based on analytical models for handling heterogeneous data to help decision makers allocate medical staff minimizing costs (Chaovalitwongse et al., 2017).

Regarding the improvement of the performance of medical staff as well as the service quality, and the patient satisfaction, some of the factors to consider are:

- a good compensation package based on an objective criterion not only impact the employees' interest but also the operational efficiency and the organization development. Therefore, it is important to make well use of the existing technologies to innovate and to modernize the way in which employees are being compensated (Qi et al., 2020).
- enough information about the level of satisfaction among employees as well as on the work environment, and overall working conditions to allow decision makers implement initiatives and action plans to improve employees' experience within the organization (Kim and Weng, 2018).
- information about the work satisfaction to implement an action plan considering elements such as professional development opportunities, motivation to achieve goals, compensation packages, working

environment and good communication allow decision maker understand and improve the overall quality of the healthcare centers (Paschalidou and Mpogiatzidis, 2018).

1.1 Objective
To develop models based on machine learning algorithms to estimate medical specialist required by patients received at an intensive care unit (ICU).

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, regression, and forecasting
In machine learning there several important task: classification, regression, and forecasting. Classification can be understood as the determination of the target class to be predicted, 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 when holding out data. 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.

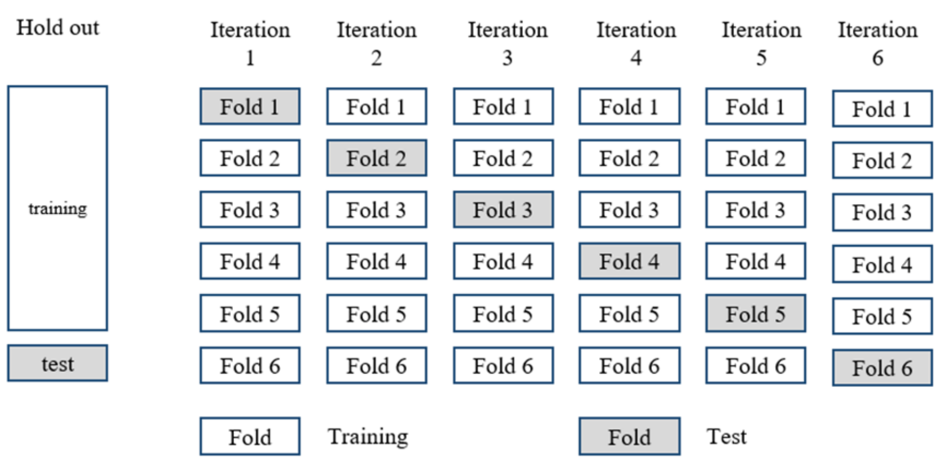

Figure 1. Hold out and cross-validation (k=6)

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).

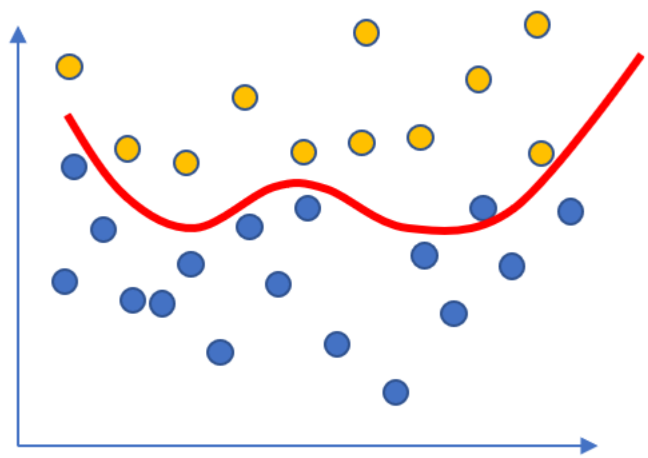

Figure 2. Overfitting

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.

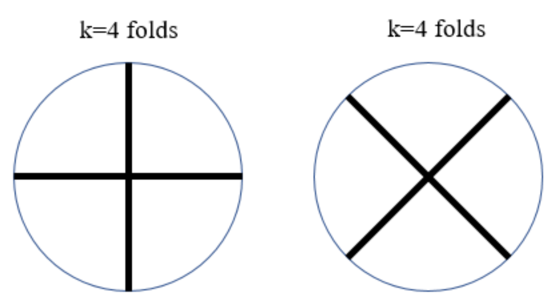

Figure 3. Different folds in cross-validation (k=4)

2.6 Artificial neural network (ANN)

An ANN is a construction made of nodes, referred as neurons, that are combined in an interconnected layered structure (Figure 4). The input layer corresponds to the nodes that receive the externa data. In the second level contains the hidden layers that transform the input data for the output layer, whose neurons are responsible for delivering the results generated by the network (Morano and Tajani 2013). The topology of an ANN is determined by the number of layers, their nodes, and a transfer function.

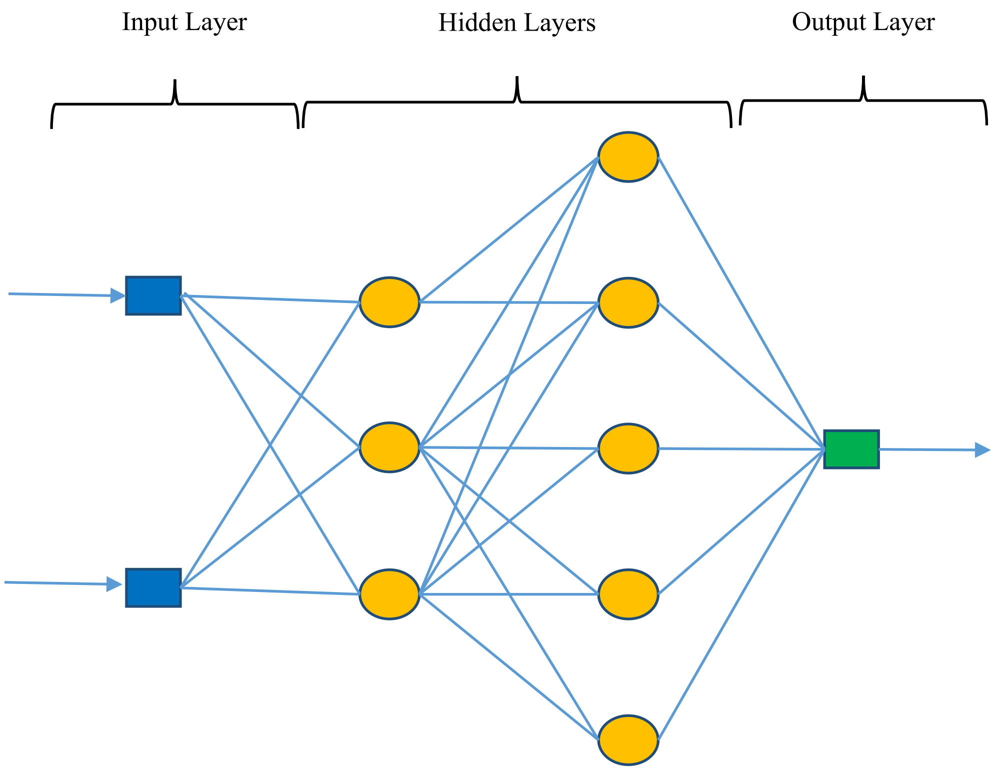

Figure 4. ANN's input, hidden, and output layers

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

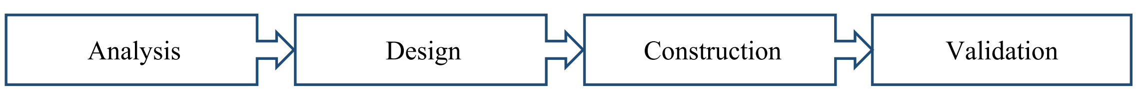

Figure 5. Four-stage model

3.1 Analysis
During the analysis, the preprocessing of ICU records was completed. For the purposes of this investigation, 15,000 records corresponding to patients attended during 2020 were considered. The complete dataset had over 76,000 records from which a fraction was used in this work. Each record contains the information of a patient who received medical care at ICU. Thus, every time a patient arrived at ICU a new record was appended to the database. The data contained in a single record goes from the arrival time to the health insurance type.

The patient record fields considered in this work are the following: age, gender, reason for visit, arrival mode, source of admission, presence of trauma, alcohol test result, and medical specialist. 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	The source of admission to ICU.
Trauma	Presence of trauma
Alcohol test result	Test to detect the presence of alcohol in the patient.
Medical specialist	Doctor focused on a defined group of patients, diseases, skills, or philosophy.

The data preprocessing revealed that most of the patients required systematically certain type of medical specialists. For instance, only 4 medical specialties were enough to cover 80% of all cases received at ICU during 2020. In the same way, 7 specialties concentrated almost 95% of all cases (Table 2).

Table 2. Record distribution per medical specialty

Number of specialties	Number of records	Percentage of records
4	56,609	80 %
7	66,155	94 %
10	69,814	99 %
15	70,066	100 %

A summary of the number of specialties and the number of corresponding records is presented in Table 3 and Figure 6. It can be seen that 80% of all records, or patients, corresponded to only 4 medical specialties.

Table 3. Medical specialties v/s patient records

Number of specialties	Percentage of specialties	Number of records	Percentage of total records
1	7 %	32,974	47 %
2	13 %	43,638	62 %
3	20 %	52,867	75 %
4	27 %	56,609	80 %
5	33 %	60,196	85 %
6	40 %	63,312	90 %
7	47 %	66,155	94 %
8	53 %	68,560	97 %
9	60 %	69,282	98 %
10	67 %	69,814	99 %
11	73 %	70,276	99 %
12	80 %	70,461	100 %
13	87 %	70,642	100 %
14	93 %	70,688	100 %
15	100 %	70,699	100 %

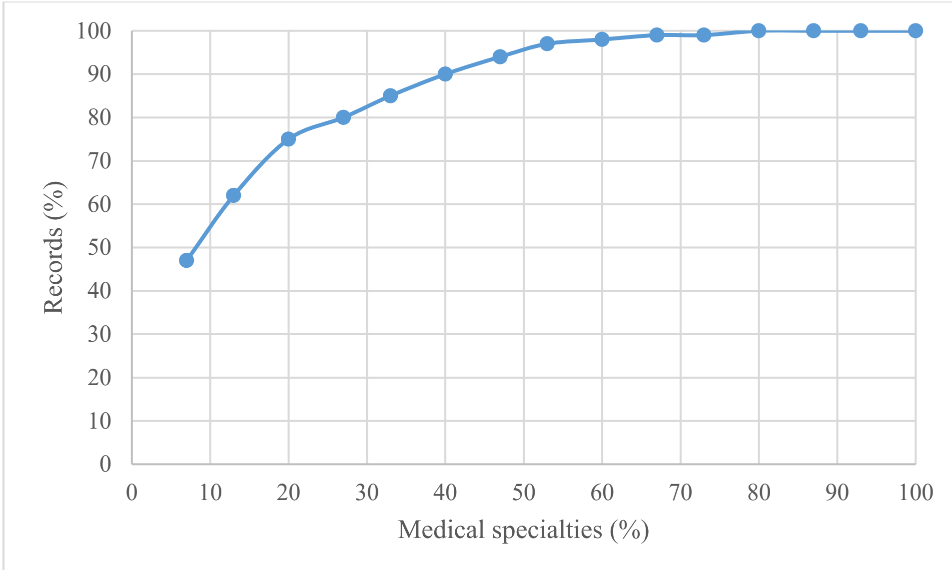

Figure 6. Medical specialties v/s patient records (%)

3.2 Design

The complete ICU 2020 dataset can be seen as a matrix composed by of rows representing patient records and columns representing fields or attributes. The challenge is to classify each record into a certain class. Being the medical specialties the only available classes.

During the design, three datasets were prepared. The first one containing only the top four most recurrent specialties, which corresponds to 80% of the records. The second one including the top seven specialties. And the last one having 15 specialties (Table 4).

Table 4. Dataset creation

Dataset	Number of specialties	Number of records
DS-04	4	15,000
DS-07	7	15,000
DS-15	15	15,000

Datasets were divided to create two subsets in a proportion of 80% and 20% respectively having 12,000 records (80%) for training and test with 80% of data, and 3,000 records (20%) for validation (Figure 7 and Table 5).

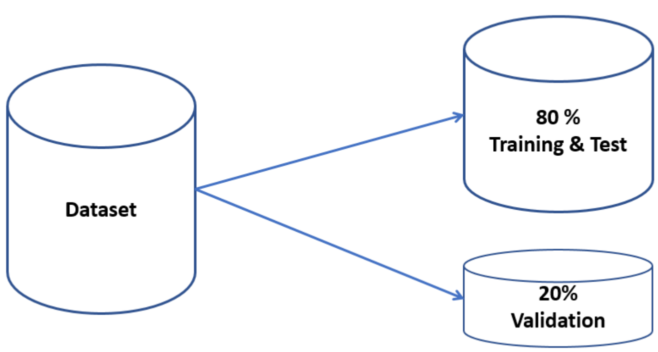

Figure 7. Dataset split up

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

Dataset	Records for training and test	Records for validation
DS-04	12,000	3,000
DS-07	12,000	3,000
DS-15	12,000	3,000

A validation dataset is used to confirm whether the prediction model, prepared with the dataset for training and test, is able to generalized properly when it is used to predict with unknown data from the dataset for validation.
In this work, several prediction models were developed using either artificial neural networks or decision trees. Namely: M-04, M-07, and M-15. As aforementioned, the scope of the research is limited to a fraction of the complete ICU database of 2020. The corresponding specialties, number of records, and prediction models are shown in Table 6.

Table 6. Specialties and prediction models

ID	Diagnose code	Number of records	M-04	M-07	M-15
S-01	General surgery	32,974	✓	✓	✓
S-02	Obstetrics and gynecology	10,664	✓	✓	✓
S-03	Pediatrics	9,229	✓	✓	✓
S-04	Midwifery	3,742	✓	✓	✓
S-05	Adult Trauma	3,587		✓	✓
S-06	Emergency medicine	3,116		✓	✓
S-07	General medicine	2,843		✓	✓
S-08	Neurologist	2,405			✓
S-09	Neurosurgery	722			✓
S-10	Internal medicine	532			✓
S-11	Pediatric surgery	462			✓
S-12	Psychiatry	185			✓
S-13	Odontology	181			✓
S-14	Urology	46			✓
S-15	Gynecology	11			✓

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, 2016).

The resulting classification models were 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. A deeper explanation on the interpretation of these curves can be found in ht work of Davis and Goadrich (Davis and Goadrich, 2006).

3.3 Construction
The objective was to build models capable of predicting the medical specialty required by arriving patients at ICU by means of applying machine learning algorithms. In this case, 8 fields (attributes) were used to predict the class *medical specialty*.

To help overcome 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 which 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	**Specialties**	**Records**	**Correct prediction (%)**	**Precision**	**Recall**	**ROC AUC**
M-04	4	12,000	77.7 %	0.778	0.777	0.909
M-07	7	12,000	60.2 %	0.499	0.602	0.799
M-15	15	12,000	56.5 %	?	0.565	0.788

Although cross-validating helped reduce the risk of overfitting, the effect of the fold partitioning remained. This fact was analyzed with more detail 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	78.59 %	2.57
M-07	60.78 %	3.77
M-15	57.69 %	4.55

3.4 Validation
The validation of the proposed prediction 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 preprocessed for each model (Table 9).

Table 9. Correct predictions rates with unknown validation data

				Weighted average		
Model	**Specialties**	**Records**	**Correct predictions**	**Precision**	**Recall**	**ROC AUC**
M-04	4	3,000	78.8 %	0.782	0.788	0.913
M-07	7	3,000	64.4 %	?	0.644	0.810
M-15	15	3,000	55.6 %	?	0.556	0.806

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). Even though, correct prediction rates decrease systematically as the number of specialties increased. Nevertheless, when the number of specialties is limited to four, the results are promising.

Table 10. Prediction model comparison

		Train and test dataset		Validation dataset	
Model	**Codes**	**Records**	**Correct predictions**	**Records**	**Correct predictions**
M-04	4	12,000	77.7 %	3,000	78.83 %
M-07	7	12,000	60.2 %	3,000	64.4 %

M-15	15	12,000	56.5 %	3,000	55.6 %

5. Results and Discussion

As the number of specialties (classes) increases, the rate of correct prediction decreases. Regardless the size of the dataset, the inclusion of additional classes would make the prediction more difficult.

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. Confusion matrices corresponding to M-04, M-07, and M-15 are presented in Table 12, Table 13, Table 14, Table 15, Table 16, and Table 17.

Table 12. Confusion matrix of M-04. Training and test dataset with cross-validation (12,000 records)

Specialty	**S-04**	**S-03**	**S-01**	**S-02**
S-04	222	5	25	489
S-03	0	2,630	175	8
S-01	20	891	4,697	323
S-02	253	30	458	1,774

Table 13. Confusion matrix of M-04. Validation dataset (3,000 records)

Specialty	**S-04**	**S-03**	**S-01**	**S-02**
S-04	55	0	11	99
S-03	0	699	98	4
S-01	7	125	1305	90
S-02	31	0	170	306

It is important to keep in mind the additional pieces of information were obtained by means of transforming the field *Reason for visit* from a string to a vector of words. Furthermore, it is possible that patients with the same *Reason for visit* required a different specialist. Thus, the probability of misclassification is significant as shown in Table 14 and Table 15.

Table 14. Confusion matrix of M-07. Training and test dataset with cross-validation (12,000 records)

	S-04	**S-03**	**S-07**	**S-01**	**S-02**	**S-06**	**S-05**
S-04	**215**	2	0	23	362	1	0
S-03	2	**1,805**	4	322	90	1	5
S-07	0	216	**0**	614	60	13	0
S-01	7	573	5	**3,651**	334	88	4
S-02	196	6	7	371	**1,519**	5	0
S-06	0	16	1	739	37	**30**	0
S-05	0	161	3	470	40	2	**0**

Table 15. Confusion matrix of M-07. Validation dataset (3,000 records)

	S-04	S-03	S-07	S-01	S-02	S-06	S-05
S-04	**39**	6	0	9	85	0	0
S-03	2	**594**	0	20	2	0	0
S-07	16	22	**0**	124	36	0	0
S-01	16	196	0	**1,055**	57	0	0
S-02	32	12	0	137	**243**	0	0
S-06	0	8	0	186	0	**0**	0
S-05	1	20	0	81	1	0	**0**

Table 16. Confusion matrix of M-15. Training and test dataset with cross-validation (12,000 records)

	S-04	S-03	S-07	S-01	S-02	S-11	S-06	S-08	S-10	S-05	S-12	S-09	S-14	S-13	S-15
S-04	**236**	2	0	18	330	0	3	0	0	0	0	0	0	0	0
S-03	1	**1,693**	12	369	44	0	20	6	0	1	0	0	0	0	0
S-07	1	218	**5**	577	39	0	21	7	0	4	0	0	0	0	0
S-01	11	552	12	**3,376**	206	0	143	51	0	14	0	0	0	0	0
S-02	264	14	5	371	**1,292**	0	46	0	0	1	0	0	0	0	0
S-11	0	31	0	22	1	**0**	0	0	0	0	0	0	0	0	0
S-06	0	15	0	657	28	0	**47**	19	2	3	0	0	0	0	0
S-08	1	3	0	205	7	0	19	**131**	3	2	0	0	0	0	0
S-10	0	0	0	88	3	0	8	5	**1**	1	0	0	0	0	0
S-05	0	159	3	433	20	0	6	1	0	**4**	0	0	0	0	0
S-12	0	0	0	21	2	0	5	2	0	0	**0**	0	0	0	0
S-09	0	7	0	57	2	0	2	2	0	0	0	**0**	0	0	0
S-14	0	0	0	6	0	0	1	0	0	0	0	0	**0**	0	0
S-13	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0
S-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**

Table 17. Confusion matrix of M-15. Validation dataset (3,000 records)

	S-04	S-03	S-07	S-01	S-02	S-11	S-06	S-08	S-10	S-05	S-12	S-09	S-14	S-13	S-15
S-04	**41**	0	0	2	71	0	0	3	0	0	0	0	0	0	0
S-03	3	**532**	0	5	5	0	0	3	0	2	0	0	0	0	0
S-07	5	13	**0**	75	22	0	0	24	0	0	0	0	0	0	0
S-01	21	205	0	**729**	161	0	0	174	0	3	0	0	0	0	0
S-02	42	1	0	8	**333**	0	0	17	0	0	0	0	0	0	0
S-11	0	39	0	1	0	**0**	0	0	0	0	0	0	0	0	0
S-06	1	4	0	135	13	0	**0**	54	0	1	0	0	0	0	0
S-08	0	0	0	46	0	0	0	**34**	0	0	0	0	0	0	0
S-10	0	0	0	4	0	0	0	1	**0**	0	0	0	0	0	0
S-05	0	23	0	86	6	0	0	10	0	**0**	0	0	0	0	0
S-12	0	0	0	4	0	0	0	6	0	0	**0**	0	0	0	0
S-09	0	2	0	18	0	0	0	9	0	0	0	**0**	0	0	0
S-14	0	0	0	3	0	0	0	0	0	0	0	0	**0**	0	0
S-13	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0
S-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**

6. Conclusion

ICU database with patient records gathered during 2020 provided a valuable insight of the daily operation at a public healthcare hospital.

The preprocessing of data during the phase of analysis along with the initial analysis revealed that an important percentage of the records were associated to a few number of medical specialties. ICU 2020 database showed that four specialties concentrated up to 80% of all patient records.

Prediction models based on machine learning algorithms are interesting and innovative tool to support the decision making process. Specially during the planning and scheduling. Although counting with almost 75,000 records, the proposed prediction models minimized the requirements of data and the processing time to necessary produce results. Therefore, only a fraction of the database was used during the investigation. In total, only eight fields per record were considered to predict the target class *Medical specialty*. The transformation of the field *Reason for visit* from a string to a vector of words provided more pieces of information by means of creating additional attributes to build larger models.

When working with machine learning algorithms, it is a common practice to apply cross-validation schemes to minimize the influence of the dataset split up, instead of simply holding out a fraction of the dataset. Averaging several results is expected to better than having only one result. Furthermore, running replications help reduce the bias caused by the fold partitioning. Besides that, having a collection of unknown data only for validation help confirming whether the models can generalize properly or not.

The experiments carried out confirmed that models having fewer target class values produced systematically better correct prediction rates. For example, with four class values corresponding to the top four medical specialties, the correct prediction rate was over 80%. In contrast, with 15 class values the correct prediction rate was close to 55%. It is worthwhile to mention the consistency of the experimental results when applying a cross-validation scheme, a validation dataset with unknown records, and a battery of 10 replications. The predictions made with unknown data proved that at least the proposed model M-04 was able to generalize well.

Finally, correct prediction ratios and confusion matrices confirm that the approach presented in this work can be useful tool to estimate staffing needs and to support decisions making when allocating specialists at an ICU.

References

Chaovalitwongse P., Somprasonk K., Phumchusri N., Heim J., Zabinsky Z.B., and Chaovalitwongse W.A., A decision support model for staff allocation of mobile medical service. *Annals of Operations Research*, vol. 249, pp. 433–448, 2017.

Chen H., Tao Z., Zhou C., Zhao S., Xing Y., and Lu M., The Effect of Comprehensive Use of PDCA and FMEA Management Tools on the Work Efficiency, Teamwork, and Self-Identity of Medical Staff: A Cohort Study with Zhongda Hospital in China as an Example. *Contrast Media & Molecular Imaging*, vol. 2022, 2022.

Chen P.-S., Huang W.-T., Chiang T.-H., and Chen G.Y.-H., Applying heuristic algorithms to solve inter-hospital hierarchical allocation and scheduling problems of medical staff. *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 318-331, 2020.

Chen P.-S., Lin Y.-J., and Peng N.-C., A two-stage method to determine the allocation and scheduling of medical staff in uncertain environments. Computers & Industrial Engineering, vol. 99, pp. 174-188, 2016.

Chen W., Guo H., and Tsui K.-L., A new medical staff allocation via simulation optimization for an emergency department in Hong Kong. *International Journal of Production Research*, vol. 58, no. 19, pp. 6004-6023, 2020.

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.

Hamidi Y., Soltanian A., Mohammadfam I., and Rabbani T., Studying the status of knowledge management aspects by using Bukowitz & William model in Staff and Deputy of Hamadan University of medical sciences *Bulletin de la Société Royale des Sciences de Liège*, vol. 85, pp. 1529-1544, 2016.

Henríquez, M., Informe de análisis: Eficiencia del gasto hospitalario en Chile. *Fundación Observatorio Fiscal*, 2020.

Hu F., Lu J., Lin S., Hua N., Wu Q., Liang J., and Song Y., Human resources allocation and turnover intention of community child health care medical staff in Guangzhou, *Chinese Journal of General Practitioners*, vol. 6, pp. 1025-1030, 2020.

Kim, SH. and Weng, SJ., Incorporating work experience of medical staff into patient safety climate management: a multi-group analysis, *BMC Health Serv. Res.,* vol. 18, no. 919, 2018.

Morano, P., and Tajani, F., Bare ownership evaluation. Hedonic price model vs artificial neuronal network, *International Journal of Business Intelligence and Data Mining*, vol. 8, no. 4, pp. 340-360, 2013.

Paschalidou I., and Mpogiatzidis P., Investigation of the relationship between management techniques and job satisfaction of the medical and nursing staff in public hospitals, *Archives of Hellenic Medicine*, vol. 35, no. 6, pp. 778-783, 2018.

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

Qi B., Wang J., Ji M., Xu S., Li J. and Shi C., Design of the management system for salary diagnosis and optimization of medical staff in Shaanxi public hospitals of traditional Chinese medicine, *Proceedings of Conference: Big Data Analytics for Cyber-Physical System in Smart City*, pp.1733-173, 2020.

Saito, T. and Rehmsmeier, M., The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets, *PLoS ONE*, vol. 10, no.3, 2015.

Shavarani, S., Golabi, M., and Vizvari, B., Assignment of Medical Staff to Operating Rooms in Disaster Preparedness: A Novel Stochastic Approach, *IEEE Transactions on Engineering Management*, vol. 67, no. 3, pp. 593-602, 2020.

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. Through the years 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 multinational companies such as 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.

Dafne Lagos is an industrial engineer and university professor. She obtained a Bachelor of Science in Engineering and a Master's in Management from the Universidad de La Frontera (Temuco, Chile) and a Doctor of Science from Atlantic International University (USA). She has lectured in Engineering Economics, Probability Modeling, Logistics, and Operations Research for engineering students. She has also written several scientific articles in the area of operations research. During her academic tenure, she has been appointed to different managerial positions and has mentored more than forty students. Her research interests include financial planning, supply chain management, and logistics.