IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 3, September 2024, pp. 3129~3140
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i3.pp3129-3140

Emergency patient forecasting with models based on support vector machines

Carlos Hernandez, Dafne Lagos, Paola Leal, Jaime Castillo
Department of Industrial Processes, Faculty of Engineering, Universidad Católica de Temuco, Temuco, Chile

Article Info

Article history:

Received Dec 31, 2023
Revised Mar 4, 2024
Accepted Mar 14, 2024

Keywords:

Emergency room
Forecasting
Linear regression
Machine learning
Support vector machine

ABSTRACT

Understanding the dynamic nature of the influx of patients is crucial for efficiently managing supplies, medical personnel, and infrastructure in an emergency room (ER). While overestimation can lead to resource wastage, underestimation can result in shortages and compromised service quality. This study addresses emergency patient forecast by means of implementing support vector machine (SVM) algorithms. Along four phases (analysis, design, development, and validation), more than 50,000 ER records were preprocessed and analyzed. Traditional error metrics such as mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and mean absolute percentage error (MAPE) were utilized alongside monthly consolidated forecasts. To benchmark performance, actual values and forecasts derived from linear regression (LR) models were used. Experiments revealed that LR models had lower errors compared to SVM models. However, monthly consolidated forecasts showed that SVM-based models underestimated less than LR-based models. In conclusion, SVM-based models could help planners to accurately estimate the requirements for supplies and medical personnel during the period under study.

Corresponding Author:

Carlos Hernandez
Department of Industrial Processes, Faculty of Engineering, Universidad Católica de Temuco
Rudecindo Ortega 02950, Edificio Adalberto Salas, Temuco 4781312, Chile
Email: carlos.hernandez.zavala@uct.cl

1. INTRODUCTION

Public healthcare systems encompass hospitals, clinics, and health centers, each varying in size and complexity. Large hospitals typically have an emergency room (ER) to provide immediate care to patients in critical condition. However, the inherent unpredictability of emergencies, such as strokes, heart attacks, and traffic accidents, makes the estimation of resources a complicated challenge for planners [1]. It is not uncommon to see overcrowded waiting rooms filled with dissatisfied patients who have to wait long before receiving care, increasing the adverse effects of their condition [2].

In many instances, ERs serve as the primary point of entry for a significant number of patients, increasing the hospital congestion and extending wait times [3]. Accurate forecasting of patient influx would allow planners improve procurement and allocation of resources [4], thereby minimizing the risk of shortages and facilitating informed medical decision-making through the timely availability of supplies for diagnostic procedures and therapies. Typically, public healthcare systems have a centralized department responsible for the procurement process, generating economies of scale through negotiations involving large volumes of supplies. Annual procurement plans undergo periodic reviews to ensure alignment with budgetary constraints. Accurate estimation is imperative due to the consequences of overestimation resulting in operational inefficiencies and underestimation leading to deficiencies and shortages.

According to the Organisation for Economic Co-operation and Development [5], the USA ranks as the foremost global spender on healthcare, surpassing other important economies like Germany and Switzerland. In Chile, approximately 53% of the annual health budget is allocated for financing the public hospital network. Despite a sustained increase in budgetary allocation over the past decade, the system's productivity has not exhibited a comparable growth.

Over time, machine learning (ML) algorithms have gained increasing recognition and application in different fields. In healthcare, ML techniques have revolutionized the diagnostic process by combining a multitude of factors to estimate the likelihood of specific events [6]. ML-based models diverge from traditional approaches applied in the healthcare industry by using statistical techniques that enable specialists to improve diagnostic precision through the analysis of massive datasets, constructing intricate non-linear relationships rather than relying solely on linear regressions (LR) [7], [8]. While identifying influential factors in diagnosis is crucial, treating patients in critical condition presents an even greater challenge, as it involves the risk to life.

Various ML-based models serve different purposes, including predicting the likelihood of disease occurrence and forecasting future events [9]. In the healthcare sector, ML algorithms, such as artificial neural networks (ANN), have been increasingly applied to enhance services and reduce costs [10]. In diabetes prevention, a range of ML algorithms including k-nearest neighbor (KNN), support vector machine (SVM), decision trees, naive Bayes, and logistic regression have been employed too [11]. Similarly, for predicting heart diseases, deep learning techniques have been applied to analyze data and to compare the results obtained with other algorithms such as SVM, naive Bayes, and KNN [12]. By means of analyzing electronic health records (EHR), specialists can identify and interpret patterns to facilitate timely decision-making based on non-trivial predictions. Furthermore, combining predictions from multiple algorithms has shown advances in the overall performance and accuracy [13]. While statistical models and autoregressive integrated moving average (ARIMA)-based approaches have been extensively studied and utilized in forecasting over the years [14], the COVID-19 sanitary crisis has catalyzed the adoption of ML algorithms and other forms of artificial intelligence (AI) for predicting, diagnosing, and detecting positive cases [15]. Additionally, ML algorithms have been useful in identifying biomarkers associated with patient mortality [16].

Significant enhancements in patient classification have been achieved through the application of ML-based classifiers [17]. Moreover, research has demonstrated that ML-based models can outperform classical moving averages (MA) in predicting wait times in large queues [18]. Furthermore, ML-based models have been applied to deal with the complexity and randomness in delay and wait time patterns [19].

SVM-based models have demonstrated efficacy in dealing prediction challenges when handling large datasets [20]. It is no coincidence that both ANN and SVMs are among the most used forms of AI in the healthcare industry today [21]. While ANN-based models are widely employed in image recognition [22], SVM-based models have found application in diagnosing renal diseases [23].

Despite the extensive exploration of SVM-based forecast models in time series analysis, particularly in financial fields [24]–[27], there appears to be limited research on the utility of SVM for forecasting patient influx, to the best of the authors' knowledge. Existing studies predominantly focus on forecasting utilizing classical methods [28]–[30], or ML algorithms other than SVM [31]. Hence, the contribution of this work is, on the one hand, the analysis of a vast database collected during the COVID crisis and, on the other hand, the application of SVM-based models to forecast service demand in the healthcare industry.

Understanding the nature of the emergency patient influx is crucial for planners to accurately estimate the requirements for supplies, medical personnel, and infrastructure. This investigation explores the utilization of SVM-based models for iteratively forecasting the monthly influx of ER patients with 30-day horizons. In addition to employing classical error metrics such as mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and mean absolute percentage error (MAPE), overestimation or underestimation in monthly patient counts are analyzed too. To establish a benchmark for comparison, LR based forecasts serve as the baseline, while error metrics are computed against actual values.

2. METHOD

The investigation advanced through four distinct phases: analysis, design, construction, and validation, as depicted in Figure 1. The initial phases encompassed activities related to data preprocessing, overall planning, and comparing alternatives, while the latter phases focused on programming and result comparison.

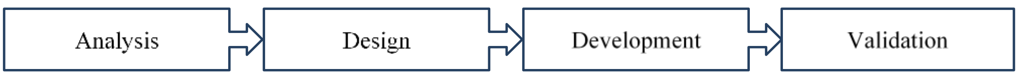


Figure 1. 4-phase research

2.1. Phase of analysis

Throughout the analysis phase, significant effort was dedicated to collecting data from diverse sources and integrating them into a comprehensive dataset. This task was executed by a specialized management unit at Hospital Hernan Henriquez Araneda, situated in Temuco, Chile. The investigation encompassed approximately 50,000 records of patients treated in the ER.

Each record includes patient demographics, reason for ER admission, and a number of fields ranging from admission time to health insurance type. Given the unpredictable nature of emergencies, daily patient influxes exhibit considerable variability. To elucidate this dynamic, the data was organized into a chronological sequence, forming daily time series. As depicted in Figure 2, a significant decline in patient numbers is observed following the summer months of January, February, and early March.

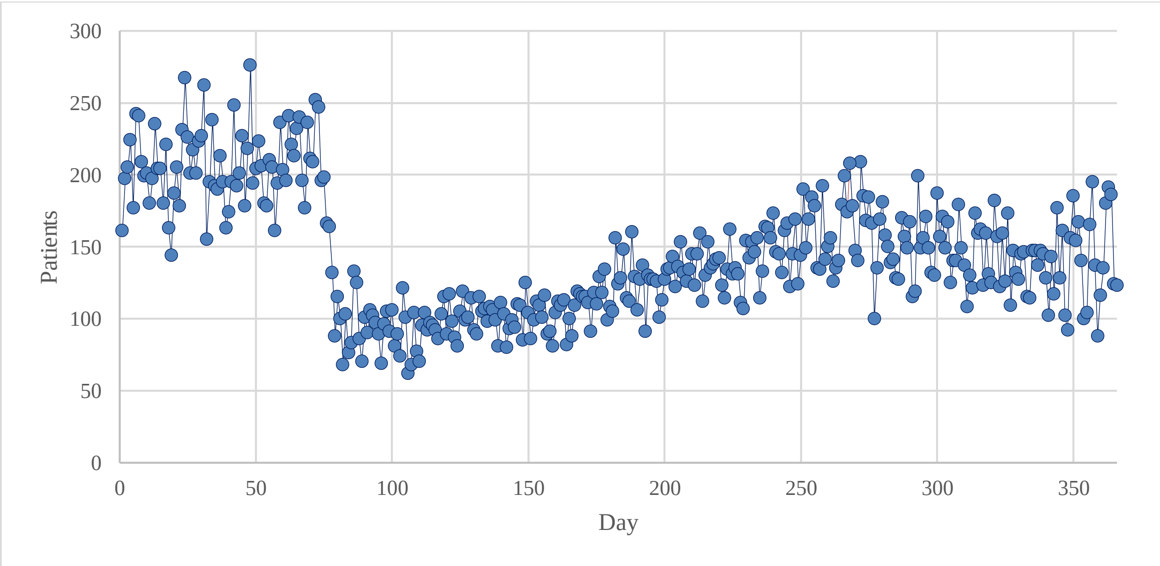


Figure 2. Daily number of incoming ER patients over the course of a year

It is intriguing to observe the manifestation of one of the most known principles in engineering, namely the Pareto principle. In this context, Table 1 and Figure 3 show the relationship between the number of patients and the corresponding number of medical specialties required for their treatment. Remarkably, it is found that approximately 75% of all patients required only 20% of the total medical specialties. Only 3 out of 15 specialties. Given the focus of this study on enhancing the accuracy of forecasts to assist planners in estimating the needs for supplies, medical personnel, and infrastructure, the focus was set in the most demanded specialties. By narrowing down the list, planners can concentrate their efforts on the predominant needs, thereby optimizing efficiency and resource allocation.

Table 1. Number of patients sorted by medical specialty during a 12-month period.

Medical specialty	Number of patients	Accumulated patients (%)
General surgery	32,974	47
Obstetrics and gynecology	10,664	62
Pediatrics	9,229	75
Midwifery	3,742	80
Adult Trauma	3,587	85
Emergency medicine	3,116	90
General medicine	2,843	94
Neurologist	2,405	97
Neurosurgery	722	98
Internal medicine	532	99
Psychiatry	462	99
Pediatric surgery	185	100
Odontology	181	100
Urology	46	100
Gynecology	11	100

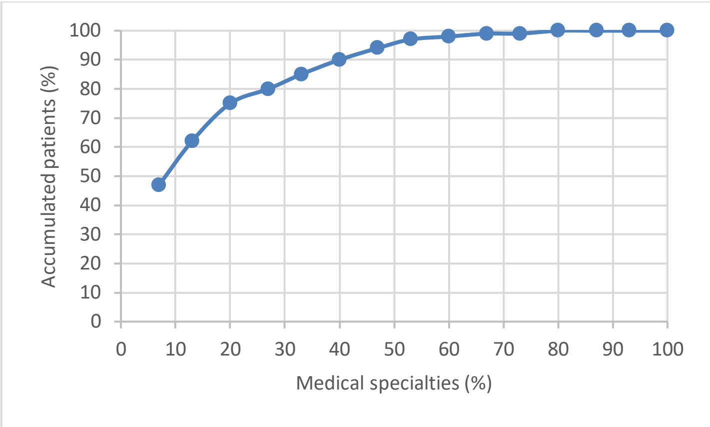


Figure 3. Principle of Pareto: 75% of all incoming patients required 20% of all medical specialties

2.2. Phase of design

Public health systems typically depend on a centralized procurement office for acquiring supplies. Planners have to estimate future needs periodically, be it annually, semiannually, or monthly, to ensure the provision of adequate service levels. In this research, the forecast horizon of 30 days was used. Consequently, over the span of one year, assuming a monthly review of procurement plans, iterative 30-day forecasts were generated from March to December. January and February were arbitrarily excluded to facilitate the development of initial forecast models.

For the purposes of this work, the total number of patients in a day must be understood as a single observation. Two alternative approaches were explored. The first involves utilizing all accumulated observations, while the second focuses solely on data from previous month. For instance, as illustrated in Figure 4, the 31-day forecast for December used information either from the 335 available observations spanning from January to November or from the latest 30-day period, specifically November. Forecasts derived from LR models served as the baseline for comparison with those generated by SVM models. While the simplicity and widespread use of LR justify its inclusion, the selection of SVMs is based on their documented efficacy in time series forecasting.

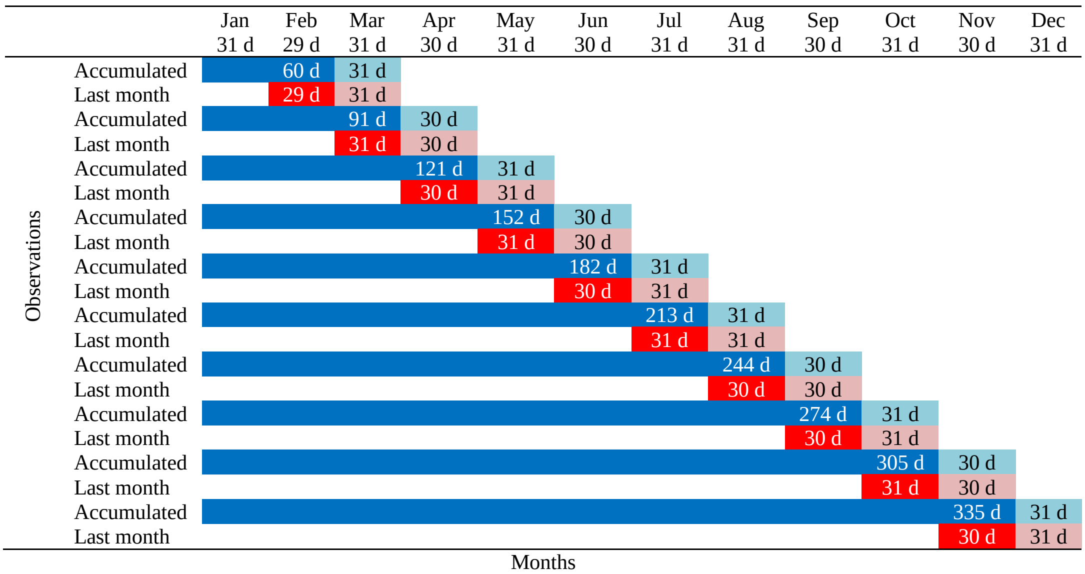


Figure 4. Monthly forecasts with accumulated and last month’s observations

Standard error metrics such as MAE, MSE, RMSE, and MAPE were employed to assess the accuracy of the forecasts. However, rather than calculating errors retrospectively from past values, they were computed based on the difference between the resulting 30-day forecast and the corresponding actual values. For instance, December's 31-day forecast, calculated from either 335 or 30 observations, was compared with the actual patient influx received during December.

While overestimating patient influx may lead to resource over allocation and operational inefficiencies, underestimating needs could cause shortages of critical supplies, potentially resulting in loss of life. Therefore, as an additional metric, the frequency of months where the forecast falls below the actual value was incorporated into the investigation.

2.3. Phase of development

The initial step of model development involves dataset preparation. Data, in the form of daily time series, were organized into separate workable files to facilitate further calculations. Given the forecast horizon of one month (30 or 31 days), ten forecasting models were created, going from March to December. The first two months (60 days) were set aside to develop the initial models for forecasting March. Table 2 provides a summary of the monthly forecast models, indicating their respective observation periods and forecast horizons.

For this phase, two software packages were utilized: Microsoft Excel 2016 and WEKA v3.9.6. Initially, LR models were constructed using the regression analysis capabilities within Microsoft Excel. Subsequently, SVM models were developed using the forecasting package offered by WEKA.

Table 2. Summary of monthly forecast models, specifying the start and end dates of the observation period, as well as the forecast horizon

	Observations						Forecast horizon		
	Accumulated			Last month					
	Days	From	To	Days	From	To	Days	From	To
March	60	1 Jan	29 Feb	31	1 Feb	29 Jan	31	1 Mar	31 Mar
April	91	1 Jan	31 Mar	30	1 Mar	31 Mar	30	1 Apr	30 Apr
May	121	1 Jan	30 Apr	31	1 Apr	30 Apr	31	1 May	31 May
June	152	1 Jan	31 May	30	1 May	31 May	30	1 Jun	30 Jun
July	182	1 Jan	30 Jun	31	1 Jun	30 Jun	31	1 Jul	31 Jul
August	213	1 Jan	31 Jul	31	1 Jul	31 Jul	31	1Aug	31Aug
September	244	1 Jan	31Aug	30	1Aug	31Aug	30	1 Sep	30 Sep
October	274	1 Jan	30 Sep	31	1 Sep	30 Sep	31	1 Oct	31 Oct
November	305	1 Jan	31 Oct	30	1 Oct	31 Oct	30	1 Nov	30 Nov
December	335	1 Jan	30 Nov	31	1 Nov	30 Nov	31	1 Dec	31 Dec

Given their simplicity and widespread usage, LR models served as the initial method for forecasting the patient influx expected over the subsequent 30 or 31 days. These LR models served as the baseline for comparing forecasts generated by SVM models. Table 3 displays the monthly LR forecast models alongside their respective forecast horizon, number of observations, and mathematical expression. For example, the LR model utilized to forecast June (30 days) necessitated either 152 or 31 observations, depending on the chosen approach: accumulated or last month. In this section, SVM models are not included due to their complexity and limited readability. Instead, comprehensive summary tables containing forecasts and error metrics will be presented in subsequent sections.

Table 3. Summary of monthly LR forecast models with their corresponding horizon, number of observations, and mathematical expression

Forecast		Accumulated		Last month	
Month	Days	Observations	LR Model	Observations	LR Model
March	31	60	Y=203.40 X + 0.03	29	Y=184.64 X + 0.37
April	20	91	Y=229.59 X - 0.88	31	Y=608.50 X - 5.90
May	31	121	Y=242.20 X -1.26	20	Y=73.36 X + 0.19
June	30	152	Y=232.67 X - 1.05	31	Y=97.52 X + 0.02
July	31	182	Y=218.75 X - 0.81	30	Y=-24.74 X + 0.80
August	30	213	Y=202.14 X - 0.56	31	Y=-45.24 X + 0.81
September	30	244	Y=188.33 X - 0.38	30	Y=-90.28 X + 0.97
October	31	274	Y=175.14 X - 0.22	30	Y=129.88 X + 0.07
November	30	305	Y=167.95 X - 0.15	31	Y=284.91 X + 0.45
December	31	335	Y=164.69 X - 0.12	30	Y=-10.12 X + 0.43

2.4. Phase of validation

The validation process encompassed two distinct approaches. Firstly, forecasts were reviewed to ensure their positivity and identify any anomalous values. Secondly, correlation analyses were conducted to ascertain the fidelity of the proposed models' forecast values with existing data. A comparison of monthly correlation coefficients for both approaches-accumulated and last month's observations-is presented in Table 4.

Table 4. Summary of monthly correlation coefficients between actual and forecasted values

Forecast		SMV based forecast		LR based forecast	
Month	Days	Accumulated	Last month	Accumulated	Last month
March	31	-0.65	-0.44	-0.86	-0.86
April	30	0.43	-0.17	-0.17	-0.17
May	31	0.10	0.26	-0.01	0.01
June	30	0.27	0.03	-0.45	0.45
July	31	0.38	0.14	-0.32	0.32
August	31	0.45	0.38	-0.42	0.42
September	30	0.42	0.34	-0.34	0.34
October	31	0.05	0.25	-0.03	0.03
November	30	-0.06	0.40	0.19	0.19
December	31	0.02	0.06	-0.13	0.13

3. RESULTS AND DISCUSSION

Public procurement refers to the systematic acquisition of supplies and services by the public healthcare system from specialized vendors. Typically, a group of authorized providers is established for procurement purposes. When multiple providers offer the same goods, a request for quotation is typically issued, prompting providers to compete with their bids and quotations. Alternatively, direct contracts may be negotiated in certain instances.

In order to reduce costs and avert potential shortages, healthcare institutions typically prepare annual or semiannual procurement plans well in advance. Nevertheless, the emergence of outbreaks and unforeseen events may force planners to conduct monthly reviews of both their present and upcoming requirements. Budgetary constraints often prohibit excessive purchases, yet inadequacy in procurement can have negative consequences. A flawed estimation, therefore, risks either operational inefficiencies or critical deficiencies. In the context of this study, it is assumed that planners routinely conduct monthly reviews of their needs to ensure appropriate resource management. While AI, particularly ML algorithms, has witnessed a surge in popularity in recent years, estimations calculated with LR or MA remain prevalent. Consequently, in this study, forecasts generated by LR models were employed as the baseline for comparison with those produced by SVM models.

Two approaches were investigated concerning the number of observations. Firstly, utilizing the accumulated observations (long memory). Secondly, restricting the observations to the most recent 30 days (short memory). With LR-based forecasts, the smoothing effect is visible when older data are incorporated. As depicted in Figure 5, July's daily LR forecasts exhibit significant deviation from the actual values. Conversely, in Figure 6, July's forecasts show closer alignment with the actual values (depicted by the red dotted line), irrespective of whether accumulated or last month's observations were considered. In both scenarios—whether accumulated or last month's observations-SVM forecasts exhibit less deviation from the actual values compared to LR forecasts.

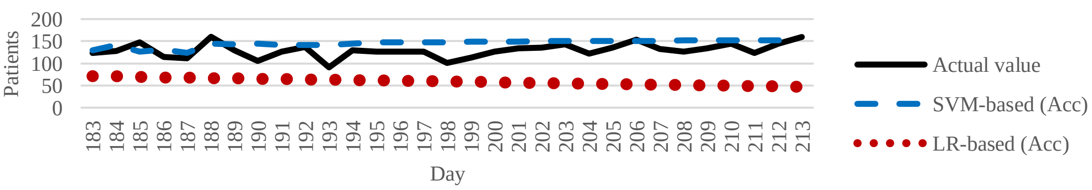


Figure 5. July's forecast with daily accumulated observations from January to June

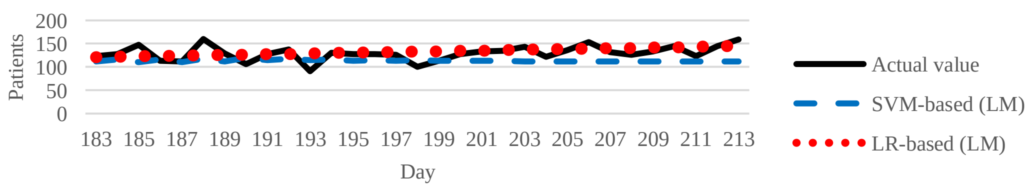
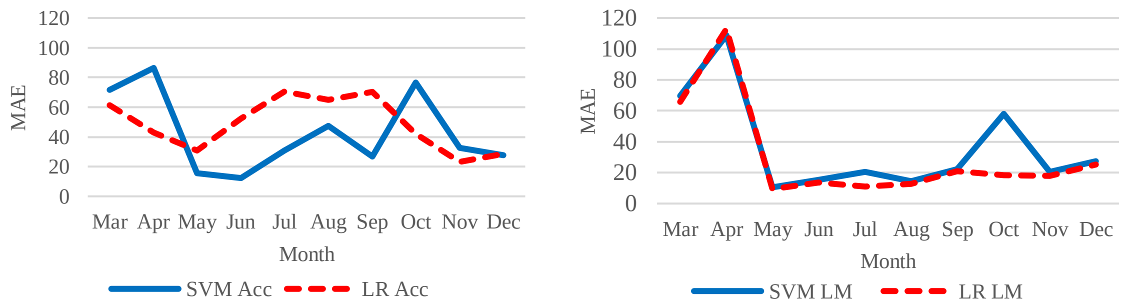


Figure 6. July's forecast with only June's observations (last month)

3.1. Error metrics

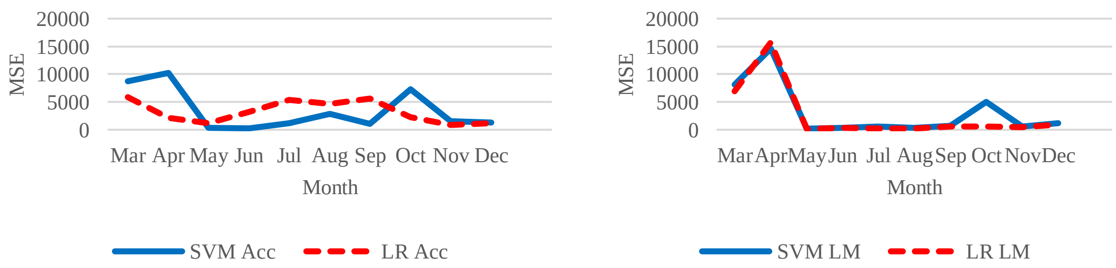
Using the widely accepted error metrics—namely MAE, MSE, RMSE, and MAPE—the forecasts for each month were evaluated and compared. Given the familiarity with these error metrics, their definitions and formulae have been omitted for brevity. Figures 7(a) and (b) illustrates the monthly MAE values obtained using either accumulated or last month's observations. These values denote the difference between the monthly forecasts derived from SVM and LM models. While not definitive, discernible differences are observed, particularly when incorporating all available observations.



(a) (b)

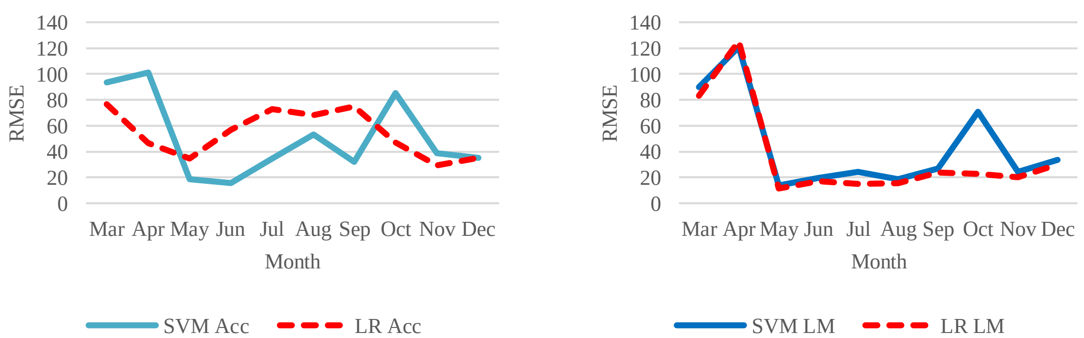
Figure 7. Monthly MAE values: (a) accumulated and (b) last month's observations

Figures 8(a)-(b) and 9(a)-(b) depict the disparity between monthly MSE values, monthly RMSE values, and actual values. Once more, discernible differences are evident. Particularly when employing a comprehensive set of accumulated observations (long memory approach).



(a) (b)

Figure 8. Monthly MSE values: (a) accumulated and (b) last month's observations



(a) (b)

Figure 9. Monthly RMSE values: (a) accumulated and (b) last month's observations

Figures 10(a) and 10(b) illustrates the difference between monthly MAPE values and actual values. It appears evident that, for this specific dataset, there is not a significant difference in error metrics when adopting a short memory approach (last month). However, there is less divergence when a smoothing effect from older data is incorporated (long memory approach).

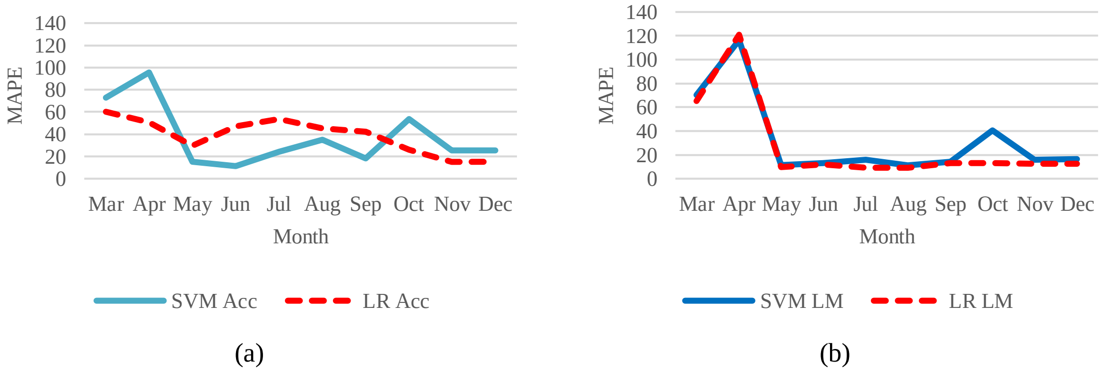


Figure 10. Monthly MAPE values: (a) accumulated and (b) last month's observations

As summarized in Table 5, MAE values do not decisively favor either SVM models or LR models when utilizing a full set of accumulated observations. In 5 out of 10 months, either SVM models outperformed LR models or vice versa. However, with 30 observations, LR models exhibit lower MAE in 9 out of 10 months.

Table 5. Monthly MAE and MSE values with daily accumulated and last month's observations

	MAE				MSE			
	Accumulated obs.		Last month obs.		Accumulated obs.		Last month obs.	
	SVM	LR	SVM	LR	SVM	LR	SVM	LR
March	71.8	61.4	69.5	65.9	8702.8	5860.8	8098.4	6885.2
April	86.4	43.0	108.8	112.9	10224.5	2144.9	14610.1	15723.1
May	15.6	30.7	10.5	9.4	351.5	1180.7	189.4	130.6
June	12.3	52.5	15.1	13.6	242.6	3201.4	379.0	296.5
July	30.7	70.5	20.6	11.2	1209.9	5334.6	580.3	230.5
August	47.6	65.0	14.4	12.6	2829.1	4620.3	346.2	241.0
September	26.7	70.3	22.2	20.8	1015.8	5634.5	725.8	551.1
October	76.4	42.1	58.0	18.5	7250.6	2222.5	4970.7	514.7
November	32.7	23.2	20.6	17.7	1502.3	858.5	591.7	405.5
December	27.5	28.5	27.2	25.3	1234.9	1231.1	1116.4	883.7

As depicted in Table 6, monthly RMSE values reveal an equivalence between SVM and LR models when utilizing a full set of accumulated observations. However, LR models outperform SVM models in 9 out of 10 months when considering the last 30 days. Monthly MAPE values exhibit a similar pattern, with no discernible difference observed when utilizing accumulated observations. However, LR models yield superior results when last month's observations are considered, prevailing in 9 out of 10 months.

Table 6. Monthly RMSE and MAPE values with daily accumulated and last month's observations

	RMSE				MAPE			
	Accumulated obs.		Last month obs.		Accumulated obs.		Last month obs.	
	SVM	LR	SVM	LR	SVM	LR	SVM	LR
March	93.3	76.6	90.0	83.0	72.6	60.2	70.1	65.2
April	101.1	46.3	120.9	125.4	95.6	50.4	116.4	120.9
May	18.7	34.4	13.8	11.4	15.2	29.9	11.3	9.5
June	15.6	56.6	19.5	17.2	11.3	46.8	13.2	12.1
July	34.8	73.0	24.1	15.2	24.4	53.5	16.1	9.2
August	53.2	68.0	18.6	15.5	35.1	45.5	11.2	9.4
September	31.9	75.1	26.9	23.5	18.1	42.2	14.4	13.2
October	85.2	47.1	70.5	22.7	53.7	26.3	40.4	12.9
November	38.8	29.3	24.3	20.1	25.6	15.1	16.0	12.7
December	35.1	35.1	33.4	29.7	25.7	15.2	16.2	12.8

When adopting a slightly different approach and analyzing the averaged value of monthly error metrics, the data revealed that, as illustrated in Table 7, SVM models generated lower errors when utilizing accumulated observations. However, when the number of observations is limited to 30 days, LR models demonstrated lower errors. In general, LR models developed with only 30 observations generated lower averaged error metrics.

Table 7. Averaged monthly error metrics: MAE, MSE, RMSE, and MAPE

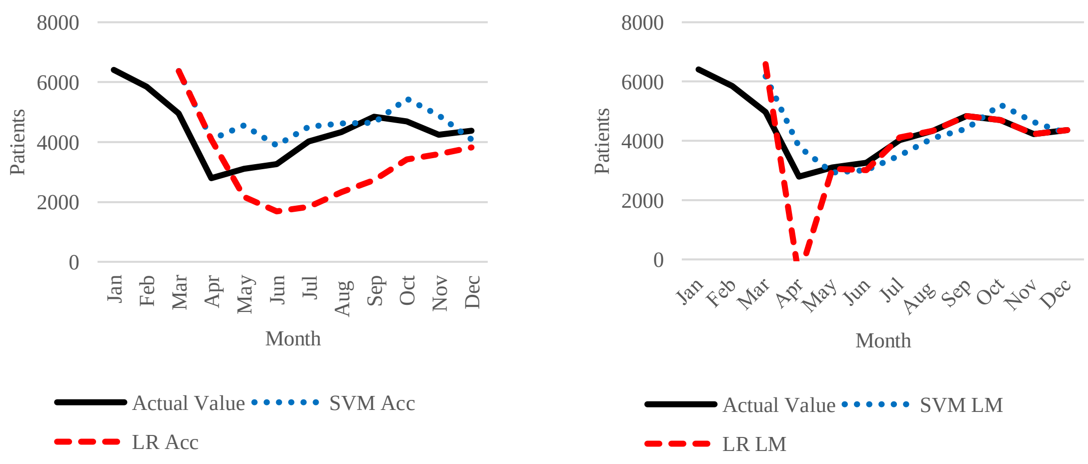
	Accumulated obsservations		Last month´s obsservations	
	SVM	LR	SVM	LR
Averaged MAE	42.8	48.7	36.7	30.8
Averaged MSE	3,456.4	3,228.9	3,160.8	2,586.2
Averaged RMSE	50.8	54.2	44.2	36.7
Averaged MAPE	37.7	38.5	32.5	27.8

3.2. Shortage months

In contrast to previous sections where error metrics were computed based on 30-day forecasts, this section utilizes monthly consolidated summaries instead. This alternative approach enables planners to identify those months in which forecasts underestimate actual values, potentially leading to shortages. In public healthcare systems, while overestimation may incur negative financial implications, shortages resulting from underestimation could have negative consequences.

When the patient influx is aggregated monthly, mirroring the approach planners would take during monthly reviews of current needs, new insights emerge. Utilizing all accumulated observations, SVM models produced only 2 instances of underestimation out of 10 months, contrasting with LR models, which underestimated in 8 out of 10 months. Conversely, when considering only last month's observations, these figures were 6 and 3 out of 10 months, respectively. However, it's important to note that the magnitude of these underestimations was relatively small. Figures 11(a) and 11(b) visually depicts that SVM models forecasted monthly figures closer to the actual values, while LR models consistently tended to underestimate the number of patients. It is essential to clarify that the numbers presented in Table 8 depict the summation of monthly actual values and the corresponding summation of 30-day forecasts conducted using both the accumulated observations and the last month's observations.

Upon closer examination of the percentage difference between actual values and forecasts, it becomes evident that even when SVM models generated underestimations, the magnitude of the difference was relatively small, approximately 4% and 6% in September and December, respectively. Despite the potential misleading nature of error metrics, when data are aggregated on a monthly basis, it is apparent that forecasts derived from SVM models closely align with actual values compared to those produced by LR models. Figures 12(a) and 12(b) illustrates the magnitude (percentage) of the difference between actual values and forecasts, further reinforcing the preference for SVM models.



(a) (b)

Figure 11. Monthly forecast: (a) accumulated and (b) last month’s observations

Table 8. Summary of monthly patient influx: actual values versus forecasts

		Accumulated observations		Last month's observations	
	Actual value	SVM	LR	SVM	LR
March	4965	6371	6368	6177	6588
April	2793	4079	4083	3790	-593
May	3102	4564	2160	2930	3061
June	3261	3886	1685	3010	3019
July	4014	4503	1828	3515	4123
August	4337	4620	2321	4097	4337
September	4836	4638	2727	4416	4836
October	4697	5443	3419	5223	4697
November	4237	4881	3603	4635	4237
December	4372	4097	3815	4259	4372

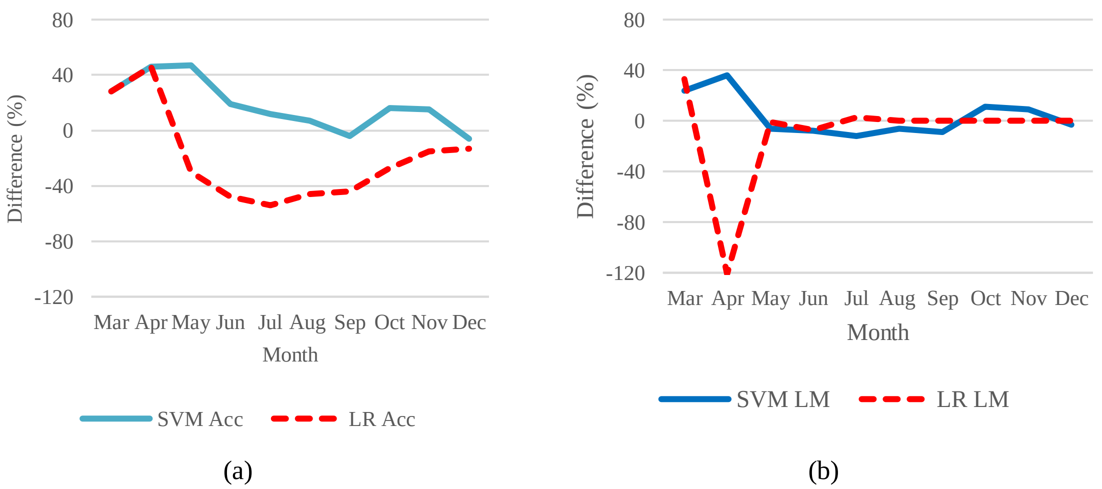


Figure 12. Percentage monthly difference (%): (a) accumulated and (b) last month observations

As mentioned earlier, underestimating the number of patients may result in an underestimation of the requirements for supplies and medical staff, along with its associated implications. Table 9 provides a quantification of the differences illustrated in Figure 12. Despite the notable distortion observed in the first two months, SVM models generated a closer forecast to actual values compared to LR models. It is important to emphasize that all forecasts, summaries, error metrics, and percentage comparisons presented in this study are valid solely for the dataset under analysis, which ultimately determines the mathematical form of the forecast models. Even with the same dataset, altering the number of observations could yield significantly different models.

ML algorithms are known for their sensitivity to input data. While comparing forecasts obtained using different algorithms is undoubtedly intriguing, evaluating forecast accuracy, measured in terms of error metrics, across experiments conducted by independent researchers working with different datasets can be misleading and confusing. Therefore, LR models were utilized as a reference point for comparison in this study.

Table 9. Percentage monthly difference between actual values and forecasts

	Accumulated obs.		Last month obs.	
	SVM (%)	LR (%)	SVM (%)	LR (%)
March	28	28	24	33
April	46	46	36	-121
May	47	-30	-6	-1
June	19	-48	-8	-7
July	12	-54	-12	3
August	7	-46	-6	0
September	-4	-44	-9	0
October	16	-27	11	0
November	15	-15	9	0
December	-6	-13	-3	0

4. CONCLUSION

Forecasting models based on time series have been a subject of study for decades. In recent years, innovative approaches based on ML algorithms have gained acceptation. This study delved into the efficacy of employing SVM-based models for forecasting ER patient influx to estimate the requirements for supplies, medical personnel, and infrastructure. Experimental findings indicate that when error metrics are computed using a 30-day forecast, LR-based models produced lower errors compared to SVM-based models. However, alternative analyses based on monthly consolidated forecasts revealed that underestimations were less likely to occur with SVM-based models, and when they did, the magnitude of the difference between forecast and actual values was typically below 10%. After analyzing over 50,000 ER patient records, it can be concluded that implementing SVM-based forecast models could aid planners in estimating the needs for supplies and medical staff during the study period. Furthermore, utilizing a reduced set of recent observations (last month) led to more accurate estimations.

REFERENCES

[1] O. Yanmaz and O. Kabak, "A two phase approach based on multi objective programming and simulation for physician scheduling in emergency rooms," *International Journal of Applied Decision Sciences*, vol. 1, no. 1, 2023, doi: 10.1504/IJADS.2023.10049933.

[2] T. Susnjak and P. Maddigan, "Forecasting patient demand at urgent care clinics using explainable machine learning," *CAAI Transactions on Intelligence Technology*, vol. 8, no. 3, pp. 712–733, 2023, doi: 10.1049/cit2.12258.

[3] M. Afilal, F. Yalaoui, F. Dugardin, L. Amodeo, D. Laplanche, and P. Blua, "Forecasting the emergency department patients flow," *Journal of Medical Systems*, vol. 40, no. 7, 2016, doi: 10.1007/s10916-016-0527-0.

[4] R. D. Riley *et al.*, "A guide to systematic review and meta-analysis of prognostic factor studies," *BMJ*, vol. 364, 2019, doi: 10.1136/bmj.k4597.

[5] Organisation for Economic Co-operation and Development, "Health spending," *OECD*, pp. 120–121, 2016, doi: 10.1787/soc_glance-2016-24-en.

[6] A. C. Alba *et al.*, "Discrimination and calibration of clinical prediction models: Users' guides to the medical literature," *JAMA - Journal of the American Medical Association*, vol. 318, no. 14, pp. 1377–1384, 2017, doi: 10.1001/jama.2017.12126.

[7] T. Panch, P. Szolovits, and R. Atun, "Artificial intelligence, machine learning and health systems," *Journal of Global Health*, vol. 8, no. 2, 2018, doi: 10.7189/jogh.08.020303.

[8] G. S. Collins and K. G. M. Moons, "Reporting of artificial intelligence prediction models," *The Lancet*, vol. 393, no. 10181, pp. 1577–1579, 2019, doi: 10.1016/S0140-6736(19)30037-6.

[9] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement," *Circulation*, vol. 131, no. 2, pp. 211–219, 2015, doi: 10.1161/CIRCULATIONAHA.114.014508.

[10] N. Shahid, T. Rappon, and W. Berta, "Applications of artificial neural networks in health care organizational decision-making: A scoping review," *PLoS ONE*, vol. 14, no. 2, 2019, doi: 10.1371/journal.pone.0212356.

[11] S. Goyal, N. Batra, and K. Chhabra, "Diabetes disease diagnosis using machine learning approach," *Lecture Notes in Networks and Systems*, vol. 473, pp. 229–237, 2023, doi: 10.1007/978-981-19-2821-5_19.

[12] N. H. Harlapur and V. Handur, "Text-based prediction of heart disease doctor chatbot using machine learning," *Smart Innovation, Systems and Technologies*, vol. 311, pp. 231–245, 2023, doi: 10.1007/978-981-19-3571-8_24.

[13] A. Al Ahdal and P. Chawla, "A review on machine learning-based patient scanning, visualization, and monitoring," *Lecture Notes in Networks and Systems*, vol. 473, pp. 487–497, 2023, doi: 10.1007/978-981-19-2821-5_40.

[14] R. A. Taylor *et al.*, "Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach," *Academic Emergency Medicine*, vol. 23, no. 3, pp. 269–278, 2016, doi: 10.1111/acem.12876.

[15] S. Kejriwal and N. Rajagopalan, "A technical review on machine learning-based prediction on COVID-19 diagnosis," *Smart Innovation, Systems and Technologies*, vol. 311, pp. 529–539, 2023, doi: 10.1007/978-981-19-3571-8_50.

[16] J. H. Chen and S. M. Asch, "Machine learning and prediction in medicine-beyond the peak of inflated expectations," *New England Journal of Medicine*, vol. 376, no. 26, pp. 2507–2509, 2017, doi: 10.1056/nejmp1702071.

[17] J. Miles, J. Turner, R. Jacques, J. Williams, and S. Mason, "Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review," *Diagnostic and Prognostic Research*, vol. 4, no. 1, 2020, doi: 10.1186/s41512-020-00084-1.

[18] A. Pak, B. Gannon, and A. Staib, "Predicting waiting time to treatment for emergency department patients," *International Journal of Medical Informatics*, vol. 145, 2021, doi: 10.1016/j.ijmedinf.2020.104303.

[19] C. Curtis, C. Liu, T. J. Bollerman, and O. S. Pianykh, "Machine learning for predicting patient wait times and appointment delays," *Journal of the American College of Radiology*, vol. 15, no. 9, pp. 1310–1316, 2018, doi: 10.1016/j.jacr.2017.08.021.

[20] X. Ma, J. Wu, and X. Xue, "Identification of DNA-binding proteins using support vector machine with sequence information," *Computational and Mathematical Methods in Medicine*, vol. 2013, 2013, doi: 10.1155/2013/524502.

[21] F. Jiang *et al.*, "Artificial intelligence in healthcare: Past, present and future," *Stroke and Vascular Neurology*, vol. 2, no. 4, pp. 230–243, 2017, doi: 10.1136/svn-2017-000101.

[22] J. K. H. Andersen *et al.*, "Neural networks for automatic scoring of arthritis disease activity on ultrasound images," *RMD Open*, vol. 5, no. 1, 2019, doi: 10.1136/rmdopen-2018-000891.

[23] H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," *Journal of Medical Systems*, vol. 41, no. 4, 2017, doi: 10.1007/s10916-017-0703-x.

[24] K. J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1–2, pp. 307–319, 2003, doi: 10.1016/S0925-2312(03)00372-2.

[25] L. Cao, "Support vector machines experts for time series forecasting," *Neurocomputing*, vol. 51, pp. 321–339, 2003, doi: 10.1016/S0925-2312(02)00577-5.

[26] L. J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1506–1518, 2003, doi: 10.1109/TNN.2003.820556.

[27] J. V. Hansen, J. B. McDonald, and R. D. Nelson, "Some evidence on forecasting time-series with support vector machines," *Journal of the Operational Research Society*, vol. 57, no. 9, pp. 1053–1063, 2006, doi: 10.1057/palgrave.jors.2602073.

[28] T. Morzadec *et al.*, "Forecasting the patients flow at pediatric emergency departments," *Journal of Medical Systems*, vol. 45, no. 3, 2021, doi: 10.1007/s10916-021-01712-z.
[29] L. M. Schweigler, J. S. Desmond, M. L. McCarthy, K. J. Bukowski, E. L. Ionides, and J. G. Younger, "Forecasting models of emergency department crowding," *Academic Emergency Medicine*, vol. 16, no. 4, pp. 301–308, 2009, doi: 10.1111/j.1553-2712.2009.00356.x.
[30] S. McRae, "Long-term forecasting of regional demand for hospital services," *Operations Research for Health Care*, vol. 28, 2021, doi: 10.1016/j.orhc.2021.100289.
[31] X. Zhao *et al.*, "A deep learning architecture for forecasting daily emergency department visits with acuity levels," *Chaos, Solitons and Fractals*, vol. 165, 2022, doi: 10.1016/j.chaos.2022.112777.

BIOGRAPHIES OF AUTHORS

Carlos Hernandez is an industrial engineer, consultant, and university professor. He holds a Master of Science in Engineering and a Doctorate of Engineering from Technische Universität Braunschweig in Brunswick, Germany. He has delivered lectures on subjects such as Discrete Event Simulation, Engineering Economics, Corporate Finance, Data Mining, and Machine Learning for engineering students. He has worked for multinational companies including PricewaterhouseCoopers, BHP Billiton, and Merck Sharp & Dohme. Additionally, he was part of the scientific research group at the Institut für Produktionsmesstechnik at TU Braunschweig, Germany. His research interests include manufacturing process simulation, supply chain design and simulation, as well as machine learning for financial applications. She can be contacted at email: carlos.hernandez.zavala@uct.cl.

Dafne Lagos is an industrial engineer and university professor. She holds a Bachelor of Science in Engineering and a Master in Management from Universidad de La Frontera in Temuco, Chile, as well as a Doctor of Science from Atlantic International University in the USA. She has delivered lectures in Engineering Economics, Probability Modeling, Logistics, and Operations Research. She has authored scientific articles in the field of operations research. In addition, she has held various managerial roles and has provided mentorship to over forty students during her academic tenure. Her research interests include financial planning, supply chain management, and logistics. She can be contacted at email: dlagos@uct.cl.

Paola Leal 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 delivered lectures in Operations Research, Logistics, and Supply Chain Management for engineering students. Throughout her academic career, she has held various management positions and has provided mentorship to over sixty students. Her research interests lie in the areas of logistics, optimization, and supply chain management. She can be contacted at email: pleal@uct.cl.

Jaime Castillo is an industrial engineer, consultant, and university professor. He holds a Licentiate Degree in Forest Engineering from Universidad de Concepción in Concepción, Chile, and a Master in Industrial Engineering from Universidad del Desarrollo in Santiago, Chile. He has delivered lectures in Project Planning & Management, Evaluation, Decision Theory, and Process Simulation for engineering students. His extensive professional background includes working on sustainable development projects within the local forest industry. He has held various management positions and has provided mentorship to over fifty students. His research interests include project management, logistic risk assessment, and decision theory. He can be contacted at email: jcastill@uct.cl.