Original Article

Predicting Patient Length of Stay in a Neurosurgical Intensive Care Unit of a Large Teaching Hospital

Behrouz Alizadeh Savareh[1], Ahmad Alibabaei[2,3], Soleiman Ahmady[1,3], Majid Mokhtari[4], Mohammadreza Hajiesmaeili[2], Saeedeh Nateghinia[2*]

Abstract

Background: The intensive care unit (ICU) has the highest mortality and admission rates compared to other wards. Therefore, to increase the performance of hospital services, it is very important to evaluate indicators such as mortality and length of stay of patients in ICU. The present study aimed to investigate the neural network analysis method and Particle Swarm Optimization - Support Vector Machine to predict the length of stay in the neurosurgical intensive care unit.

Materials and Methods: This descriptive research deals with data mining and modeling of intensive care unit processes, leading to a practical example of the application of health systems engineering knowledge, using MATLAB software. Data of 1200 patients admitted during the years 2017 to 2019 in the intensive care unit of neurosurgery. Then we evaluated all data with SVM + PSA and NCA.

Results: Identifying the important features and using them has gradually reduced the LOS prediction error from 40% to 7%. Using the NCA technique makes better results for predicting ICU LOS.

Conclusion: PSO + SMV in addition to NCA is a good predictor of ICU LOS screening in patients after neurosurgery and can provide more accurate prognostic factors.

Keywords: Neuro-ICU, Length of Stay, PSO, SVM, feature selection

1. National Agency for Strategic Research in Medical Education, Tehran, Iran
2. Anesthesiology and Critical Care Department, Critical Care Quality Improvement Research Center, Loghman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3. Virtual School of Medical Education and Management, Shahid Beheshti University of Medical Sciences, Tehran, Iran
4. Department of Pulmonary and Critical Care Medicine, Loghman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Corresponding Author: Saeedeh Nateghinia, Skull Base Research Center, Loghman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
Email: s.nateghinia@gmail.com

Please cite this article as: Alizadeh Savareh B, Alibabaei A, Ahmady S, Mokhtari M, Hajiesmaeili MR, Nateghinia S. Predicting Patient Length of Stay in a Neurosurgical Intensive Care Unit of a Large Teaching Hospital. J Cell Mol Anesth. 2021;6(2):132-40. DOI: https://doi.org/10.22037/jcma.v6i2.32395

Introduction

The intensive care unit (ICU) is one of the most specialized and costly parts of a hospital (1) in comparison to other sections, it has one of the highest mortality rates and admits a relatively large number of patients (2). On the other hand, as hospitals are always struggling to improve their services and reduce costs, ICUs are constantly working on the evaluation, comparison, and improvement of their performance to achieve these goals. For this purpose, measurements of outcome indices, including hospital mortality and length of stay, are usually carried out (3). A patient's length of stay (LOS) in a hospital or any of the general or specialized sections is an index that can be expressed as the days of admission, and it is usually reported as an average. This index may reflect the number of resources used in a hospital (4).

Hospital LOS is one of the most useful indices for various purposes, such as hospital care management, quality control, appropriateness of using hospital services, hospital planning, determination of efficiency, and the rate of using hospital resources. Therefore, due to the underlying role of the patients' ICU LOS, this index has been well received by researchers; however, various studies have been conducted worldwide to predict the ICU LOS moe appropriately. The exact prediction of patients' stay in ICUs enables doctors to provide more accurate information for patient satisfaction.

In addition, doctors can more accurately and deliberately regulate patient care plans and provide more assistance to the authorities and planners to determine, prepare, and allocate financial resources. Moreover, this prediction enables doctors to calculate the length of stay adjusted to risk, and calculate and compare it in and among organizations before and after making changes in everyday hospital management (5). This prediction is an important tool for better service delivery and more patient satisfaction.

Hence, designing accurate and reliable models to predict the length of patients' stay is not only helpful for hospital management, but also for prioritizing macro-policies in the health sector, improving the quality of healthcare services, proper prioritization of resource allocation based on the difference in the patients' length of stay on the one hand, and simultaneous attention to the status of the patients' economic and social indices at a micro-level, on the other hand (6).

Neurosurgical patients may have higher LOS as well as higher readmission rates when compared to the non-neurosurgical patients with a similar overall health condition. In the neurosurgical literature, there is growing interest in predicting outcomes of surgeries and postoperative care, but we are lacking adequate supporting data (7, 8).

Modeling ICU LOS as an outcome variable is complex (9). Different methods have been used to predict LOS including, arithmetic models, statistical methods, multi-stage models, and data-driven approaches. Most LOS prediction research is based on numerical values, however, by a massive increase in the use of hospital information systems, a large amount of electronic data had been made accessible for analysis of patient clinical information. As a result, utilizing data mining and machine learning methods has increased in recent years to help improve healthcare performance. (10-12)

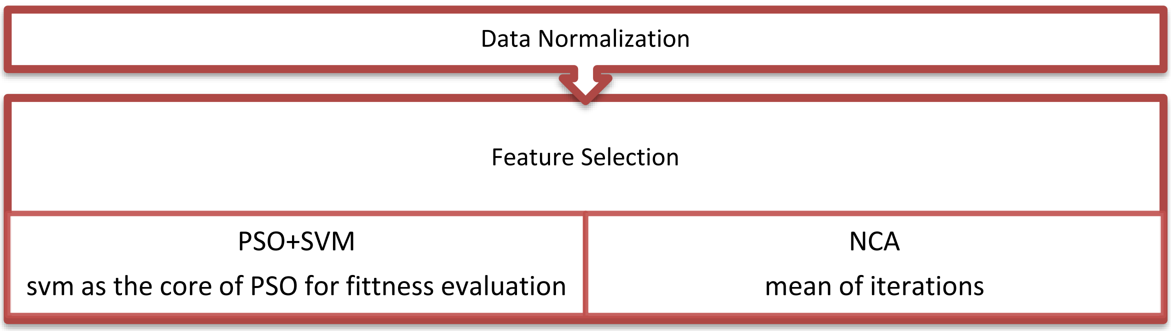
In this research, we used a data-driven approach, neural network analysis, and PSO-SVM (Particle Swarm Optimization - Support Vector Machine) feature selection method (9) to predict neurosurgical ICU LOS.

Methods

The present study is a descriptive applied research in the field of data mining and modeling of intensive care unit processes and a practical example of the application of health systems engineering knowledge that using MATLAB software. Data of 1200 patients admitted during the years 2017 to 2019 In the intensive care unit of neurosurgery, Loghman Hakim Educational, and Medical Center affiliated to Shahid Beheshti University of Medical Sciences, Tehran has been performed.

To create a predictive model for a patient’s length of stay in neurosurgical ICU, we took the steps illustrated in the following flowchart. The implementation of these steps was carried out entirely in MATLAB 2019.

One of the limitations faced by the researchers in this study was the low quality and quantity of data in the computer records of the hospital patients under study. About the quality of the data, because the values of most variables are entered into the system manually, human errors in recording reduce the accuracy of the data, which will have a great impact on the results of the study.



Normalization

Normalization is an important step of data mining applied for the preparation of data when features have different ranges of values. Normalization has a positive effect on the accuracy of the models because mathematical techniques used in data mining have an assumption of values fall between 0 and 1 or -1 to 1. Normalization changes the values to equal scale, without losing the useful information. Min-Max Normalization is commonly used which normalizes values between 0 and 1, as the following formula:

$$\boldsymbol{X\_norm} = \frac{\boldsymbol{X\_org - Min}}{\boldsymbol{Max - Min}}$$

Eq1. MinMax normalization

Feature Selection: In this step, the most useful features in the dataset were selected, based on their role in the output prediction. The goal of this technique is to decrease training time, increase model accuracy and interpretability, and generalization of performance on the test set. To select the most useful features in the dataset, two paths were considered; 1- a combination of Particle Swarm Optimization (PSO) and Support Vector Machine (SVM) used to select features and 2- iterations of NCA.

First path: a combination of PSO and SVM; Particle Swarm Optimization (PSO)**:** as the number of features grows, choosing the more important features with the lowest classification error requires exponential order of computations. Because the problem space has 22 dimensions (features representing neurosurgical ICU patient's status), selecting the most optimal subset requires a high number of calculations. The computation burden for these features is in the exponential order $(2^{22} * classification\ train/test\ time)$.

Therefore, an appropriate approach in this situation is selecting more important features using evolutionary algorithms such as Particle Swarm Optimization (PSO). Its basic principle was inspired by observing the birds flying patterns while searching for food. We applied PSO in this study to avoid exhaustive ***computation***. The method searches the problem space based on the location and speed of the particles. During the development of several steps, only the most optimist particle can transmit information into the others particles; hence the convergence speed is remarkably high. In addition, the calculation in PSO is quite simpler than other methods and can be completed with more speed and ease. (13).

Every particle in the population has two vectors; the velocity vector and the position vector (14). The PSO prompts social search behavior among particles in the ***space***, where every particle indicates one point in n-dimensional space as a possible solution for a given problem. In comparison with other evolutionary algorithms such as the Genetic Algorithm (GA), the PSO has improved search efficacy with a faster and more stable convergence manner. (15) The approach for PSO used in this study is based on the standard PSO (16), but according to our particular situation, the following changes were made. Feature selection was encoded in the binary format, and in the evolutionary iterations, the PSO tries to find a binary code that minimizes the classification error in the LOS classifications. Then, features that have a stronger correlation with LOS classification were identified through the PSO iterations. The PSO selected the more important features and injected them for fitness evaluation (LOS classification by SVM).

Support Vector Machine (SVM): Support Vector Machine (SVM) is one of the most efficient machine learning techniques that has a vast variety ***of*** applications; such as classification and regression applications (17). There are three major advantages of SVM; having a few numbers of setting parameters (easy for tuning), optimal for solving a linearly constrained quadratic problem, and good generalization performance (18). These were the reasons we were persuaded to use SVM in the core of fitness function in LOS (short or long) classification.

PSO + SVM combination: The figure1 shows the PSO+SVM combination for selecting the features. As illustrated in the figure, PSO controls the overall process of the feature selection and SVM serves the fitness evaluation inside the PSO.

Some terms should be addressed about the figure. As any particle in PSO is a candidate solution (binary coding of feature selection), *pbest* indicates a personal best solution for any particles achieved, and *gbest* standing for general best indicates the best solution for all of the particles. As mentioned in the flowchart above, SVM evaluation was considered as the core of decision-making about the fitness of possible solutions. The following are some of the settings used in the PSO+SVM combination.

- options = optimoptions ('particleswarm','Display','iter','SwarmSize',50,'MaxIterations',1000,'PlotFcn','pswplotbestf');

With the above settings, PSO is searching for the reduction of error (10-fold SVM classification error). In the above process, the features that play an important role in the prediction of LOS are identified. By considering the successive rounds of PSO, effective features in predicting LOS are identified.

NCA: After dimensionality reduction using the PSO+SVM algorithm, the next step involved reducing the feature number with the help of the Neighboring Component Analysis (NCA). In NCA, the importance of each feature is calculated based on its power in the prediction of the output. NCA learns features' importance by maximizing the expected leave-one-out (LOO) classification accuracy (19). As the real distribution of data is unknown, NCA attempts to optimize the performance based on the training data. The algorithm is restricted to find quadratic distance metrics. It can always be represented by symmetric positive semi-definite matrices. If it is denoted by a transformation matrix A, a metric is effectively learned as $Q = A > A$ as equation

$$d(x,y) = (x - y) > Q(x - y) = (Ax - Ay) > (Ax - Ay)$$

Eq. 1. Q matrix calculation in NCA algorithm

The goal of this algorithm is to reduce f(A), which is defined by Eq. 2.

$$f(A) = \sum_{i} \sum_{j \in Ci} pij = \sum_{i} pi$$

Eq. 2. f(A): NCA maximization goal (class separability).

NCA can surpass local minima based on its non-convex computational core. Therefore, using NCA in the feature selection guarantees that the achieved results were not related to the initial starting points. Moreover, using a high number of NCA iterations and calculations can imply that the results are deterministic absolutely.

Results

The followings are the simulation results of using PSO+SVM (decreasing prediction error) (Chart 1). Identifying the important features and using them has gradually reduced the LOS prediction error from 40% to 7%. Due to the nature of the PSO (rapid convergence in the low number of rounds), it has achieved high accuracy in LOS prediction in just 60 rounds. As in the figure below, after successive rounds of PSO execution, the error variance between the CrossValidation train/test sets gradually decreased and the model became more stable in the last rounds of PSO (Chart 2)

Using the NCA technique to sort the features based on their impact in LOS prediction yielded the following results. What follows is the average NCA output values per 1000 iterations (Chart 3).

Discussion

Using the combination of PSO+SVM and NCA, as the results of the study show, lead to an appropriate feature selection result. Considering types of feature selection

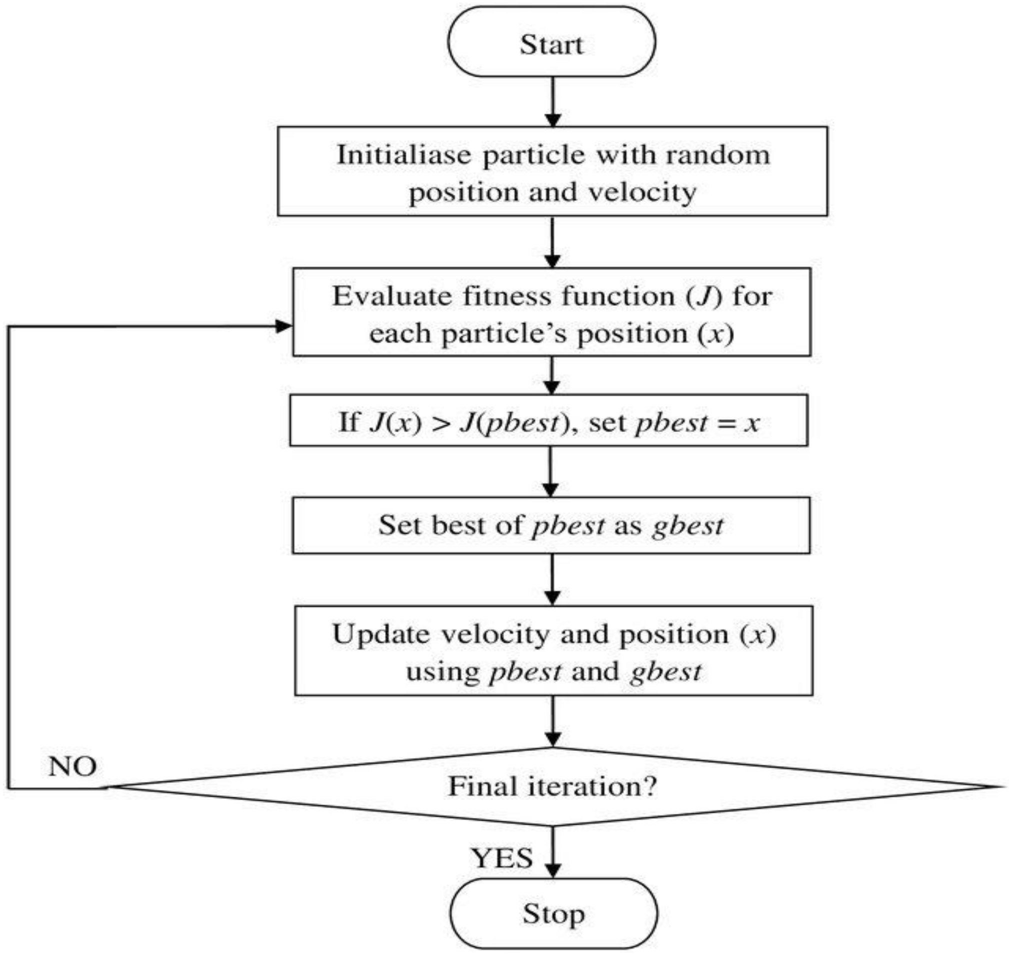


Figure 1. PSO +SVM combination.

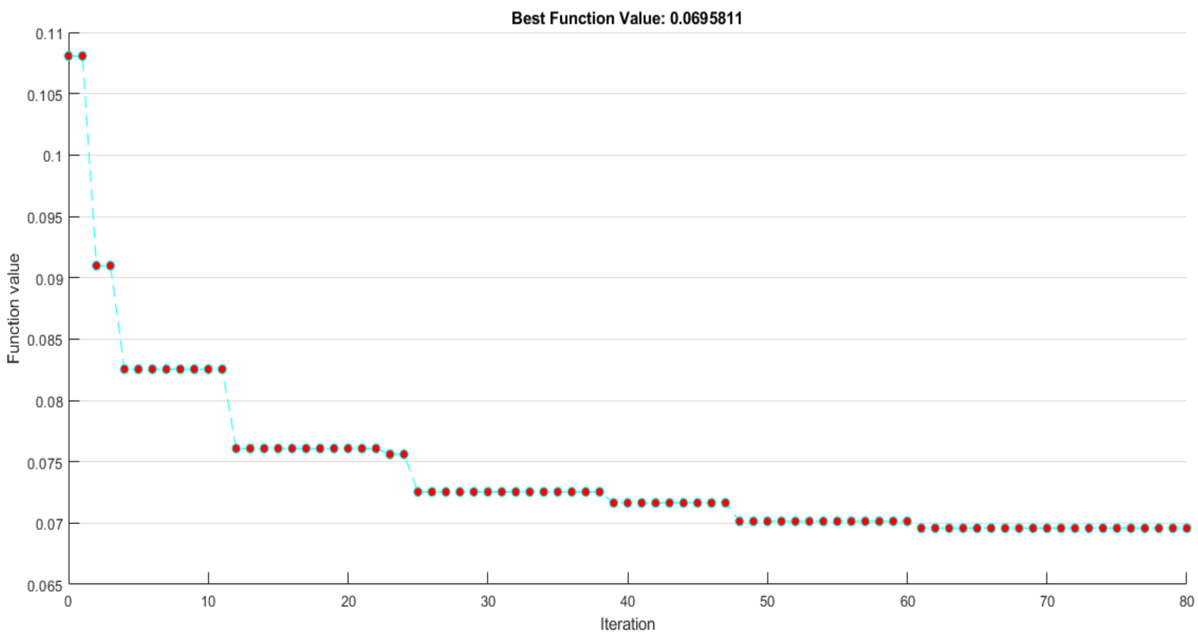


Chart 1. The simulation results of using PSO+SVM.

techniques, our method, as wrapper one, is essentially capable of solving the "real" problem (optimizing the

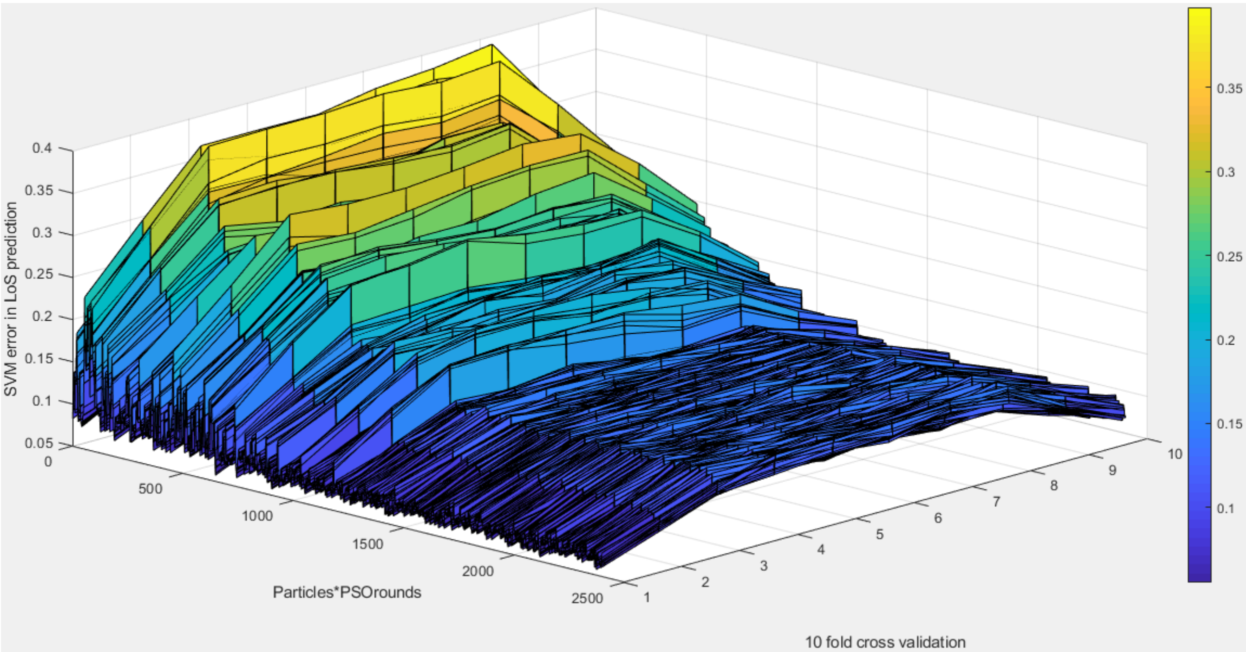


Chart 2. The error variance between the CrossValidation train/test sets.

SVM performance in LOS prediction). But the wrapper methods are computationally more expensive compared to other feature selection methods, due to the repeated learning steps and cross-validation. Given the high processing power of today's hardware, the processing burden imposed by this method is bearable and is not problematic in non-super-sized data (such as ours).

In the present study, it was found that the most common factors associated with ICU LOS were

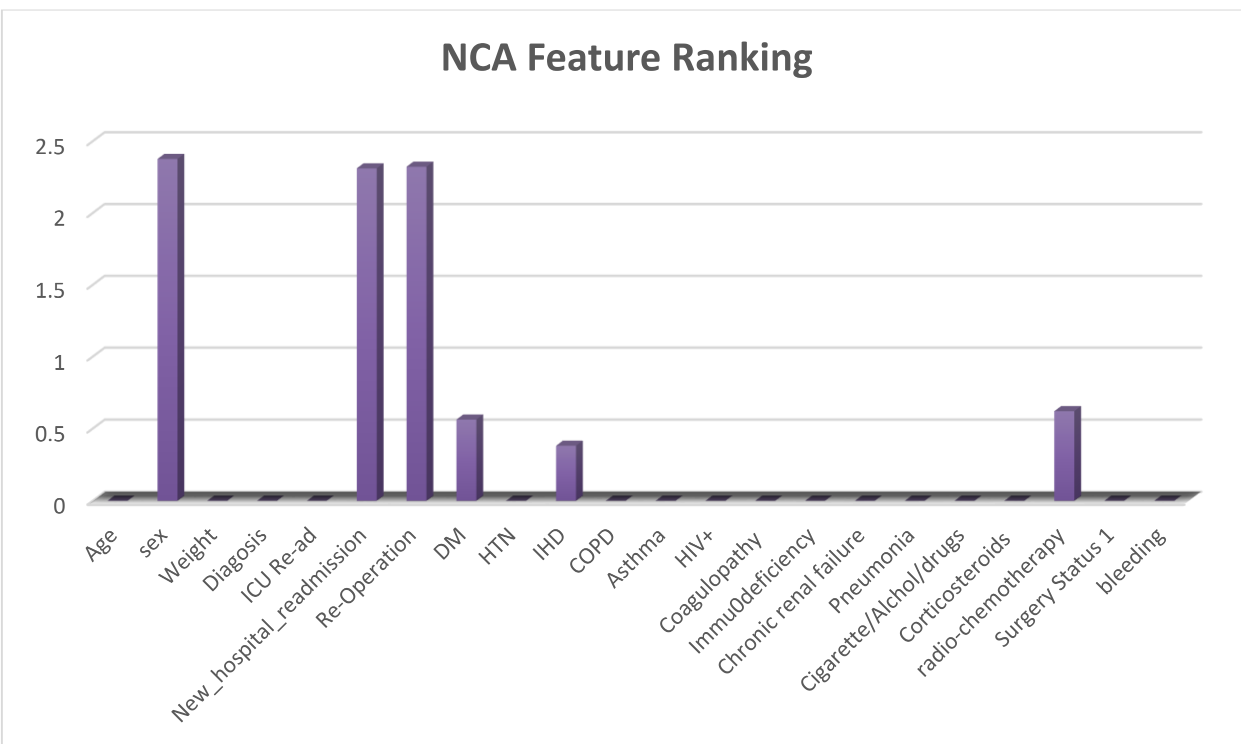


Chart 3. Ranking of NCA features.

gender, reoperation, and readmission in the ICU, followed by diabetes, ischemic heart disease, and chemotherapy. A 2015 study by Junior et al found that ICU LOS was 29% under-predicted and overestimated at 18% and Gender, admission status for surgery, chemotherapy, mechanical ventilation, vasopressor use, and active infection at the time of admission were associated with the lowest ICU LOS assessment. Males and active infection at admission were independently associated with low ICU LOS estimates in logistic regression. Type of hospitalization (immediate medical surgery), the reason for admission (no postoperative supervision), chemotherapy, mechanical ventilation, vasopressor use, delirium and infectious status at admission, serum creatinine, and readmission were associated with ICU LOS overestimation. The type of admission, the reason for admission to chemotherapy, and active infection at admission were independently associated with the overestimation of the LOS ICU. This study is similar to the present study in several factors, including chemotherapy, ICU readmission, and gender impact, but differs from other factors in that it may be due to the moderating effect of the NCA (20). In the study of Houthooft et al., It was found that the best model for predicting patient mortality and a long stay in the ICU is SVM with GA, D = 65.9% and GS, L = 73.2%.

In terms of LOS regression, the best regression performance model is vector regression, which achieves a mean absolute error of 1.79 days and a mean absolute error of 1.22 days for those patients who do not stay in the hospital for a long time. In the present study, it was found that using SMV for ICU LOS is a good method for a long time, but the difference between the present study and the Houthooft study is that SMV correction was performed with PSO and it was seen that the combination of these two methods causes better SMV performance and reduces its error (21).

In the study of Vieira et al., It was found that using the PSO method to predict ICU LOS is a good method that is similar to the current study. In this study, it was found that the best method for predicting ICU LOS is the modified binary particle swarm optimization method which optimizes the PSO well. In the present study, it was found that the best method is PSO + SMV, which makes a more accurate prediction of ICU LOS. Of course, one of the differences between that study and the current study is that the Vieira study was performed on ICU patients suffering from septic shock. The present study examined patients who were admitted to the ICU after neurosurgery, which may also cause differences in results (22).

In Maharlou et al.'s study, it was found that the conclusion of the fuzzy neural adaptive algorithm produces a more accurate model because it applies both the neural network architecture capabilities and the knowledge of experts as a hybrid algorithm. It identifies nonlinear components and provides significant results for predicting length of stay, which is a useful calculation to support ICU management, to identify higher-quality management, and cost reduction. In the present study, it was found that the best method for ICU management is the combination of PSO + SMV methods with NCA, which makes a detailed study of the effective factors in predicting the amount of ICU LOS and reduces costs and increases the quality and quantity of ICU services (23). In the study of Rooij et al., It was seen one of the important factors of ICU LOS is age and comorbidity, and disease severity, which is similar to the current study in terms of age and comorbidity and type of comorbidity, but in that study, the effect of reoperation and ICU readmission on ICU LOS was not seen, but in the present study it was found that these two factors are very effective factors that have been diagnosed by the NCA and the value of these factors is higher than sex and comorbid disease. This also indicates the greater impact of the use of NCA in the study of factors affecting ICU LOS (24).

In this study, it was seen that correcting and adapting the effective factors in increasing the duration of hospitalization, gender, and reoperation, and readmission in the ICU has the greatest effect on increasing the length of stay in the ICU of patients. In the study of Stecker et al., It was found that hospitalization in the ICU is one of the factors that affect the increase in the length of hospitalization of patients. Other factors include DVT, urinary tract infection, and intubation (25). In the study of Lazaridis et al., It was seen that age and sex are important factors in increasing the length of ICU stay, while in this study it was seen that sex is an effective factor but age is not a relevant factor (26). It seems that readmission increases the length of hospital stay because it exposes the patient to various nosocomial infections, such as urinary tract infection, which is the most common nosocomial infection. These infections are also more common in women, and readmission may increase the length of time patients stay in the hospital. On the other hand, reoperation is often associated with problems such as bleeding and death (27). Lack of management of these problems in patients increases the need for patients to stay longer in the ICU. Patients undergoing chemotherapy and diabetics are more likely to stay in the hospital due to a weakened immune system and susceptibility to nosocomial infections, especially urinary tract infections, who need more extensive treatment to return to recovery. Patients with IHD may also have problems with treatment due to the heart's inability to receive enough fluids, which requires increasing the length of hospital stay for long-term treatment due to treatment limitations.

Conclusion

PSO + SMV in addition to NCA is a good predictor of ICU LOS screening in patients after neurosurgery and can provide more accurate prognostic factors.

Acknowledgment

None.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

1. Farzanegan B, Zangi M. Predictor factors for sepsis diagnosis, length of ICU stay and mortality in ICU. J cell mol Anesth. 2017;2(2):55-62.
2. Rahmatizadeh S, Valizadeh-Haghi S, Dabbagh A. The Role of Artificial Intelligence in Management of Critical COVID-19 Patients. J Cell Mol Anesth. 2020;5(1):16-22.
3. Verburg IW, de Keizer NF, de Jonge E, Peek N. Comparison of regression methods for modeling intensive care length of stay. PLoS One. 2014;9(10):e109684.
4. Maharlou H, Niakan Kalhori SR, Shahbazi S, Ravangard R. Predicting Length of Stay in Intensive Care Units after Cardiac Surgery: Comparison of Artificial Neural Networks and Adaptive Neuro-fuzzy System. Healthc Inform Res. 2018;24(2):109-17.
5. Widyastuti Y, Stenseth R, Wahba A, Pleym H, Videm V. Length of intensive care unit stay following cardiac surgery: is it impossible to find a universal prediction model? Interact Cardiovasc Thorac Surg. 2012;15(5):825-32.
6. Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. Healthc Inform Res. 2013;19(2):121-9.
7. Ansari SF, Yan H, Zou J, Worth RM, Barbaro NM. Hospital Length of Stay and Readmission Rate for Neurosurgical Patients. Neurosurgery. 2018;82(2):173-81.
8. Missios S, Bekelis K. Association of Hospitalization for Neurosurgical Operations in Magnet Hospitals With Mortality and Length of Stay. Neurosurgery. 2018;82(3):372-7.
9. Litton E, McCann M, van Haren F. Predicting Intensive Care Unit Length of Stay After Cardiac Surgery. J Cardiothorac Vasc Anesth. 2018;32(6):2683-4.
10. Awad A, Bader-El-Den M, McNicholas J. Patient length of stay and mortality prediction: A survey. Health Serv Manage Res. 2017;30(2):105-20.
11. Alamdari NM, Afaghi S, Rahimi FS, Tarki FE, Tavana S, Zali A, et al. Mortality Risk Factors among Hospitalized COVID-19 Patients in a Major Referral Center in Iran. Tohoku J Exp Med. 2020;252(1):73-84.
12. Rajaei S, Dabbagh A. Risk Factors for Postoperative Respiratory Mortality and Morbidity in Patients Undergoing Coronary Artery Bypass Grafting. Anesth Pain Med. 2012;2(2):60-5.
13. Freitas D, Lopes LG, Morgado-Dias F. Particle Swarm Optimisation: A Historical Review Up to the Current Developments. Entropy (Basel). 2020;22(3).
14. Nagaoka Y, Miyazaki T, Sugaya Y, Omachi S. Text Detection Using Multi-Stage Region Proposal Network Sensitive to Text Scale. Sensors (Basel). 2021;21(4).
15. Lv X, Wang Y, Deng J, Zhang G, Zhang L. Improved Particle Swarm Optimization Algorithm Based on Last-Eliminated Principle and Enhanced Information Sharing. Comput Intell Neurosci. 2018;2018:5025672.
16. Zhang R, Zhang Y, Sun J, Li Q. Pattern Synthesis of Linear Antenna Array Using Improved Differential Evolution Algorithm with SPS Framework. Sensors (Basel). 2020;20(18).
17. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics Proteomics. 2018;15(1):41-51.
18. Kuusisto F, Santos Costa V, Nassif H, Burnside E, Page D, Shavlik J. Support Vector Machines for Differential Prediction. Mach Learn Knowl Discov Databases. 2014;8725:50-65.
19. Sun Y, Todorovic S, Goodison S. Local-learning-based feature selection for high-dimensional data analysis. IEEE Trans Pattern Anal Mach Intell. 2010;32(9):1610-26.
20. N Junior AP, Caruso P. Analysis of ICU physician predictions for

ICU length of stay. Crit Care. 2015;19(Suppl 2):P17-P.
21. Houthooft R, Ruyssinck J, van der Herten J, Stijven S, Couckuyt I, Gadeyne B, et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. Artif Intell Med. 2015;63(3):191-207.
22. Teng AK, Wilcox AB. A Review of Predictive Analytics Solutions for Sepsis Patients. Appl Clin Inform. 2020;11(3):387-98.
23. Sun LY, Bader Eddeen A, Ruel M, MacPhee E, Mesana TG. Derivation and Validation of a Clinical Model to Predict Intensive Care Unit Length of Stay After Cardiac Surgery. J Am Heart Assoc. 2020;9(21):e017847.
24. de Rooij SE, Abu-Hanna A, Levi M, de Jonge E. Factors that predict outcome of intensive care treatment in very elderly patients: a review. Crit Care. 2005;9(4):R307-14.
25. Stecker MM, Stecker M, Falotico J. Predictive model of length of stay and discharge destination in neuroscience admissions. Surg Neurol Int. 2017;8:17.
26. Lazaridis C, Yang M, DeSantis SM, Luo ST, Robertson CS. Predictors of intensive care unit length of stay and intracranial pressure in severe traumatic brain injury. J Crit Care. 2015;30(6):1258-62.
27. Kim TK, Yoon JR, Choi YN, Park UJ, Kim KR, Kim T. Risk factors of emergency reoperations. Anesth Pain Med (Seoul). 2020;15(2):233-40.