Iran University of Medical Sciences

Journal of Health Administration

https://jha.iums.ac.ir/

Original article

Predicting the length of hospital stay in patients with congestive heart failure using data mining techniques

Mozhgan Esmaeili[a], Hadi Lotfnezhad Afshar[b,c*], Bahlol Rahimi[a,c], Kamal Khademvatani[d], Shirin Samadzad Qushchi[c], Vahid Hoseinpour[e]

[a] *Department of Medical Informatics, School of Allied Medical Sciences, Urmia University of Medical Sciences, Urmia, Iran.*
[b] *Department of Health Information Technology, School of Allied Medical Sciences, Urmia University of Medical Sciences, Urmia, Iran.*
[c] *Health and Biomedical Informatics Research Center, Urmia University of Medical Sciences, Urmia, Iran.*
[d] *Department of Cardiology, School of Medicine, Urmia University of Medical Sciences, Urmia, Iran.*
[e] *Department of Emergency Medicine, School of Medicine, Urmia University of Medical Sciences, Urmia, Iran.*

ARTICLE INFO

Corresponding Author:
Hadi Lotfnezhad Afshar

e-mail addresses:
lotfnezhadafshar.h@ums u.ac.ir

Received:09/Feb/2025
Revised: 16/Aug/2025
Accepted: 03/Sep/2025
Published: 28/Sep/2025

Keywords:
Congestive heart failure
Length of stay
Data mining
Machine learning
Random forest
Association rule mining

10.61882/jha.28.2.53

ABSTRACT

Introduction: Congestive heart failure (CHF) is a significant global challenge for healthcare systems, with its prevalence increasing due to an aging population. Accurate prediction of the length of stay (LOS) for CHF patients is critical for optimizing hospital resource management, reducing treatment costs, and improving the quality of medical care. This study aimed to develop a data mining-based predictive model to estimate the LOS of CHF patients and identify the most influential factors.

Methods: This cross-sectional study was conducted using the data of 3,421 CHF patients hospitalized at Seyed Al-Shohada and Ayatollah Taleghani hospitals in Urmia, Iran, between 2018 and 2020. Data from Seyed Al-Shohada Hospital were used for model training (80%) and testing (20%). The LOS was categorized into short-term and long-term classes using K-means clustering. Random forest, Decision tree (C5.0), Artificial neural network (ANN), and adaptive neuro-fuzzy inference system (ANFIS) were applied to classify LOS. Techniques such as oversampling, undersampling, and SMOTE were applied to balance the classes, and 10-fold cross-validation was used to ensure model reliability. The Apriori algorithm was also used to discover association rules.

Results: The random forest achieved the best performance with an accuracy of 87.14%, a sensitivity of 97.56%, and an AUC of 85.40%. Key predictors of LOS included elevated creatinine levels, low hemoglobin, male gender, and underlying comorbidities. The Apriori algorithm also revealed significant clinical and meaningful associations among variables.

Conclusion: The proposed model can serve as an effective tool for predicting LOS in CHF patients and support clinical and administrative decision-making in hospital settings.

What was already known about this topic:

- Congestive heart failure (CHF) is among the costliest chronic diseases with a high rate of hospital readmission.
- Accurate prediction of the length of stay (LOS) in CHF patients is essential for bed management, discharge planning, and optimal allocation of hospital resources.
- Traditional statistical methods often lack the accuracy and interpretability required in real-world clinical settings for LOS prediction.

What this study added to our knowledge:

- The prolonged hospital stay in CHF patients is significantly associated with elevated creatinine levels, low hemoglobin, and the presence of comorbidities.

Copyright: © 2025 The Author(s); Published by Iran University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0) (https://creativecommons.org/licenses/by-nc/4.0/), which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source.

Extended Abstract

Introduction

Congestive heart failure (CHF) is one of the most common and severe chronic diseases worldwide, significantly contributing to increased mortality rates and reduced quality of life [1–3]. Owing to frequent and prolonged hospitalizations, CHF imposes substantial pressure on hospital systems, including inpatient beds, medical staff, and healthcare equipment. Global estimates indicate that CHF affects approximately 64.3 million people, with a prevalence of 1%-2% among adults in developed countries and over 25 million cases worldwide [4–6]. In the United States alone, healthcare costs associated with CHF are projected to rise from $39 billion to over $153 billion by 2030 [7, 8]. The growing prevalence of CHF is particularly evident in developing countries such as Iran, where an aging population is accelerating disease incidence [9, 10]. Accurate prediction of the length of stay (LOS) for CHF patients enables healthcare providers to better estimate bed occupancy rates and optimize hospital operations. For CHF patients, LOS prediction also facilitates better discharge planning, which is critical for improving patient outcomes and minimizing the risk of readmission [11].

Data mining techniques provide innovative methods to analyze large-scale healthcare data and can be effectively used to develop LOS prediction models [12–17]. Machine learning and data mining have been applied effectively for health-related predictions, particularly with models such as support vector machines (SVM) and random forests (RF). For instance, Hache-Sou et al. [18] applied machine learning algorithms to predict LOS in cardiac patients, achieving 96.4% accuracy. Similarly, Turgeman et al. [17] used regression trees (Cubist) and SVM for LOS prediction, achieving 84% accuracy. However, these studies often lack external validation and focus primarily on general cardiac patients rather than CHF populations. Moreover, previous research has rarely employed association rule mining techniques, such as the Apriori algorithm, to identify specific factors influencing prolonged LOS. The Apriori algorithm enables the extraction of actionable clinical insights by identifying associations between patient features and LOS patterns [19, 20].

In the context of LOS prediction for CHF patients, a significant research gap remains in applying advanced data mining techniques to improve predictive accuracy and identify key clinical predictors. Studies by Luo et al. [21] and Dagistani et al. [22] have demonstrated the potential of algorithms such as RF for LOS prediction. However, only a few studies have specifically targeted CHF patients, and fewer still have applied a combined approach of predictive modeling and association rule mining to provide comprehensive clinical insights. To address these gaps, the present study proposes a data mining framework that combines various machine learning algorithms for accurate LOS prediction, along with the Apriori algorithm to uncover hidden associations in CHF patient data.

Methods

Data collection: This study was conducted using a retrospective cross-sectional design. Data were collected from 3,421 patients diagnosed with CHF discharged between 2018 and 2020 from Seyed Al-Shohada and Ayatollah Taleghani Hospitals in Urmia, Iran. A total of 1,690 records from Seyed Al-Shohada Hospital were used as the primary dataset (Dataset 1) for model development, while 1,719 records from Ayatollah Taleghani Hospital were used as an external validation dataset (Dataset 2).

The dataset included 27 variables covering demographic information (e.g., age, gender) and clinical characteristics such as hypertension history, length of stay (LOS), family history, diabetes, dyslipidemia, history of valve replacement, coronary artery bypass grafting, angioplasty, mitral balloon valvuloplasty, chronic pulmonary disease, asthma, stroke, atrial fibrillation, myocardial infarction, pericardial effusion, comorbidities, smoking, drug addiction, alcohol use, underlying etiology, elevated creatinine, low hemoglobin, number of CHF-related hospitalizations, and number of cardiovascular-related hospitalizations. Only patients with an ICD-10 code of I50.0 (CHF) were included.

Data cleaning: Variables such as body mass index (BMI) were excluded due to more than 70% missing data (1,183 cases), which could significantly bias model performance. For variables with less than 1% missing data, such as elevated creatinine (4 missing data) and low hemoglobin (17 missing data), mode imputation was applied. These strategies were implemented to preserve model accuracy and minimize errors from missing data.

Feature selection: In consultation with expert cardiologists and by referencing medical guidelines (e.g., ESC 2021 for heart failure), 27 out of 35 available variables were selected for analysis. This selection was validated through literature review [21, 23–27], manual review of patients' electronic health records, and expert inputs. A panel of four cardiologists (average age: 52 years; average experience: 19 years; three males, one female) participated in this process.

Clustering: To transform LOS into a classification-ready format, K-means clustering was applied, guided by Silhouette coefficient (0.65) and

the Elbow method to determine the optimal number of clusters. Clustering served as an unsupervised pre-processing step to uncover natural patterns in the data. Results indicated that K = 2 was optimal, consistent with prior studies that suggested a 7-day threshold to distinguish short and long hospital stays [28]. Accordingly, LOS was categorized into short stay: ≤ 7 days and long stay: > 7 days.

K-means clustering was performed after initial pre-processing (removing invalid records and imputing missing values) but before data balancing, to preserve the natural distribution of the data.

Balancing the data: To balance the binary LOS classes, resampling techniques including over-sampling, under-sampling, and synthetic minority over-sampling technique (SMOTE) were applied. SMOTE achieved the best performance an AUC of 85% and F1-score of 78%, providing more diverse and generalizable synthetic samples. Resampling was performed only during model training for supervised classifiers.

Apriori rule mining: The Apriori algorithm was employed to identify significant associations among variables. It was applied directly to the binary-labeled dataset (short vs long LOS) derived from K-means clustering, avoiding potential biases from machine learning classification outputs.

Model training: The initial dataset consisted of 1,690 records. After preprocessing and removing incomplete cases, 1,248 records remained. The dataset was then split into 80% training (1,000 records) and 20% testing (248 records). Modeling was conducted using SPSS Clementine 12 and R. Machine learning algorithms including decision tree (DT), neural network (NN), and adaptive neuro-fuzzy inference system (ANFIS) were evaluated. Random forest (RF) outperformed other models and was fine-tuned using grid search with 10-fold cross-validation. The final optimized hyperparameters for RF included number of trees: 100, features per split: 5, maximum tree depth: 30, splitting criterion: Gini index, minimum samples per split: 2.

While all algorithms were fine-tuned, RF showed superior improvements and resistance to overfitting compared to other models.

External validation: To evaluate the model generalizability, the trained model was validated on an independent dataset of 1,719 records from Ayatollah Taleghani Hospital. Pre-processing was applied in the same way as for Dataset 1. The BMI was excluded as well. The validation set included 1,133 patients (65.9%) with LOS ≤ 7 days and 586 patients (34.1%) with LOS > 7 days.

Evaluation: Model performance on both the training and testing datasets was assessed using the standard metrics, such as accuracy, sensitivity (recall), specificity, precision, Cohen's kappa, F1-score, ROC curve and AUC [29-33].

Results

Dataset 1 (Seyed Al-Shohada Hospital): The dataset was divided into two classes: short and long stay. The Table 1 outlines the characteristics of each class.

Table 1. Comparison of clinical and demographic features (Dataset 1)

Feature	Short-term (n=1171)	Long-term (n=519)	p-value
Mean age (years)	68.2 ± 12.2	75.3 ± 10.8	<0.001
Male gender (%)	52.1	58.7	0.013
Hypertension (%)	65.4	78.2	<0.001
Diabetes (%)	32.1	45.6	<0.001
Atrial fibrillation (%)	15.3	28.9	<0.001
Elevated creatinine (%)	48.2	72.4	<0.001
Low hemoglobin (%)	53.1	68.9	<0.001
History of angioplasty (%)	8.7	12.5	0.021

Of the 27 variables examined, only those showing statistically significant differences between the short and long stay ($p < 0.05$) are reported in Table 1.

Baseline feature analysis revealed that patients with long hospital stays were significantly older and had a higher prevalence of comorbidities (diabetes, hypertension) and laboratory abnormalities (elevated creatinine, low hemoglobin).

The algorithm implementation on dataset 1 (Seyed Al-Shohada Hospital) showed that RF outperformed decision trees, ANN, and ANFIS. As shown in Table 2, RF achieved accuracy of 87.14%, sensitivity of 97.56%, specificity of 23.24%, AUC of 55.40%, and F1-score of 71.13%.

Table 2. Algorithm performance on dataset 1 (Seyed Al-Shohada Hospital)

Algorithm	Accuracy (%)	Sensitivity(%)	Specificity(%)	AUC (%)	Kappa (%)	F1-score
Decision tree (C5.0)	67.56	94.18	07.08	51.03	01.55	65.61
Artificial neural network (ANN)	64.24	92.24	26.88	59.56	20.75	69.36
ANFIS	67.15	84.44	27.86	56.15	13.73	65.82
Random forest (RF)	87.14	97.56	23.24	55.40	22.95	71.13

After obtaining the above metrics, three balancing techniques (SMOTE, over-sampling, and under-sampling) were applied. SMOTE was applied only to the training data to balance class distribution. Evaluation was performed on the original, imbalanced test data. As a result, no significant

improvement in accuracy was observed, but sensitivity and F1-score improved compared to the case without SMOTE. The slight decrease in specificity reflects the model's shifted focus toward the long-stay class. As summarized in Table 3, Figure 1, and Figure 2, in the balanced dataset, RF demonstrated the best separation as well (AUC = 0.854). Table 4 shows the important features for LOS prediction based on the final RF model.

Table 3. Comparison of C5.0, ANFIS, and Random Forest after SMOTE balancing (Test Set - Seyed Al-Shohada Hospital)

Metric	Accuracy%	Sensitivity%	Specificity%	AUC%	Kappa%	F1-score%
C5.0	85.13	91.09	70.00	80.55	64.63	79.16
ANFIS	73.17	85.28	57.04	71.16	43.59	68.36
Random Forest	81.45	98.39	64.52	85.40	62.90	84.14

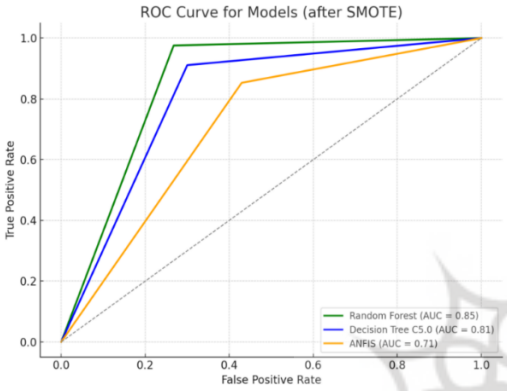


Figure 1. ROC curves of the algorithms (Dataset 1)

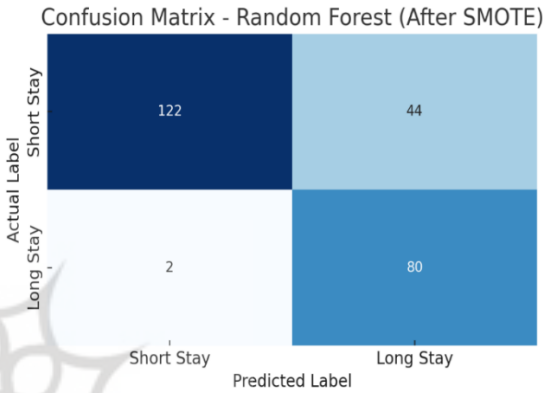


Figure 2. Confusion matrix for dataset 1

Table 4. Important predictive features

Rank	Variable	Predictive Role	Explanation
1	History of CABG	Positive	Patients with a history of coronary artery bypass grafting (CABG) were more likely to have a hospital stay longer than 7 days.
2	Diabetes	Positive	Diabetic patients were more frequently found in the long-term stay group.
3	Dyslipidemia	Positive	Dyslipidemia was associated with longer hospital stays
4	Male Gender	Positive	The proportion of male patients was higher in the >7-day stay group.
5	Hypertension	Negative	Hypertension was more common among patients with shorter stays.
6	History of PCI	Positive	Patients with a history of percutaneous coronary intervention (PCI) had longer lengths of stay.
7	Elevated Creatinine	Negative	High creatinine levels, particularly when combined with hypertension, were associated with shorter hospital stays.

The Apriori algorithm was used to extract rules identifying key factors affecting length of stay. Support and confidence thresholds were selected empirically, based on literature and expert validation. The following two rules were considered most clinically meaningful:

- Rule 1: Male patients with hypertension, no valve replacement history, and elevated creatinine are more likely to have shorter hospital stays (Support: 0.107; Confidence: 0.923)
- Rule 2: Patients with atrial fibrillation and elevated creatinine, but no angioplasty, no stroke, and no addiction history, are more likely to have longer stays (Support: 0.104; Confidence: 0.864)

Dataset 2 (Taleghani Hospital): This dataset included 1719 patients, divided into two groups: short stay (n=1133 patients/ 65.9%, long stay (n= 586 patients /34.1%). This dataset was used only for evaluation of previously developed models. Table 5, Figure 3, and Figure 4 confirm the generalizability and real-world applicability of RF model. The consistency in accuracy (77.40%) and AUC (84.82%) supports its use in clinical environments.

Table 5. Algorithm performance for dataset 2 (Taleghani Hospital)

Algorithm	Accuracy%	Sensitivity%	Specificity%	AUC%	Kappa%	F1-score%
Random Forest (RF)	77.40	94.32	68.61	84.82	65.40	73.90
Decision Tree (C5.0)	74.05	93.24	70.29	81.47	65.40	72.91
ANFIS	76.03	83.68	70.30	76.99	52.50	71.90

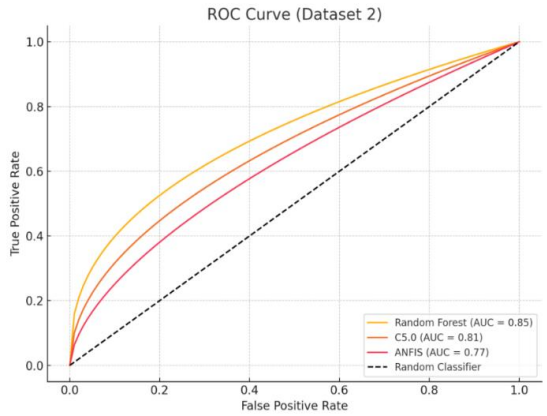


Figure 3. ROC curve for dataset 2(Taleghani Hospital)

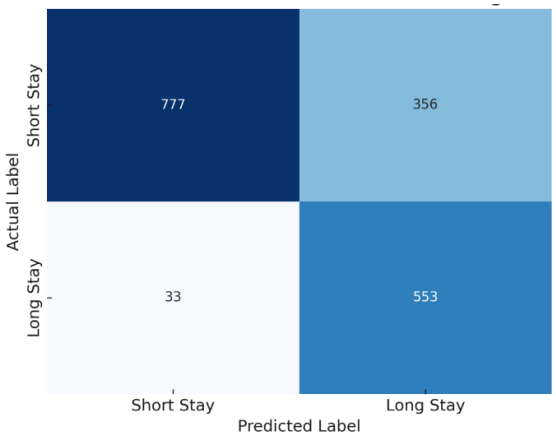


Figure 4. Confusion matrix for dataset 2(Taleghani Hospital)

Discussion

The proposed model was successfully classified the length of stay (LOS) for patients with congestive heart failure (CHF) with high accuracy by employing the random forest algorithm. The model performance was acceptable on both the internal and external validation datasets, and results indicated that machine learning-based approaches are effective tools for predicting LOS in CHF patients.

These findings are consistent with the the study by Dagistani et al. [22], which employed data-driven algorithms to analyze the medical records of cardiac patients. However, that study did not report how missing data were handled, whereas in the present study, careful preprocessing and systematic handling of incomplete data were crucial for enhancing the model performance. The use of the K-means algorithm for clustering LOS facilitated a more precise separation grouping of patients and contributed to an improved classification. This method, combined with advanced algorithms, such as random forest, outperformed simpler models like decision trees or artificial neural networks (ANN), a point that should be further explored in comparison with similar studies.

In comparison with the study by Aghajani et al. [34], which focused on the factors affecting LOS in the general surgery ward in Tehran and reported a decision tree accuracy of 84.69%, the random forest model in the current study showed superior performance. Moreover, although Maharloo et al. [35] reported high performance of ANFIS for predicting LOS in ICU patients after cardiac surgery, in our study, this algorithm underperformed compared with RF and C5.0. These discrepancies may result from differences in patient populations, data characteristics, or preprocessing stages.

In a similar study, Gholipour et al. [36] employed an artificial neural network algorithm to predict trauma patients' survival and LOS in the ward and ICU. Although their model predicted patients' clinical outcomes with good accuracy (93.33%), LOS prediction was relatively error-prone. In contrast, in the present study, the RF model accurately classified patients into short- and long-term stay groups with high accuracy and acceptable AUC. Another notable aspect was the use of the SMOTE technique for data balancing in the present study. Unlike studies that used simpler methods such as undersampling, this approach improved model accuracy. Overall, employing advanced machine learning algorithms, especially random forest, combined with proper data preprocessing and class balancing improves LOS prediction in patients.

Ultimately, proposed model identified variables such as gender, hypertension, comorbidities, and creatinine level as key predictors of LOS in CHF patients. Specifically, higher creatinine levels and the presence of comorbidities were associated with longer hospital stays, whereas male patients with hypertension but without a history of heart valve replacement were more likely fall into the short-stay group. These findings are consistent with confirming the role of comorbidities and impaired kidney function in increasing hospitalization duration. For example, Dagistani et al. [22], identified chronic diseases such as diabetes and hypertension as factors contributing increased LOS. Moreover, previous studies have shown that impaired kidney function, through its effect on fluid and electrolyte balance, may delay the recovery process in CHF patients and increase LOS [37]. Therefore, considering these variables at the time of admission can play a key role in predicting LOS and optimizing hospital resource management.

Hypertension also emerged as a significant predictor, consistent with studies suggesting that it exacerbates CHF and, due to its association with comorbid conditions, results in longer hospital stays [26, 38]. Particularly when combined with other chronic diseases, hypertension can complicate patient condition management and delayed

discharge. This finding aligns with Gottlieb et al. [39], who showed that CHF patients with hypertension often remain hospitalized longer due to the need for more intensive management and the higher risk of complications.

Creatinine level was another strong predictor of LOS. Elevated creatinine level indicates impaired kidney function, which can complicate CHF treatment. Poor renal function leads to longer hospital stays because these patients require closer monitoring, more precise drug therapy, and more complex management [39].

Atrial fibrillation was also found to be associated with longer LOS. This cardiac rhythm disorder usually co-occurs with heart failure and, due to the need for monitoring, multi-drug therapy, and higher risk of complications, results in greater resource use and delayed discharge [40]. Overall, these results highlight the importance of identifying high-risk patients at admission so that accurate LOS prediction can enable more efficient hospital resource allocation.

Through two complementary approaches (random forest algorithm for LOS prediction and Apriori for association rules extraction), this study proposed a comprehensive model for analyzing LOS in CHF patients. While previous studies such as Hachesu et al. [18] and Torgeman et al. [17] focused mainly on precise LOS prediction, the present study enhanced interpretability by association rule analysis. The Apriori algorithm was applied to binary-classified data (LOS ≤7 and >7 days), identifying specific combinations of patient features. For example, male patients with hypertension, and no valve replacement" are more associated with short stays. These rules complement the random forest model and can help interpret results and design targeted intervention programs.

The practical implications of these findings are significant for care planning and resource allocation in CHF management. Physicians can use these two approaches to identify patients at risk of longer stays early and plan targeted care accordingly. For instance, feature combinations such as atrial fibrillation and high creatinine levels, which are associated with longer LOS, can be applied to design personalized treatment pathways. Moreover, external validation using an independent dataset enhanced robustness and generalizability, indicating that applicability of this model beyond the study site.

Limitations

Despite its strengths, this study has several limitations. The data were collected from only two hospitals in Iran, which may limit the generalizability of the findings to other healthcare systems or populations. Furthermore, the dataset lacked variables such as detailed echocardiographic data or medication history, which could have provided a more comprehensive picture of patients' condition and potentially improved prediction accuracy.

Future research should include additional clinical data, particularly imaging and medication-related variables, to enhance model performance and clinical relevance. Finally, exploring advanced machine learning techniques such as ensemble learning or deep learning may provide deeper insights into complex interactions within patient data and further improve LOS prediction in CHF and related conditions.

Conclusion

This study demonstrated the effectiveness of data mining techniques for predicting the length of stay (LOS) for patients with congestive heart failure (CHF) and highlighted its practical implications for resource management and patient care. By integrating predictive modeling with association rule mining, we proposed a comprehensive approach that can be adapted to other chronic diseases as well. Accurate LOS prediction facilitates improved planning and resource allocation, thereby enhancing the efficiency of healthcare delivery for CHF patients. The findings from this model can assist clinicians in identifying high-risk patients who may require prolonged care and facilitate timely interventions.

Declarations

Ethical considerations: This study was conducted under the ethical approval code IR.UMSU.REC.1398.012 issued by the Ethics Committee in Biomedical Research at Urmia University of Medical Sciences.

Funding: This study was part of a Master's thesis supported by the Vice Chancellor for Research and Technology at Urmia University of Medical Sciences. The funding body had no role in data collection, analysis, or manuscript preparation.

Conflict of interest: The authors declare no conflicts of interest related to this manuscript.

Authors' contributions: ME: Conceptualization, study design, data collection, methodology, software, validation, data analysis, data curation, writing – original draft, writing – review & editing. **HLA:** Conceptualization, study design, data collection, methodology, software, validation, data analysis, data curation, writing – original draft, writing – review & editing, Project administration, Fund acquisition. **BR:** Methodology, software, validation, data analysis. **KKH:** data collection**,** Methodology, validation, data analysis. **ShSG:** software, writing – review & editing, writing – review & editing, Visualization. **VH:** Methodology, validation, Financing.

Consent for publication: Not applicable**.**

Data availability: The datasets and codes used in this study are available from the corresponding author upon reasonable request via email
AI declaration: The English text of this article was edited using InstaText software. All content revised with the software was reviewed and approved by the authors.
Acknowledgments: The authors wish to thank all healthcare providers who supported this study at Seyed Al-Shohada and Ayatollah Taleghani Hospitals in Urmia. This article is based on a Master's thesis titled *"Predicting Length of Stay in Congestive Heart Failure Patients Using Data Mining Techniques at Seyed Al-Shohada and Ayatollah Taleghani Teaching Hospitals in Urmia,"* approved by Urmia University of Medical Sciences in 2020 (Project Code: 2509 ,Tracking code:3144).

References

1. Alemzadeh-Ansari MJ, Ansari-Ramandi MM, Naderi N. Chronic pain in chronic heart failure: a review article. The Journal of Tehran University Heart Center. 2017;12(2):49-56. Available from: /https://pmc.ncbi.nlm.nih.gov/articles/PMC5558055
2. Keyhani D, Razavi Z, Shafiee A, Bahadoram S. Autonomic function change following a supervised exercise program in patients with congestive heart failure. ARYA Atherosclerosis. 2013;9(2):150-156. PMCID: PMC3653242. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3653242
3. Writing Group Members, Rosamond W, Flegal K, et al. Heart disease and stroke statistics—2009 update: a report from the American heart association statistics committee and stroke statistics subcommittee. Circulation. 2009;119(3):e21–e181 https://doi.org/10.1161/CIRCULATIONAHA.108.191261
4. Ahmadi A, Soori H, Mobasheri M, Etemad K, Khaledifar A. Heart failure: the outcomes, predictive and related factors in Iran. Journal of Mazandaran University of Medical Sciences. 2014;24(118):180–188. [In Persian]. Available from: http://jmums.mazums.ac.ir/article-1-4636-en.html
5. Liu LC, Voors AA, van Veldhuisen DJ, van der Meer P. Heart failure highlights in 2012–2013. European Journal of Heart Failure. 2014;16(2):122–32. https://doi.org/10.1002/ejhf.43
6. Bowen RES, Graetz TJ, Emmert DA, Avidan MS. Statistics of heart failure and mechanical circulatory support in 2020. Annals of Translational Medicine. 2020;8(13):827. https://doi.org/10.21037/atm-20-1127
7. Nomali M, Mohammadrezaei R, Keshtkar AA, Roshandel G, Ghiyasvandian S, Alipasandi K, et al. Self-monitoring by traffic light color coding versus usual care on outcomes of patients with heart failure reduced ejection fraction: protocol for a randomized controlled trial. JMIR Research Protocols. 2018;7(11):e9209. https://doi.org/10.2196/resprot.9209
8. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. Nature Reviews Cardiology. 2016;13(6):368–78. https://doi.org/10.1038/nrcardio.2016.25
9. Mirdamadi A, Shafiee A, Ansari-Ramandi M, Garakyaraghi M, Pourmoghaddas A, Bahmani A, Mahmoudi H, Gharipour M. Beneficial effects of testosterone therapy on functional capacity, cardiovascular parameters, and quality of life in patients with congestive heart failure. BioMed Research International. 2014;2014:392432. https://doi.org/10.1155/2014/392432
10. Mori J, Krantz MJ, Tanner J, Horwich TB, Yancy C, Albert NM, Hernandez AF, Dai D, Fonarow GC. Influence of hospital length of stay for heart failure on quality of care. The American Journal of Cardiology. 2008;102(12):1693–1697. https://doi.org/10.1016/j.amjcard.2008.08.015
11. Azari A, Janeja VP, Mohseni A. Predicting hospital length of stay (PHLOS): a multi-tiered data mining approach. In: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW). 2012. p. 17–24. https://doi.org/10.1109/ICDMW.2012.69
12. Mehdipour Y, Ebrahimi S, Karimi A, Alipour J, Khammarnia M, Siasar F. Presentation a model for prediction of cerebrovascular accident using data mining algorithm. Sadra Medical Journal. 2016;4(4):255–266. Available from: https://smsj.sums.ac.ir/article_43946_en.html
13. Ristevski B, Chen M. Big data analytics in medicine and healthcare. Journal of Integrative Bioinformatics. 2018;15(3):20170030. https://doi.org/10.1515/jib-2017-0030
14. Pasupathi C, Kalavakonda V. Evidence based healthcare system using big data for disease diagnosis. In: 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and BioInformatics (AEEICB). 2016. p. 370–4. https://doi.org/10.1109/AEEICB.2016.7538393
15. Sarafi Nejad A, Saeid A, Mohammed Rose I, Rowhanimanesh A. Modeling a data mining decision tree and propose a new model for the diagnosis of skin cancer by immunohistochemical staining methods. Journal of Health and Biomedical Informatics. 2014;1(1):54–62. Available from: http://jhbmi.ir/article-1-62-en.html
16. Tekieh MH, Raahemi B. Importance of data mining in healthcare: a survey. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2015. p. 1057–62. https://doi.org/10.1145/2808797.2809367
17. Turgeman L, May JH, Sciulli R. Insights from a machine learning model for predicting the hospital length of stay at the time of admission. Expert Systems with Applications. 2017;78:376–85. https://doi.org/10.1016/j.eswa.2017.02.023
18. Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. Healthcare Informatics Research. 2013;19(2):121–9. https://doi.org/10.4258/hir.2013.19.2.121

19. Thuraisingham B. A primer for understanding and applying data mining. IT Professional. 2002;2(1):28–31. https://doi.org/10.1109/6294.819936
20. Zhao J, Feng X, Pang Q, Fowler M, Lian Y, Ouyang M, et al. Battery safety: machine learning-based prognostics. Progress in Energy and Combustion Science. 2024;102:101142. https://doi.org/10.1016/j.pecs.2023.101142
21. Luo L, Lain S, Feng C, Huang D, Zhang W. Data mining-based detection of rapid growth in length of stay on COPD patients. In: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). 2017. p. 319–23. https://doi.org/10.1109/ICBDA.2017.8078819
22. Daghistani TA, Elshawi R, Sakr S, Ahmad A, Al-Thwayee A, Al-Mallah. Predictors of in hospital length of stay among cardiac patients: a machine learning approach. International Journal of Cardiology. 2019; 288:140–7. https://doi.org/10.1016/j.ijcard.2019.01.046
23. Neri L, Oberdier MT, van Abeelen KCJ, Menghini L, Tumarkin E, Tripathi H, et al. Electrocardiogram monitoring wearable devices and artificial-intelligence-enabled diagnostic capabilities: a review. Sensors. 2023;23(10):4805. https://doi.org/10.3390/s23104805
24. Dai W, Brisimi TS, Adams WG, Mela T, Saligrama V, Paschalidis IC. Prediction of hospitalization due to heart diseases by supervised learning methods. International Journal of Medical Informatics. 2015;84(3):189-197. https://doi.org/10.1016/j.ijmedinf.2014.10.002
25. Natale J. A strategy for reducing congestive heart failure readmissions through the use of interventions targeted by machine learning [Doctoral dissertation]. University of Akron; 2015. OhioLINK Electronic Theses and Dissertations Center. Available from: http://rave.ohiolink.edu/etdc/view?acc_num=akron1428233380
26. Messerli FH, Rimoldi SF, Bangalore S. The transition from hypertension to heart failure: contemporary update. JACC: Heart Failure. 2017;5(8):543–51. https://doi.org/10.1016/j.jchf.2017.04.012
27. Berkhin P, Becher JD. Learning simple relations: theory and applications. In: Proceedings of the 2002 SIAM International Conference on Data Mining. 2002. p. 420–36. https://doi.org/10.1137/1.9781611972726.25
28. Zebin T, Rezvy S, Chaussalet TJ. A deep learning approach for length of stay prediction in clinical settings from medical records. In: 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). 2019. p. 1–6. https://doi.org/10.1109/CIBCB.2019.8791477
29. Flach P, Blockeel H, Ferri C, Orallo JH, Struyf J. Decision support for data mining: an introduction to ROC analysis and its applications. In: Data Mining and Decision Support: Integration and Collaboration. Springer; 2003. p. 81–90. https://doi.org/10.1007/978-1-4615-0286-9_7
30. Galdi P, Tagliaferri R. Data mining: accuracy and error measures for classification and prediction. Encyclopedia of Bioinformatics and Computational Biology. 2018;1:431–6. https://doi.org/10.1016/B978-0-12-809633-8.20474-3
31. Ben-David A. About the relationship between ROC curves and Cohen's kappa. Engineering Applications of Artificial Intelligence. 2008;21(6):874–81. https://doi.org/10.1016/j.engappai.2007.09.009
32. Preda S, Oprea SV, Bâra A, Belciu (Velicanu) A. PV forecasting using support vector machine learning in a big data analytics context. Symmetry. 2018;10(12):748. https://doi.org/10.3390/sym10120748
33. Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering. 2005;17(3):299–310. https://doi.org/10.1109/TKDE.2005.50
34. Levy D, Larson MG, Vasan RS, Kannel WB, Ho KK. The progression from hypertension to congestive heart failure. JAMA. 1996;275(20):1557–62. https://doi.org/10.1001/jama.1996.03530440037034
35. Maharlou H, Niakan Kalhori S.R, 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. Healthcare Informatics Research. 2018;24(2):109–17. https://doi.org/10.4258/hir.2018.24.2.109
36. Gholipour C, Rahim F, Fakhree A, Ziapour B. Using an artificial neural networks (ANNs) model for prediction of intensive care unit (ICU) outcome and length of stay at hospital in traumatic patients. Journal of Clinical and Diagnostic Research. 2015;9(4):OC19–23. https://doi.org/10.7860/JCDR/2015/9467.5828
37. Bleumink GS, Knetsch AM, Sturkenboom MC, Straus SM, Hofman A, Deckers JW, et al. Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure: the Rotterdam Study. European Heart Journal. 2004;25(18):1614–9. https://doi.org/10.1016/j.ehj.2004.06.038
38. Sud M, Yu B, Wijeysundera HC, Austin PC, Ko DT, Braga J, et al. Associations between short or long length of stay and 30 day readmission and mortality in hospitalized patients with heart failure. JACC: Heart Failure. 2017;5(8):578–88. https://doi.org/10.1016/j.jchf.2017.03.012
39. Gottlieb SS, Abraham W, Butler J, Forman DE, Loh E, Massie BM, et al. The prognostic importance of different definitions of worsening renal function in congestive heart failure. Journal of Cardiac Failure. 2002;8(3):136–41. https://doi.org/10.1054/jcaf.2002.125289
40. Heist EK, Ruskin JN. Atrial fibrillation and congestive heart failure: risk factors, mechanisms, and treatment. Progress in Cardiovascular Diseases. 2006;48(4):256–69. https://doi.org/10.1016/j.pcad.2005.09.001

نشریه مدیریت سلامت

https://jha.iums.ac.ir/

دانشگاه علوم پزشکی ایران

مقاله اصیل

پیش‌بینی طول مدت بستری بیماران مبتلا به نارسایی قلبی با داده‌کاوی

مژگان اسماعیلی[1]، هادی لطف نژاد افشار[2,3*]، بهلول رحیمی[1,3]، کمال خادم وطنی[4]، شیرین صمد زاد قوشچی[3]، وحید حسین پور[5]

[1] گروه انفورماتیک پزشکی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی ارومیه، ارومیه، ایران.

[2] گروه فناوری اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی ارومیه، ارومیه، ایران.

[3] مرکز تحقیقات انفورماتیک سلامت و زیست پزشکی، دانشگاه علوم پزشکی ارومیه، ارومیه، ایران.

[4] گروه قلب و عروق، دانشکده پزشکی، دانشگاه علوم پزشکی ارومیه، ارومیه، ایران.

[5] گروه طب اورژانس، دانشکده پزشکی، دانشگاه علوم پزشکی ارومیه، ارومیه، ایران.

اطلاعات مقاله

نویسنده مسئول:
هادی لطف نژاد افشار

رایانامه:
lotfnezhadafshar.h@umsu.ac.ir

وصول مقاله: ۱۴۰۳/۱۱/۲۱
اصلاح نهایی: ۱۴۰۴/۰۵/۲۵
پذیرش نهایی: ۱۴۰۴/۰۶/۱۲
انتشارآنلاین: ۱۴۰۴/۰۷/۰۶

واژه‌های کلیدی:
نارسایی احتقانی قلب
طول مدت بستری
داده‌کاوی
یادگیری ماشین
الگوریتم جنگل تصادفی

چکیده

مقدمه: نارسایی احتقانی قلب یکی از چالش‌های اصلی نظام‌های سلامت در جهان است که با افزایش جمعیت سالمند، شیوع آن رو به افزایش است. پیش‌بینی دقیق طول مدت بستری بیماران مبتلا می‌تواند نقش مهمی در مدیریت بهینه منابع بیمارستانی، کاهش هزینه‌های درمان و ارتقای کیفیت مراقبت‌های پزشکی ایفا کند. این مطالعه با هدف ارائه مدل پیش‌بینی مبتنی بر داده‌کاوی جهت تخمین مدت بستری بیماران مبتلا به نارسایی احتقانی قلب و شناسایی عوامل کلیدی مؤثر بر آن انجام شد.

روش‌ها: این پژوهش تحلیلی-کاربردی به‌صورت مقطعی بر روی داده‌های ۳۴۲۱ بیمار بستری در دو بیمارستان سیدالشهدا و آیت‌الله طالقانی ارومیه طی سال‌های ۱۳۹۷ تا ۱۳۹۹ انجام شد. داده‌های بیمارستان سیدالشهدا برای آموزش (۸۰٪) و آزمون مدل (۲۰٪)استفاده شد. طول مدت بستری با الگوریتم K-means خوشه‌بندی و به دو خوشه کوتاه‌مدت و بلندمدت تفکیک شد. الگوریتم‌های یادگیری ماشین شامل جنگل تصادفی، درخت تصمیم(C5.0)، شبکه عصبی مصنوعی و سیستم استنتاج عصبی-فازی تطبیقی (ANFIS) برای طبقه‌بندی مدت بستری استفاده شد. همچنین از روش‌های بیش‌نمونه‌گیری، کم نمونه‌گیری و SMOTE برای متعادل‌سازی کلاس‌ها، اعتبارسنجی متقاطع ۱۰ تایی برای پایایی مدل‌ها و الگوریتم اپریوری برای کشف قوانین همبستگی بین متغیرها استفاده شد.

یافته‌ها: جنگل تصادفی با صحت ۸۷/۱۴٪، حساسیت ۹۷/۵۶٪ و AUC معادل ۸۵/۴۰٪ بهترین عملکرد را داشت. متغیرهایی مانند کراتینین بالا، هموگلوبین پایین، جنسیت مرد و بیماری‌های زمینه‌ای عوامل مؤثر بر مدت اقامت بودند.

نتیجه‌گیری: مدل پیشنهادی می‌تواند ابزاری مؤثر برای طبقه‌بندی مدت بستری بیماران نارسایی احتقانی قلب و پشتیبان تصمیم‌گیری‌های مدیریتی و بالینی در بیمارستان‌ها باشد.

آنچه می‌دانیم:

- نارسایی احتقانی قلب (CHF) یکی از بیماری‌های پرهزینه با نرخ بالای بستری مجدد است.
- پیش‌بینی دقیق طول مدت بستری بیماران برای مدیریت تخت، برنامه‌ریزی ترخیص و تخصیص منابع بیمارستانی ضروری است.
- روش‌های سنتی آماری دقت و تفسیرپذیری کافی برای محیط واقعی بالینی ندارند.

آنچه این مطالعه اضافه کرده است:

- افزایش طول مدت بستری در بیماران CHF با سطح کراتینین بالا، هموگلوبین پایین و وجود بیماری‌های زمینه‌ای ارتباط معنی‌دار دارد.

مقدمه

نارسایی احتقانی قلب (Congestive Heart Failure/CHF) یکی از شایع‌ترین و شدیدترین بیماری‌های مزمن در سراسر جهان است که نقش عمده‌ای در افزایش نرخ مرگ و کاهش کیفیت زندگی ایفا می‌کند [۱-۳]. این بیماری به دلیل نیاز به بستری‌های مکرر و طولانی‌مدت، فشار قابل‌توجهی بر سیستم‌های بیمارستانی (شامل تخت‌های بستری، نیروی انسانی و تجهیزات پزشکی) وارد می‌کند و چالش‌های زیادی را به همراه دارد. برآوردهای جهانی نشان می‌دهد که CHF حدود ۶۴/۳ میلیون نفر را تحت تأثیر قرار داده و شیوع آن در کشورهای توسعه‌یافته بین ۱ تا ۲ درصد در میان بزرگسالان و بیش از ۲۵ میلیون نفر در سطح جهانی است [۴-۶]. در ایالات‌متحده، پیش‌بینی می‌شود که هزینه‌های بهداشتی مرتبط با CHF از ۳۹ میلیارد دلار به بیش از ۵۳/۱ میلیارد دلار تا سال ۲۰۳۰ افزایش یابد [۷،۸]. شیوع این بیماری به‌ویژه در کشورهای درحال‌توسعه مانند ایران با افزایش جمعیت سالمند در حال افزایش است[۹،۱۰].

پیش‌بینی دقیق طول مدت بستری (Length of Stay/LOS) در بیمارستان به ارائه‌دهندگان خدمات بهداشتی این امکان را می‌دهد که نرخ اشغال تخت را بهتر تخمین بزنند و عملیات بیمارستانی را بهینه کنند. برای بیماران CHF، پیش‌بینی LOS به برنامه‌ریزی مؤثرتر ترخیص کمک می‌کند که برای مدیریت نتایج بیمار و کاهش خطر بستری مجدد حیاتی است [۱۱].

داده‌کاویروش‌های جدیدی برای تحلیل داده‌های بهداشتی در مقیاس بزرگ ارائه می‌دهد و می‌تواند برای توسعه مدل‌های پیش‌بینی LOS استفاده شود [۱۲-۱۷]. یادگیری ماشین و داده‌کاوی در پیش‌بینی‌های بهداشتی مؤثر بوده‌اند، به‌ویژه از طریق مدل‌هایی مانند ماشین‌های بردار پشتیبان و جنگل تصادفی. به‌عنوان نمونه، هاچه سو و همکاران [۱۸] در مطالعه‌ای بر روی بیماران قلبی از الگوریتم‌های یادگیری ماشین مانند ماشین بردار پشتیبان، درخت تصمیم و شبکه عصبی با هدف پیش‌بینی طول مدت اقامت بیماران در بیمارستان استفاده کردند و دقت ۹۶/۴ درصدی را به دست آوردند. به همین ترتیب، تورگمان و همکاران [۱۷] از مدل‌های یادگیری ماشین برای پیش‌بینی LOS استفاده کردند و با روش‌هایی مانند درخت‌های رگرسیون Cubist و ماشین‌های بردار پشتیبان دقت ۸۴ درصدی به دست آوردند. بااین‌حال، این مطالعات اغلب فاقد اعتبارسنجی خارجی هستند و تمرکز آنها عمدتاً به‌جای بیماران خاص CHF بر بیماران قلبی است. علاوه‌براین، پژوهش‌های گذشته به‌ندرت از روش‌های استخراج قوانین همبستگی مانند اپریوری (Apriori) برای شناسایی عوامل خاص مؤثر بر طولانی‌تر شدن LOS استفاده کرده‌اند. اپریوری امکان استخراج بینش‌های بالینی قابل‌اجرا را با نشان دادن ارتباطات بین ویژگی‌های بیمار و الگوهای LOS فراهم می‌کند [۱۹،۲۰].

درزمینه پیش‌بینی LOS در مورد CHF، هنوز فاصله تحقیقاتی در استفاده از داده‌کاوی ب وجود دارد. مطالعات لو و همکاران [۲۱] و داغیستانی و همکاران [۲۲] توان الگوریتم‌هایی مانند جنگل تصادفی را برای پیش‌بینی LOS نشان داده‌اند. بااین‌حال، مطالعات کمی به‌طور خاص بر بیماران CHF تمرکز داشته‌اند و از روش‌های پیش‌بینی و استخراج قوانین همبستگی به‌صورت ترکیبی کمتر استفاده شده است.

در این مطالعه با هدف پر کردن شکاف‌های موجود در پیش‌بینی LOS، از الگوریتم‌های مختلف یادگیری ماشین (مانند جنگل تصادفی، درخت تصمیم، شبکه عصبی و سیستم استنتاج عصبی-فازی تطبیقی) برای پیش‌بینی مدت بستری و الگوریتم اپریوری برای شناسایی همبستگی‌های پنهان در داده‌های بیماران CHF استفاده شده است.

روش ها

جمع‌آوری داده‌ها: مطالعه حاضر از نوع مقطعی، تحلیلی و کاربردی بود. داده‌های ۳۴۲۱ بیمار مبتلا CHF بستری شده بین سال‌های ۱۳۹۷ تا ۱۳۹۹ در بیمارستان‌های سیدالشهدا و آیت‌الله طالقانی در ارومیه، ایران جمع‌آوری شد. ۱۶۹۰ مورد از این داده‌ها مربوط به بیمارستان سیدالشهدا (مجموعه داده ۱) و داده اصلی مطالعه بود. ۱۷۱۹ مورد مربوط به بیمارستان طالقانی (مجموعه داده ۲) برای اعتبارسنجی خارجی استفاده شد. این مجموعه داده شامل ۲۷ متغیر است که داده‌های دموگرافیک (مانند سن، جنسیت) و ویژگی‌های بالینی (مانند سابقه فشارخون، طول مدت اقامت، سابقه فامیلی، سابقه بیماری دیابت، دیس لیپیدمی، سابقه تعویض دریچه قلب، سابقه پیوند عروق کرونر، سابقه آنژیوپلاستی، باز کردن دریچه میترال با بالون، سابقه بیماری مزمن ریوی، سابقه آسم، سابقه سکته مغزی، فیبریلاسیون دهلیزی، سابقه سکته قلبی، ازدحام مایع پریکارد، بیماری همراه، مصرف سیگار، اعتیاد به مواد مخدر، اعتیاد به الکل، علت زمینه‌ای، کراتنین بالا، هموگلوبین پایین، تعداد دفعات بستری بیمار به علت نارسایی احتقانی قلب، تعداد دفعات بستری به علت بیماری قلبی) را در برمی‌گیرد. تنها موارد دارای کد ICD-10 برای(I50.0) CHF در این مطالعه لحاظ شدند.

پاک‌سازی داده‌ها: مجموعه داده‌ها تحت یک فرآیند پاک‌سازی دقیق قرار گرفتند. متغیرهایی مانند BMI (Body Mass Index) به دلیل داشتن بیش از ۷۰٪ مقدار مفقوده (۱۱۸۳ مورد) به‌طور کامل از تحلیل حذف شد. در مقابل، برای متغیرهایی با مقدار مفقوده کمتر از ۱٪ مانند «کراتینین بالا» (۴ مورد مفقوده) و «هموگلوبین پایین» (۱۷ مورد مفقوده)، از روش جایگذاری با مقدار مُد (mode) استفاده شد. این راه‌کارها برای حفظ دقت مدل و خطا ناشی از داده‌های ناقص به‌کار گرفته شد.

انتخاب ویژگی‌ها: با مشاوره پزشکان متخصص و بررسی منابع علمی، ویژگی‌های بالینی مرتبط با طول مدت بستری در بیماران CHF انتخاب شدند. از ۳۵ متغیر موجود، ۲۷ متغیر انتخاب شد که با استانداردهای پزشکی کنونی (استفاده از راهنماهای پزشکی معتبر مانند ESC 2021 برای نارسایی قلبی و تائید تیم متخصص) اعتبارسنجی شدند [۲۳-۲۵]. این ویژگی‌ها با ارجاع به مطالعات پیشین مرتبط [۲۱، ۲۶]، حذف داده‌های مفقوده با بررسی پرونده‌های الکترونیک بیماران در بیمارستان و همچنین مشاوره با متخصصان انتخاب شدند. تیم متخصص شامل چهار نفر از متخصصان قلب و عروق با میانگین سنی ۵۲ سال و میانگین سابقه خدمت ۱۹ سال (سه مرد و یک زن) بود.

خوشه‌بندی: برای تعیین ساختار بهینه طول مدت بستری، از الگوریتم خوشه‌بندی K-means همراه با معیارهای ارزیابی سیلهوئت (۰/۶۵) و روش البو استفاده شد [۲۷]. این خوشه‌بندی طبقه‌بندی

باینری طول مدت بستری را تسهیل کرد و تمایز بین اقامت‌های کوتاه‌مدت و بلندمدت را آسان‌تر کرد. تحلیل خوشه‌بندی (به‌عنوان یک روش بدون نظارت) صرفاً برای شناسایی ساختار طبیعی داده‌ها انجام شد و تمام مراحل خوشه‌بندی بر روی داده‌های با توزیع واقعی اجرا شد. نتایج هر دو روش سیلهوئت و البو نشان داد که بهینه‌ترین تعداد خوشه برابر ۲ است. این یافته با مطالعات پیشین در خصوص حد آستانه هفت روز برای تفکیک دوره‌های بستری همسو بود [۲۸]. بر این اساس، LOS به دو گروه کوتاه‌مدت (۰-۷ روز) و بلندمدت (بیش از ۷ روز) تقسیم شد. خوشه‌بندی K-means در این مطالعه پس از انجام پیش‌پردازش اولیه (شامل حذف موارد نادرست و جایگزینی مقادیر مفقوده) ولی قبل از متعادل‌سازی داده‌ها به دلیل حفظ ساختار طبیعی داده‌ها و شناسایی الگوهای واقعی موجود در بیماران انجام شد.

متعادل‌سازی: برای متعادل‌سازی متغیر LOS پس از پیش‌پردازش داده‌ها از روش‌های بیش‌نمونه‌گیری، و SMOTE استفاده شد تا روشی با بهترین پیش‌بینی انتخاب شود. بر اساس یافته‌ها، SMOTE به علت ایجاد داده‌های متنوع و غیرتکراری و تولید نتایج قابل‌تعمیم‌تر و همچنین کسب نتایج بهتر انتخاب شد. متعادل‌سازی نمونه‌ها تنها در مرحله آموزش مدل‌های طبقه‌بندی نظارت‌شده اعمال شد و الگوریتم‌های دسته‌بندی نیز بر روی داده‌های متعادل‌سازی شده اعمال شدند.

استخراج قوانین همبستگی: الگوریتم اپریوری برای شناسایی همبستگی‌ها بین متغیرها استفاده شد تا قوانین مربوط به روابط بین عواملی مانند جنسیت، فشارخون و طول مدت بستری شناسایی شود. این الگوریتم مستقیماً بر روی داده‌های اصلی با برچسب‌های دو کلاسه (کوتاه‌مدت / بلندمدت) به‌دست آمده از خوشه‌بندی اعمال شد. این کار به‌منظور جلوگیری از تأثیر خطاهای طبقه‌بندی الگوریتم‌های یادگیری ماشین بر قوانین استخراج‌شده بود.

آموزش مدل طبقه‌بند: مجموعه داده اولیه شامل ۱۶۹۰ مورد بود. پس از پیش‌پردازش و حذف داده‌های ناقص، داده‌های ۱۲۴۸ بیمار باقی ماند. سپس، مجموعه داده به نسبت ۸۰ درصد برای آموزش (۱۰۰۰ بیمار) و ۲۰ درصد برای آزمون (۲۴۸ بیمار) تقسیم شد. از نرم‌افزار SPSS Clemntine 12 و زبان برنامه‌نویسی R برای مدل‌سازی استفاده شد. چندین الگوریتم یادگیری ماشین ازجمله درخت تصمیم، شبکه عصبی مصنوعی و سیستم استنتاج عصبی-فازی تطبیقی (ANFIS) ارزیابی شدند. جنگل تصادفی به دلیل عملکرد برتر انتخاب شد و از طریق تنظیمات ابرپارامترها بهبود یافت. تنظیم ابرپارامترها با استفاده از جستجوی شبکه‌ای همراه با اعتبارسنجی متقابل ۱۰تایی انجام شد. مقادیر نهایی ابرپارامترهای تنظیم‌شده شامل: تعداد درخت‌ها (۱۰۰)، تعداد ویژگی‌ها در هر تقسیم (۵)، حداکثر عمق درخت‌ها (۳۰)، معیار تقسیم (Gini) و حداقل نمونه برای انشعاب (۲) بودند. این فرآیند برای سایر الگوریتم‌ها نیز اجرا شد ولی‌نتایج نشان داد که بهبود عملکرد سایر الگوریتم‌ها پس از تنظیم ابرپارامترها ناچیز بود و جنگل تصادفی مقاومت بیشتری در برابر بیش‌برازش داشت.

اعتبارسنجی خارجی: برای ارزیابی تعمیم‌پذیری، مدل با یک مجموعه داده مستقل از بیمارستان آیت‌الله طالقانی (شامل ۱۷۱۹ بیمار) اعتبارسنجی شد. عملیات پیش‌پردازش اولیه بر روی داده‌ها انجام شد. متغیر BMI از مجموعه داده‌های بیمارستان طالقانی حذف شد. از این داده‌ها، تعداد ۱۱۳۳ نفر (۶۵/۹٪) اقامت کمتر یا مساوی هفت روز و ۵۸۶ نفر (۳۴/۱٪) اقامت بالای هفت روز در بیمارستان داشته‌اند.**ارزیابی:** ارزیابی مدل‌ها با استفاده از معیارهایی مانند صحت، حساسیت، ویژگی، شاخص کاپا، سطح زیرمنحنی (AUC)و F1-score بر اساس روش‌های مرسوم اندازه‌گیری شد [۲۹-۳۳].

یافته ها

مجموعه داده اول (بیمارستان سیدالشهدا): از ۱۶۹۰ داد ، پس از تقسیم داده‌ها و حذف موارد ناقص ۲۴۸ داده به‌عنوان داده آزمون استفاده شد. جدول ۱ ویژگی‌های این داده را نشان می‌دهد. از بین ۲۷ متغیر بررسی‌شده، فقط متغیرهای ذکر شده در جدول ۱ تفاوت آماری معنی‌دار بین دو گروه کوتاه‌مدت و بلندمدت داشتند. سایر متغیرها فاقد اختلاف آماری معنی‌دار بودند و در جدول ذکر نشدند ولی در تحلیل‌ها لحاظ شده‌اند. بررسی ویژگی‌های پایه نشان داد بیماران با LOS بلندمدت به‌طور معناداری مسن‌تر بوده و شیوع بالاتری از بیماری‌های همراه (دیابت، فشارخون) و اختلالات آزمایشگاهی (کراتینین بالا، هموگلوبین پایین) داشتند.

جدول ۱. ویژگی‌های بالینی و دموگرافیک بین دو گروه (داده اول)

ویژگی	بلندمدت (۵۱۹)	کوتاه‌مدت (۱۱۷۱)	p-value
میانگین سن (سال)	۷۵/۳±۱۰/۸	۶۸/۲±۱۲/۲	<۰/۰۰۱
جنسیت مرد (٪)	۵۸/۷	۵۲/۱	۰/۰۰۱۳
فشارخون بالا (٪)	۷۸/۲	۶۵/۴	<۰/۰۰۱
دیابت (٪)	۴۵/۶	۳۲/۱	<۰/۰۰۱
فیبریلاسیون دهلیزی(٪)	۲۸/۹	۱۵/۳	<۰/۰۰۱
کراتینین بالا (٪)	۷۲/۴	۴۸/۲	<۰/۰۰۱
هموگلوبین پایین(٪)	۶۸/۹	۵۳/۱	<۰/۰۰۱
سابقه آنژیوپلاستی (٪)	۱۲/۵	۸/۷	۰/۰۲۱

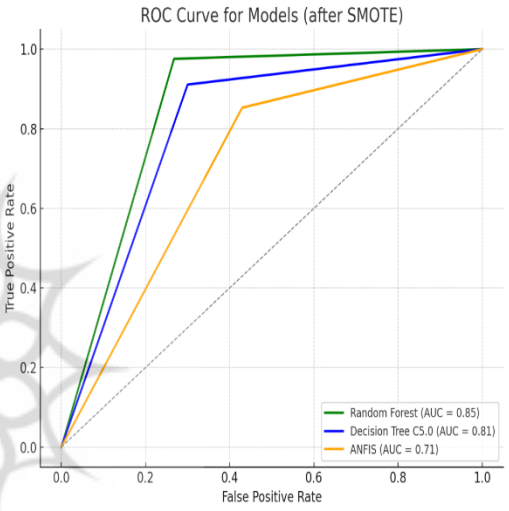
نتایج اجرای الگوریتم‌ها در مجموعه داده۱(بیمارستان سیدالشهدا) نشان داد که الگوریتم جنگل تصادفی عملکرد برتری نسبت به مدل‌های درخت تصمیم، شبکه عصبی و ANFIS داشت. همان‌طور که در جدول ۲ نشان داده شده است، جنگل تصادفی به صحت ۸۷/۱۴ ٪، حساسیت ۹۷/۵۶ ٪، ویژگی ۲۳/۲۴، AUC معادل ۵۵/۴۰ ٪ و مقدار F ۷۱/۱۳ ٪ دست‌یافت. پس از به دست آمدن این مقادیر، هر سه روش متعادل‌سازی روی داده‌ها انجام و مشخص شد که SMOTE بهترین عملکرد را دارد (جدول ۳، شکل ۱ و ۲). مدل جنگل تصادفی با مساحت زیر منحنی معادل ۰/۸۵۴ بهترین عملکرد را در تفکیک بیماران با مدت اقامت کوتاه‌مدت و بلندمدت نشان داد. بهبود زیادی در صحت بعد از متعادل‌سازی رخ نداد ولی حساسیت و F1-score نسبت به حالت نامتعادل بهبود یافت. بر اساس مدل جنگل تصادفی نهایی، مهمترین عوامل مؤثر بر پیش‌بینی مدت اقامت بیماران شناسایی شدند (جدول ۴).

جدول ۲. عملکرد مدل در مجموعه داده ۱(بیمارستان سیدالشهدا)

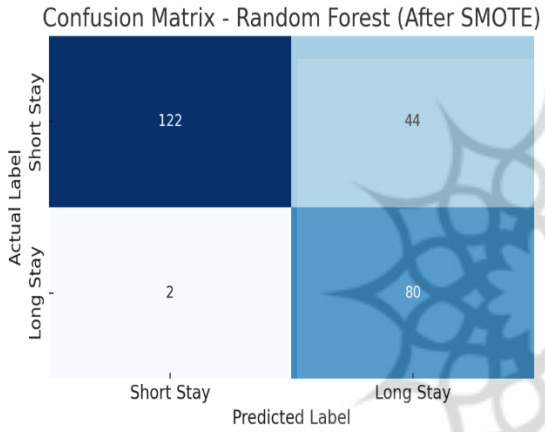
مدل	صحت (٪)	حساسیت (٪)	ویژگی (٪)	AUC (٪)	کاپا (٪)	F1 (٪)
درخت تصمیم [C5.0]	۶۷/۵۶	۹۴/۱۸	۷/۰۸	۵۱/۰۳	۰۱/۵۵	۶۵/۶۱
شبکه عصبی مصنوعی	۶۴/۲۴	۹۲/۲۴	۲۶/۸۸	۵۹/۵۶	۲۰/۷۵	۶۹/۳۶
ANFIS	۶۷/۱۵	۸۴/۴۴	۲۷/۸۶	۵۶/۱۵	۱۳/۷۳	۶۵/۸۲
جنگل تصادفی	۸۷/۱۴	۹۷/۵۶	۲۳/۲۴	۵۵/۴۰	۲۲/۹۵	۷۱/۱۳

جدول ۳. عملکرد مدل‌ها بر روی داده آزمون (بیمارستان سیدالشهدا) پس از متعادل‌سازی با روشSMOTE

مدل	صحت (٪)	حساسیت (٪)	ویژگی (٪)	AUC(٪)	کاپا (٪)	F1(٪)
C5.0	۸۵/۱۳	۹۱/۰۹	۷۰/۰۰	۸۰/۵۵	۶۴/۶۳	۷۹/۱۶
ANFIS	۷۳/۱۷	۸۵/۲۸	۵۷/۰۴	۷۱/۱۶	۴۳/۵۹	۶۸/۳۶
جنگل تصادفی	۸۱/۴۵	۹۸/۳۹	۶۴/۵۲	۸۵/۴۰	۶۲/۹۰	۸۴/۱۴



شکل ۱. منحنی ROC در مجموعه داده اول



شکل ۲. ماتریس درهم‌ریختگی در مجموعه داده آزمون اول

جدول ۴. عوامل مؤثر در پیش‌بینی مدت اقامت بیماران

متغیر	نقش در پیش‌بینی	عوامل مؤثر در پیش‌بینی مدت اقامت بیماران
سابقهCABG	مثبت	بیماران دارای سابقه جراحی بای‌پس، بیشتر از ۷ روز بستری بودند.
دیابت	مثبت	بیماران دیابتی بیشتر در گروه اقامت بلندمدت قرار داشتند.
دیس لیپیدمی	مثبت	دیس لیپیدمی با اقامت بیشتر از ۷ روز ارتباط دارد.
جنسیت مرد	مثبت	نسبت مردان در گروه اقامت >۷ روز بیشتر بود.
فشارخون بالا	منفی	فشارخون بالا بیشتر در بیماران با اقامت کمتر از ۷ روز دیده شد.
سابقهPCI	مثبت	بیماران باسابقه آنژیوپلاستی اقامت طولانی‌تری داشتند.
کراتینین بالا	منفی	کراتینین بالا همراه با فشارخون بالا با اقامت کوتاه‌تر ارتباط داشت.

با استفاده از الگوریتم اپریوری، قوانینی استخراج شدند که عوامل کلیدی مؤثر بر طول مدت بستری را نشان می‌دهند. مقادیر پشتیبانی و سطح اعتماد به‌صورت تجربی و بر اساس مقادیر مرسوم در مقالات علمی انتخاب شدند. ابتدا چندین بار الگوریتم با مقادیر مختلف اجرا شد. سپس، تیم متخصصان قوانین استخراج‌شده را بررسی کردند. مقادیر (۰/۱ و ۰/۸۵) انتخاب شد. به دلیل تمرکز بر قوانین با بالاترین مقادیر پشتیبانی و سطح اعتماد که ازنظر بالینی معنادار و مورد تائید متخصصین قلب بودند، دو قانون زیر استخراج شد:

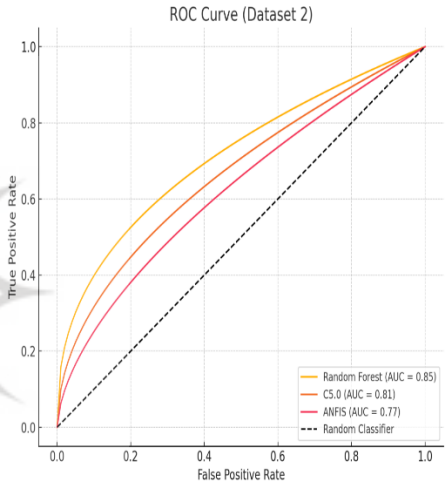
قانون ۱: بیمارانی که مرد هستند، سابقه فشارخون بالا دارند، سابقه تعویض دریچه قلب ندارند و کراتینین بالا دارند، با طول مدت بستری کوتاه‌تر مرتبط هستند. مقدار پشتیبانی این قانون ۰/۱۰۷ و سطح اعتماد آن ۰/۹۲۳ است. مقدار پشتیبانی نشان می‌دهد که ۱۰/۷ درصد از بیماران در مجموعه داده این ترکیب خاص از عوامل را نشان می‌دهند. مقدار اعتماد احتمال ۹۲/۳ درصد را نشان می‌دهد که بیماران با این شرایط طول مدت بستری کوتاه‌تری خواهند داشت.

قانون ۲: بیمارانی که سابقه فیبریلاسیون دهلیزی و کراتینین بالا دارند، ولی سابقه آنژیوپلاستی، سکته یا اعتیاد ندارند، بیشتر احتمال دارد که طول مدت بستری طولانی‌تری داشته باشند. این قانون دارای مقدار پشتیبانی ۰/۱۰۴ و سطح اعتماد ۰/۸۴۶ است. در این حالت، مقدار پشتیبانی نشان‌دهنده این است که ۱۰/۴ درصد از بیماران این الگو را نشان می‌دهند، درحالی‌که مقدار اعتماد احتمال ۸۶/۴ درصد را نشان می‌دهد که بیماران با این شرایط طول مدت بستری طولانی‌تری خواهند داشت.

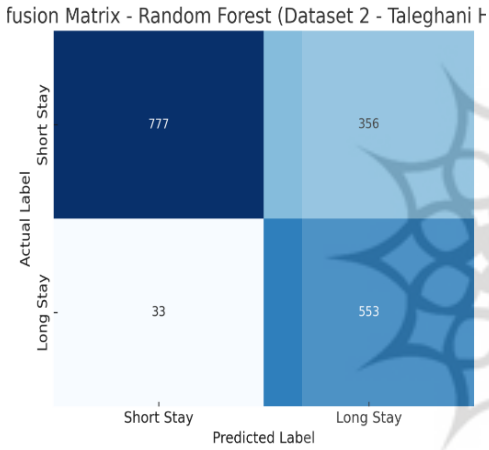
مجموعه داده دوم (بیمارستان طالقانی): داده‌های این مجموعه شامل ۱۷۱۹ بیمار بود که به دو گروه اقامت کوتاه‌مدت (۱۱۳۳ بیمار، ۶۵/۹ درصد) و بلندمدت(۵۸۶ بیمار، ۳۴/۱ درصد) تقسیم شدند. این داده‌ها در آموزش مدل نقشی نداشتند و صرفاً جهت ارزیابی استفاده شدند. جدول ۵، شکل ۳ و ۴ عملکرد الگوریتم‌ها را نشان می‌دهند. عملکرد مدل‌ها در مجموعه داده‌های اعتبارسنجی دقت مدل جنگل تصادفی را تأیید کرد، به‌طوری‌که این مدل با صحت ۷۷/۴۰ درصد و AUC معادل ۸۴/۸۲ درصد عملکرد بهتری داشت.

جدول ۵. عملکرد مدل در مجموعه داده ۲(بیمارستان طالقانی)

مدل	صحت (٪)	حساسیت (٪)	ویژگی (٪)	AUC(٪)	کاپا (٪)	F1(٪)
جنگل تصادفی	۷۷/۴۰	۹۴/۳۲	۶۸/۶۱	۸۴/۸۲	۶۵/۴۰	۷۳/۹۰
درخت تصمیم	۷۴/۰۵	۹۳/۲۴	۷۰/۲۹	۸۱/۴۷	۶۵/۴۰	۷۲/۹۱
ANFIS	۷۶/۰۳	۸۳/۶۸	۷۰/۳۰	۷۶/۹۹	۵۲/۵۰	۷۱/۹۰



شکل ۳. سطح زیر منحنی مدل‌ها در مجموعه داده دوم



شکل ۴. ماتریس درهم‌ریختگی در مجموعه داده آزمون دوم

بحث

مدل پیشنهادی در این مطالعه توانست با استفاده از الگوریتم جنگل تصادفی طول مدت اقامت بیماران مبتلا به نارسایی احتقانی قلب را با دقت بالا طبقه‌بندی کند. عملکرد این مدل در هر دو مجموعه داده آموزش و اعتبارسنجی خارجی قابل قبول بود و نتایج نشان داد که رویکردهای مبتنی بر یادگیری ماشین می‌توانند ابزار مؤثری برای پیش‌بینی مدت اقامت در بیماران CHF باشند. این یافته‌ها با نتایج مطالعه داغیستانی و همکاران [۲۲] هم‌راستا است که از الگوریتم‌های داده‌محور برای تحلیل پرونده‌های بیماران قلبی استفاده کردند. هرچند در آن مطالعه گزارشی از مدیریت داده‌های مفقوده ارائه نشده ولی در مطالعه حاضر، پیش‌پردازش دقیق و رفع داده‌های ناقص نقش مهمی در ارتقاء دقت مدل ایفا کرده است.

استفاده از الگوریتم k-means برای خوشه‌بندی طول مدت اقامت نیز امکان تفکیک دقیق‌تر گروه‌های بیمار را فراهم کرده و به ایجاد طبقه‌بندی قابل‌اتکا کمک کرده است. این روش در کنار الگوریتم‌های پیشرفته مانند جنگل تصادفی عملکرد برتری نسبت به مدل‌های ساده‌تر مانند درخت تصمیم یا ANN از خود نشان داده است. در مقایسه با پژوهش آقاجانی و همکاران [۳۴] بر روی عوامل مؤثر بر LOS در بخش جراحی عمومی در سال ۲۰۱۶ در تهران انجام‌گرفته و با استفاده از درخت تصمیم دقت ۸۴/۶۹٪ را گزارش کردند، مدل جنگل تصادفی در مطالعه حاضر عملکرد بهتری دارد. همچنین اگرچه مهارلو و همکاران [۳۵] عملکرد بالای ANFIS را برای پیش‌بینی LOS بیماران ICU پس از جراحی قلب گزارش کردند ولی در مطالعه حاضر، این الگوریتم در مقایسه با جنگل تصادفی و C5.0 عملکرد ضعیف‌تری داشت. این تفاوت‌ها ممکن است ناشی از تفاوت در جمعیت بیماران، ویژگی‌های داده‌ها یا مرحله پیش‌پردازش باشد.

در مطالعه‌ای، قلی پور و همکاران [۳۶] از الگوریتم شبکه عصبی مصنوعی برای پیش‌بینی بقای بیماران تروما و مدت اقامت در بخش و ICUاستفاده کردند. مدل آنها توانست پیامد بالینی بیماران را با دقت مناسبی (۹۳/۳۳٪) پیش‌بینی کند ولی پیش‌بینی مدت اقامت با خطای نسبی همراه بود. در مطالعه حاضر، مدل جنگل تصادفی توانست با دقت بالا و AUC مناسب، بیماران را در دو گروه اقامت کوتاه‌مدت و

بلندمدت طبقه‌بندی کند. نکته قابل‌توجه دیگر، استفاده از SMOTE برای متعادل‌سازی داده‌ها در مطالعه حاضر است که برخلاف برخی مطالعات که از روش‌های ساده‌تر مانند کم‌نمونه‌گیری استفاده کرده‌اند، به بهبود دقت مدل کمک کرده است. به‌طورکلی، استفاده از الگوریتم‌های پیشرفته یادگیری ماشین همراه با پیش‌پردازش مناسب داده‌ها و متعادل‌سازی کلاس‌ها، می‌تواند نقش مؤثری در بهبود پیش‌بینی طول مدت اقامت بیماران ایفا کند.

با استفاده از مدل پیشنهادی در این مطالعه مشخص شد که متغیرهایی مانند جنسیت، فشارخون بالا، بیماری‌های زمینه‌ای و سطح کراتینین از مهمترین پیش‌بینی کننده‌های مدت اقامت بیماران CHF در بیمارستان بودند. به‌طور خاص، سطح بالاتر کراتینین و وجود بیماری‌های زمینه‌ای با افزایش احتمال اقامت طولانی‌تر در بیمارستان همراه بودند، درحالی‌که بیماران مرد با فشارخون بالا ولی بدون سابقه تعویض دریچه قلب بیشتر در گروه اقامت کوتاه‌مدت قرار گرفتند. این یافته‌ها با مطالعات مشابهی که نقش بیماری‌های زمینه‌ای و وضعیت کلی عملکرد کلیه را در افزایش مدت بستری تأیید کرده‌اند، هم‌راستا است. برای نمونه، در پژوهشی از داغیستانی و همکاران [۲۲] نیز بیماری‌های مزمن ازجمله دیابت و فشارخون بالا به‌عنوان عواملی مرتبط با افزایش LOS معرفی شدند. به‌علاوه، مطالعات قبلی نشان داده‌اند که اختلال در عملکرد کلیه به دلیل تأثیر بر توازن مایعات و الکترولیت‌ها ممکن است روند بهبودی بیماران CHF را کند کرده و مدت اقامت را افزایش دهد [۳۷]. بنابراین، توجه به این متغیرها در هنگام پذیرش بیمار می‌تواند در پیش‌بینی مدت اقامت و مدیریت بهتر منابع بیمارستانی نقش کلیدی ایفا کند. فشارخون بالا نیز به‌عنوان یک پیش‌بینی کننده قابل‌توجه بود. مطالعاتی نشان داده‌اند که فشارخون CHF را تشدید می‌کند و به دلیل ارتباط با بیماری‌های همراه منجر به بستری طولانی‌تر می‌شود[۲۶،۳۸]. فشارخون بالا، به‌ویژه در کنار سایر بیماری‌های مزمن، می‌تواند روند کنترل وضعیت بیمار را پیچیده‌تر کند و به تأخیر در ترخیص منجر شود. این یافته با مطالعه‌ای گاتلیب و همکاران [۳۹] هم‌راستا است که نشان داد بیماران مبتلا به CHF با فشارخون بالا، به دلیل نیاز به مدیریت دقیق‌تر و خطر بالاتر عوارض، اغلب مدت‌زمان بیشتری در بیمارستان می‌مانند. همچنین، سطح کراتینین نیز به‌عنوان پیش‌بینی‌کننده قوی برای LOS شناسایی شد. افزایش سطح کراتینین نشان‌دهنده اختلال عملکرد کلیه است که می‌تواند فرآیند درمان CHF را پیچیده‌تر کند. عملکرد کلیوی ضعیف به افزایش مدت بستری منجر می‌شود زیرا این بیماران اغلب نیازمند نظارت بیشتر، دارودرمانی دقیق‌تر و مدیریت پیچیده‌تری هستند [۳۹]. فیبریلاسیون دهلیزی نیز با افزایش طول مدت اقامت مرتبط شناخته شد. این اختلال ریتم قلبی معمولاً با نارسایی قلبی هم‌زمان رخ می‌دهد و به دلیل نیاز به پایش، دارودرمانی چندوجهی و احتمال عوارض بیشتر، باعث افزایش مصرف منابع و تأخیر در ترخیص می‌شود [۴۰]. درمجموع، این نتایج اهمیت شناسایی بیماران پرخطر در ابتدای پذیرش را برجسته می‌سازد تا با پیش‌بینی دقیق LOS، مدیریت بهینه‌تری در تخصیص منابع بیمارستانی انجام گیرد.

استفاده از الگوریتم جنگل تصادفی برای پیش‌بینی طول مدت اقامت و اپریوری برای استخراج قوانین همبستگی منجر به ارایه مدلی جامع برای تحلیل LOS در بیماران CHF شد. ه مطالعات گذشته مانند هاچه‌سو و همکاران [۱۸] و تورگمان و همکاران [۱۷] بیشتر بر پیش‌بینی دقیق LOS با تکیه‌بر یادگیری ماشین متمرکز بودند ولی مطالعه حاضر با افزودن تحلیل قوانین همبستگی توانست بینش‌های تفسیری بیشتری نسبت به الگوهای مرتبط با اقامت کوتاه‌مدت و بلندمدت ارائه دهد. الگوریتم اپریوری بر روی داده‌های دوکلاسهاجرا شد و ترکیب‌های مشخصی از ویژگی‌های بیمار مانند "جنسیت مرد، فشارخون بالا و بدون تعویض دریچه قلب" را شناسایی کرد که بیشتر با اقامت کوتاه‌مدت همراه بودند. این قوانین می‌توانند مکمل خوبی برای نتایج پیش‌بینی جنگل تصادفی باشند و از آنها می‌توان در تبیین نتایج پیش‌بینی و طراحی برنامه‌های مداخه‌ای بهره گرفت.

کاربردهای عملی این یافته‌ها برای برنامه‌ریزی مراقبت و تخصیص منابع در مدیریت CHF قابل‌توجه است. پزشکان می‌توانند با استفاده از نتایج این دو رویکرد، بیماران با احتمال اقامت طولانی‌تر را زودتر شناسایی و برای مراقبت هدفمند آنها برنامه‌ریزی کنند. برای نمونه، ترکیب ویژگی‌هایی مانند فیبریلاسیون دهلیزی و سطح بالای کراتینین که با LOS طولانی‌تر مرتبط بودند، می‌تواند در طراحی مسیرهای درمانی شخصی‌سازی‌شده مؤثر باشد. همچنین، انجام اعتبارسنجی خارجی با استفاده از یک مجموعه داده مستقل، استحکام و قابلیت تعمیم‌پذیری مدل را افزایش داد و نشان می‌دهد که این مدل می‌تواند در محیط‌هایی خارج از مرکز مطالعه نیز به‌کار گرفته شود.

محدودیت‌ها

باوجود نقاط قوت، این مطالعه دارای محدودیت‌هایی است. داده‌ها از دو بیمارستان در ایران جمع‌آوری شده‌اند که ممکن است تعمیم‌پذیری به سایر جمعیت‌ها را محدود کند. علاوه‌براین، مجموعه داده شامل متغیرهایی مانند داده‌های دقیق اکوکاردیوگرافی یا سوابق دارویی نمی‌شد که می‌توانند تصویر جامع‌تری از شرایط بیمار ارائه داده و دقت پیش‌بینی را بهبود بخشند.

در تحقیقات آینده، داده‌های بالینی اضافی، به‌ویژه داده‌های مرتبط با تصویربرداری و استفاده از داروها باید برای بهبود عملکرد مدل در نظر گرفته شوند. درنهایت، بررسی روش‌های پیشرفته یادگیری ماشین، مانند یادگیری تجمعی یا یادگیری عمیق، می‌تواند بینش‌های بیشتری در مورد تعاملات پیچیده در داده‌های بیمار ارائه دهد و پیش‌بینی LOS را در CHF و شرایط مشابه بهبود بخشد.

نتیجه‌گیری

این مطالعه کارایی داده‌کاوی را در پیش‌بینی طول مدت بستری برای بیماران مبتلا به نارسایی احتقانی قلب نشان داد. پیش‌بینی دقیق LOS امکان برنامه‌ریزی بهتر و تخصیص منابع را فراهم کرده و کارایی ارائه خدمات بهداشتی برای بیماران CHF را افزایش می‌دهد. یافته‌های حاصل از این مدل می‌تواند پزشکان را در شناسایی بیماران پرخطر که

ممکن است به مراقبت‌های طولانی‌تری نیاز داشته باشند، راهنمایی کرده و مداخلات به‌موقع را تسهیل کند.

اعلان‌ها

ملاحظات اخلاقی: این پژوهش با کد اخلاق IR.UMSU.REC.1398.012 اخذشده از کمیته اخلاق در پژوهش‌های زیست پزشکی دانشگاه علوم پزشکی ارومیه انجام شده است. **حمایت مالی**: این مطالعه بخشی از پایان‌نامه کارشناسی ارشد بود که با حمایت معاونت تحقیقات و فناوری دانشگاه علوم پزشکی ارومیه انجام‌گرفته است. حامی مالی نقشی در گردآوری و تحلیل داده و نگارش مقاله نداشته است.

تضاد منافع: هیچ تضاد منافع مرتبط با این مقاله گزارش نشده است.

مشارکت نویسندگان: **مژگان اسماعیلی**: مفهوم‌سازی، طراحی مطالعه، گردآوری داده، روش‌شناسی، نرم‌افزار، اعتبارسنجی، تحلیل داده، مدیریت داده، نگارش-پیش نویس، نگارش- بررسی و ویرایش؛ **هادی لطف نژادافشار**: مفهوم‌سازی، طراحی مطالعه، گردآوری داده، روش‌شناسی، نرم‌افزار، اعتبارسنجی، تحلیل داده، مدیریت داده، نگارش-پیش نویس، نگارش- بررسی و ویرایش، سرپرستی مطالعه، تامین مالی؛ **بهلول رحیمی:** روش‌شناسی، نرم‌افزار، اعتبارسنجی، تحلیل داده؛ **کمال خادم وطنی:** گردآوری داده، روش‌شناسی، اعتبارسنجی، تحلیل داده؛ **شیرین صمدزادقوشچی:** نرم‌افزار، نگارش-پیش نویس، نگارش- بررسی و ویرایش، بصری سازی؛ **وحید حسین پور**: روش‌شناسی، اعتبارسنجی، تامین مالی. تمام نویسندگان متن نهایی مقاله را مطالعه و تایید کرده‌اند.

رضایت برای انتشار: موردی وجود ندارد.

دسترسی به داده‌ها: داده‌ها و کدهای استفاده‌شده در این مطالعه از طریق ایمیل نویسنده مسئول هادی لطف نژادافشار در دسترس است.

استفاده از هوش مصنوعی: جهت ویرایش بخش انگلیسی این مقاله از نرم‌افزار InstaText استفاده شده است. تمام محتوای ویرایش شده با این نرم‌افزار توسط نویسندگان بررسی و تائید شده است.

تشکر و قدردانی: نویسندگان از تمامی ارائه‌دهندگان خدمات بهداشتی که از این مطالعه در بیمارستان‌های سیدالشهدا و آیت‌الله طالقانی ارومیه حمایت کردند، سپاسگزارند. این مقاله حاصل بخشی از پایان‌نامه با عنوان «پیش‌بینی طول مدت اقامت بیماران نارسایی احتقانی قلب با استفاده از تکنیک‌های داده‌کاوی در بیمارستان‌های آموزشی سیدالشهدا و آیت‌الله طالقانی شهر ارومیه»، در مقطع کارشناسی ارشد مصوب دانشگاه علوم پزشکی در سال ۱۳۹۹ با کد طرح ۲۵۰۹ و کد رهگیری ۳۱۴۴ است.

منابع

1. Alemzadeh-Ansari MJ, Ansari-Ramandi MM, Naderi N. Chronic pain in chronic heart failure: a review article. The Journal of Tehran University Heart Center. 2017;12(2):49-56. Available from: /https://pmc.ncbi.nlm.nih.gov/articles/PMC5558055
2. Keyhani D, Razavi Z, Shafiee A, Bahadoram S. Autonomic function change following a supervised exercise program in patients with congestive heart failure. ARYA Atherosclerosis. 2013;9(2):150-156. PMCID: PMC3653242. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3653242
3. Writing Group Members, Rosamond W, Flegal K, et al. Heart disease and stroke statistics—2009 update: a report from the American heart association statistics committee and stroke statistics subcommittee. Circulation. 2009;119(3):e21–e181 https://doi.org/10.1161/CIRCULATIONAHA.108.191261
4. Ahmadi A, Soori H, Mobasheri M, Etemad K, Khaledifar A. Heart failure: the outcomes, predictive and related factors in Iran. Journal of Mazandaran University of Medical Sciences. 2014;24(118):180–188. [In Persian]. Available from: http://jmums.mazums.ac.ir/article-1-4636-en.html
5. Liu LC, Voors AA, van Veldhuisen DJ, van der Meer P. Heart failure highlights in 2012–2013. European Journal of Heart Failure. 2014;16(2):122–32. https://doi.org/10.1002/ejhf.43
6. Bowen RES, Graetz TJ, Emmert DA, Avidan MS. Statistics of heart failure and mechanical circulatory support in 2020. Annals of Translational Medicine. 2020;8(13):827. https://doi.org/10.21037/atm-20-1127
7. Nomali M, Mohammadrezaei R, Keshtkar AA, Roshandel G, Ghiyasvandian S, Alipasandi K, et al. Self-monitoring by traffic light color coding versus usual care on outcomes of patients with heart failure reduced ejection fraction: protocol for a randomized controlled trial. JMIR Research Protocols. 2018;7(11):e9209. https://doi.org/10.2196/resprot.9209
8. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. Nature Reviews Cardiology. 2016;13(6):368–78. https://doi.org/10.1038/nrcardio.2016.25
9. Mirdamadi A, Shafiee A, Ansari-Ramandi M, Garakyaraghi M, Pourmoghaddas A, Bahmani A, Mahmoudi H, Gharipour M. Beneficial effects of testosterone therapy on functional capacity, cardiovascular parameters, and quality of life in patients with congestive heart failure. BioMed Research International. 2014;2014:392432. https://doi.org/10.1155/2014/392432
10. Mori J, Krantz MJ, Tanner J, Horwich TB, Yancy C, Albert NM, Hernandez AF, Dai D, Fonarow GC. Influence of hospital length of stay for heart failure on quality of care. The American Journal of Cardiology. 2008;102(12):1693–1697. https://doi.org/10.1016/j.amjcard.2008.08.015
11. Azari A, Janeja VP, Mohseni A. Predicting hospital length of stay (PHLOS): a multi tiered data mining approach. In: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW). 2012. p. 17–24. https://doi.org/10.1109/ICDMW.2012.69

12. Mehdipour Y, Ebrahimi S, Karimi A, Alipour J, Khammarnia M, Siasar F. Presentation a model for prediction of cerebrovascular accident using data mining algorithm. Sadra Medical Journal. 2016;4(4):255–266. Available from: https://smsj.sums.ac.ir/article_43946_en.html
13. Ristevski B, Chen M. Big data analytics in medicine and healthcare. Journal of Integrative Bioinformatics. 2018;15(3):20170030. https://doi.org/10.1515/jib-2017-0030
14. Pasupathi C, Kalavakonda V. Evidence based healthcare system using big data for disease diagnosis. In: 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and BioInformatics (AEEICB). 2016. p. 370–4. https://doi.org/10.1109/AEEICB.2016.7538393
15. Sarafi Nejad A, Saeid A, Mohammed Rose I, Rowhanimanesh A. Modeling a data mining decision tree and propose a new model for the diagnosis of skin cancer by immunohistochemical staining methods. Journal of Health and Biomedical Informatics. 2014;1(1):54–62. Available from: http://jhbmi.ir/article-1-62-en.html
16. Tekieh MH, Raahemi B. Importance of data mining in healthcare: a survey. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2015. p. 1057–62. https://doi.org/10.1145/2808797.2809367
17. Turgeman L, May JH, Sciulli R. Insights from a machine learning model for predicting the hospital length of stay at the time of admission. Expert Systems with Applications. 2017;78:376–85. https://doi.org/10.1016/j.eswa.2017.02.023
18. Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. Healthcare Informatics Research. 2013;19(2):121–9. https://doi.org/10.4258/hir.2013.19.2.121
19. Thuraisingham B. A primer for understanding and applying data mining. IT Professional. 2002;2(1):28–31. https://doi.org/10.1109/6294.819936
20. Zhao J, Feng X, Pang Q, Fowler M, Lian Y, Ouyang M, et al. Battery safety: machine learning-based prognostics. Progress in Energy and Combustion Science. 2024;102:101142. https://doi.org/10.1016/j.pecs.2023.101142
21. Luo L, Lain S, Feng C, Huang D, Zhang W. Data mining-based detection of rapid growth in length of stay on COPD patients. In: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). 2017. p. 319–23. https://doi.org/10.1109/ICBDA.2017.8078819
22. Daghistani TA, Elshawi R, Sakr S, Ahmad A, Al-Thwayee A, Al-Mallah. Predictors of in hospital length of stay among cardiac patients: a machine learning approach. International Journal of Cardiology. 2019; 288:140–7. https://doi.org/10.1016/j.ijcard.2019.01.046
23. Neri L, Oberdier MT, van Abeelen KCJ, Menghini L, Tumarkin E, Tripathi H, et al. Electrocardiogram monitoring wearable devices and artificial-intelligence-enabled diagnostic capabilities: a review. Sensors. 2023;23(10):4805. https://doi.org/10.3390/s23104805
24. Dai W, Brisimi TS, Adams WG, Mela T, Saligrama V, Paschalidis IC. Prediction of hospitalization due to heart diseases by supervised learning methods. International Journal of Medical Informatics. 2015;84(3):189-197. https://doi.org/10.1016/j.ijmedinf.2014.10.002
25. Natale J. A strategy for reducing congestive heart failure readmissions through the use of interventions targeted by machine learning [Doctoral dissertation]. University of Akron; 2015. OhioLINK Electronic Theses and Dissertations Center. Available from: http://rave.ohiolink.edu/etdc/view?acc_num=akron1428233380
26. Messerli FH, Rimoldi SF, Bangalore S. The transition from hypertension to heart failure: contemporary update. JACC: Heart Failure. 2017;5(8):543–51. https://doi.org/10.1016/j.jchf.2017.04.012
27. Berkhin P, Becher JD. Learning simple relations: theory and applications. In: Proceedings of the 2002 SIAM International Conference on Data Mining. 2002. p. 420–36. https://doi.org/10.1137/1.9781611972726.25
28. Zebin T, Rezvy S, Chaussalet TJ. A deep learning approach for length of stay prediction in clinical settings from medical records. In: 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). 2019. p. 1–6. https://doi.org/10.1109/CIBCB.2019.8791477
29. Flach P, Blockeel H, Ferri C, Orallo JH, Struyf J. Decision support for data mining: an introduction to ROC analysis and its applications. In: Data Mining and Decision Support: Integration and Collaboration. Springer; 2003. p. 81–90. https://doi.org/10.1007/978-1-4615-0286-9_7
30. Galdi P, Tagliaferri R. Data mining: accuracy and error measures for classification and prediction. Encyclopedia of Bioinformatics and Computational Biology. 2018;1:431–6. https://doi.org/10.1016/B978-0-12-809633-8.20474-3
31. Ben-David A. About the relationship between ROC curves and Cohen's kappa. Engineering Applications of Artificial Intelligence. 2008;21(6):874–81. https://doi.org/10.1016/j.engappai.2007.09.009
32. Preda S, Oprea SV, Bâra A, Belciu (Velicanu) A. PV forecasting using support vector machine learning in a big data analytics context. Symmetry. 2018;10(12):748. https://doi.org/10.3390/sym10120748
33. Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering.

2005;17(3):299–310. https://doi.org/10.1109/TKDE.2005.50

34. Levy D, Larson MG, Vasan RS, Kannel WB, Ho KK. The progression from hypertension to congestive heart failure. JAMA. 1996;275(20):1557–62. https://doi.org/10.1001/jama.1996.03530440037034
35. Maharlou H, Niakan Kalhori S.R, 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. Healthcare Informatics Research. 2018;24(2):109–17. https://doi.org/10.4258/hir.2018.24.2.109
36. Gholipour C, Rahim F, Fakhree A, Ziapour B. Using an artificial neural networks (ANNs) model for prediction of intensive care unit (ICU) outcome and length of stay at hospital in traumatic patients. Journal of Clinical and Diagnostic Research. 2015;9(4):OC19–23. https://doi.org/10.7860/JCDR/2015/9467.5828
37. Bleumink GS, Knetsch AM, Sturkenboom MC, Straus SM, Hofman A, Deckers JW, et al. Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure: the Rotterdam Study. European Heart Journal. 2004;25(18):1614–9. https://doi.org/10.1016/j.ehj.2004.06.038
38. Sud M, Yu B, Wijeysundera HC, Austin PC, Ko DT, Braga J, et al. Associations between short or long length of stay and 30 day readmission and mortality in hospitalized patients with heart failure. JACC: Heart Failure. 2017;5(8):578–88. https://doi.org/10.1016/j.jchf.2017.03.012
39. Gottlieb SS, Abraham W, Butler J, Forman DE, Loh E, Massie BM, et al. The prognostic importance of different definitions of worsening renal function in congestive heart failure. Journal of Cardiac Failure. 2002;8(3):136–41. https://doi.org/10.1054/jcaf.2002.125289
40. Heist EK, Ruskin JN. Atrial fibrillation and congestive heart failure: risk factors, mechanisms, and treatment. Progress in Cardiovascular Diseases. 2006;48(4):256–69. https://doi.org/10.1016/j.pcad.2005.09.001

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی