Available online at www.medicinescience.org

ORIGINAL ARTICLE

Medicine Science International Medical Journal

Medicine Science 2026;15(1):25-9

Predicting intensive care unit severity: A multilayer perceptron approach to simplified acute physiology score–I (SAPS-I)

Mukadder Sanli[1], Abdulvahap Pinar[2]

[1]*İnönü University, Faculty of Medicine, Department of Anesthesiology and Resuscitation, Malatya, Türkiye*
[2]*Adıyaman University, Rectorate Unit, Adıyaman, Türkiye*

Received 05 December 2025; Accepted 21 December 2025
Available online 02 January 2026 with doi: 10.5455/medscience.2025.12.341

Abstract

Precise severity testing of illnesses in the Intensive Care Unit (ICU) is vital in the management of patients and allocation of resources. The conventional scores such as the Simplified Acute Physiology Score (SAPS-I) are based on linear models and they might not exhaust the most complicated physiological associations. Machine Learning (ML) is a promising alternative to model these non-linearities. The current study aims to construct and validate a Multilayer Perceptron (MLP) neural network to achieve the prediction of the SAPS-I score on the basis of routine clinical data gathered during the first 24 hours of ICU admission. The design was performed as a retrospective cohort study based on the use of the data created by the Death Classification ICU (n=3.600). An MLP model with one hidden layer was trained on 31 standardized variables, consisting demographics, vital signs, and laboratory parameters. The data was divided into two subsets training (70%) and testing (30%) to develop a model and to evaluate it unbiasedly. The model assessment were performed using the Sum of Squares Error and the Relative Error. The MLP model represented strong generalization, with comparable error metrics between the training (Relative Error: 0.526) and testing (Relative Error: 0.535) sets. The Sequential Organ Failure Assessment (SOFA) score was found to be the salient predictor variable by the importance analysis (100% normalized importance), followed by White Blood Cell count (63.3%) and Respiratory Rate (59.3%). Predicting the SAPS-I score with an MLP is feasible because the network captures complex nonlinear patterns in clinical data. The model's internal validation aligns with established knowledge that organ dysfunction, inflammation, and respiratory status drive severity. This matches the broader potential of ML in critical care for improved automated severity assessment, while still requiring external validation and stronger interpretability.

Keywords: Simplified acute physiology score, machine learning, multilayer perceptron, intensive care unit, severity of illness

Introduction

The Intensive Care Unit (ICU) is a highly complicated setting in which critically sick patients need constant monitoring and quick clinical decisions. Accurate evaluation of disease severity is critical for directing therapy, allocating resources, measuring institutional performance, and stratifying patients in clinical research [1]. Several prognostic scoring systems have been created to meet this requirement, the most important of which being the Simplified Acute Physiology Score (SAPS-I), which is still used in critical care.

SAPS-I, established in 1984, analyzes the worst values of 14 physiological indicators obtained during the first 24 hours of ICU admission, as well as patient age, to offer a quantitative assessment of severity [2]. Despite newer variations, such as SAPS II and SAPS III, the original SAPS-I remains a well-understood benchmark owing to its simplicity, clinical familiarity, and documented association with mortality risk [3]. Despite the later creation of updated variants such as SAPS II and III, the original SAPS-I remains a well-known benchmark and a valid depiction of the key notion that a patient's acute physiological derangement is a strong predictor of their clinical destiny.

Traditional scoring systems, like SAPS-I, are limited by the assumptions in their underlying linear statistical models. These approaches often fail to capture the intricate, nonlinear, and interactive interactions between physiological variables in critically sick patients [4]. Clinical conditions such as multiorgan dysfunction commonly entail synergistic effects that transcend linear models' representational capabilities, necessitating the investigation of more flexible, data-driven methods.

CITATION
Sanli M, Pinar A. Predicting intensive care unit severity: A multilayer perceptron approach to simplified acute physiology score–I (SAPS-I). Med Science. 2026;15(1):25-9.

Corresponding Author: Abdulvahap Pinar, Adıyaman University, Rectorate Unit, Adıyaman, Türkiye
Email: abdulvahappinar7@gmail.com

DOI: 10.5455/medscience.2025.12.341

Machine learning (ML) has enormous promise in this situation. Artificial neural networks (ANNs), particularly multilayer perceptrons (MLPs), can simulate extremely non-linear functions and automatically learn complicated variable interactions without making preconceived assumptions [5,6]. Previous research has demonstrated that ML algorithms may outperform traditional severity grading methods in predicting outcomes such as death, sepsis, and organ failure [7,8]. Nonetheless, the majority of research focuses on binary outcomes, and the application of machine learning to predict continuous severity ratings such as SAPS-I is restricted.

The purpose of this project is to create and test an MLP-based model for predicting the SAPS-I score using clinical and demographic data obtained regularly during the first 24 hours of ICU admissions. We believe that the MLP will successfully understand the nonlinear connections underlying SAPS-I and give reliable severity assessment. In addition, we investigate variable significance to determine which clinical characteristics have the most effect on the model's predictions.

Material and Methods

Study Design, Data Source, and Ethical Considerations

This study employed a retrospective cohort design utilizing the "Death Classification ICU" dataset, publicly available on the Kaggle platform [9]. The analysis involved de-identified data from patients admitted to an ICU. The primary aim was to develop a predictive model for the SAPS-I, a validated measure of disease severity in ICU patients [2]. The initial dataset contained 3600 complete patient records after the application of inclusion and exclusion criteria, which mandated the availability of data for all variables used in the model.

The research is based entirely on the secondary analysis of a publicly available, open-access dataset. The dataset used for this analysis was obtained from the Kaggle platform under the title “Intensive Care Unit Patient Outcome Prediction.” These data were derived from Electronic Health Records (EHRs) and were fully anonymized and privacy-protected by the source institution before being made publicly accessible, ensuring that they contain no personally identifiable information (PII) that could be linked to individual patients.

Accordingly, the present study does not pose any risk to the privacy or safety of individuals and is therefore considered exempt from ethical review under international ethical guidelines and Institutional Review Board (IRB/REB) policies. The dataset is published under the CC BY-NC-SA 4.0 license, which permits secondary analyses and data sharing. All ethical and legal procedures related to data collection were completed by the originating institution. As the study does not compromise patient confidentiality or security, formal ethics approval is not required.

Data Preprocessing and Variables

Prior to model development, a standard data preprocessing procedure was followed. All continuous predictor variables were standardized using Z-score normalization (mean=0, standard deviation=1) to ensure stable and efficient model training, a common practice in neural network development [7]. The categorical variable, Gender, was included as a factor. The set of predictor variables was selected based on clinical relevance and included demographic, physiological, and laboratory parameters recorded at or near ICU admission. These were: Gender, SOFA (Sequential Organ Failure Assessment) score, Age, Height, Weight, Diastolic Arterial Blood Pressure (DiasABP), Heart Rate (HR), Mean Arterial Pressure (MAP), Non-Invasive Diastolic Blood Pressure (NIDiasABP), Non-Invasive Systolic Blood Pressure (NISysABP), Respiratory Rate (RespRate), Oxygen Saturation (SpO2), Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), Albumin (Albumin), Blood Urea Nitrogen (BUN), Bilirubin (Bilirubin), Creatinine (Creatinine), Fraction of Inspired Oxygen (FiO2), Bicarbonate (HCO3), Hematocrit (HCT), Potassium (K), Lactate (Lactate), Sodium (Na), Partial Pressure of Carbon Dioxide (PaCO2), Partial Pressure of Oxygen (PaO2), Platelets (Platelets), White Blood Cell Count (WBC), pH (pH), and Mechanical Ventilation Duration (MechVentDuration). The dependent/output variable was the SAPS-I score, treated as a continuous scale variable.

Multilayer Perceptron (MLP) Model

MLP, a class of feed forward artificial neural network, was constructed to predict the SAPS-I score. The model was developed and executed using IBM SPSS Statistics software [10]. The dataset was partitioned into a training set (70%, n=2515) for model building and a testing set (30%, n=1085) for unbiased evaluation of its predictive performance; no holdout sample was reserved.

The network architecture was determined automatically by the software's algorithm. The final model consisted of an input layer with 31 units (corresponding to the 29 covariates and the coded factor of Gender), a single hidden layer with 9 units, and an output layer with a single unit. The hyperbolic tangent activation function was used in the hidden layer to model complex, non-linear relationships, while the output layer utilized an identity activation function, which is appropriate for regression tasks [6]. All covariates and the scale-dependent variable (SAPS-I) were standardized.

The model was trained using the batch training method with the scaled conjugate gradient optimization algorithm. The training hyperparameters were set as follows: initial lambda=0.0000005, initial sigma=0.00005, and an interval offset of 0.5. To prevent overfitting, a stopping rule based on the testing data was implemented, which halted training if no decrease in error was observed for one consecutive step. A maximum training time of 15 minutes was also set.

Model Evaluation

Model performance was evaluated based on its error on the testing set, which was not used during the training phase. The primary metrics for assessment were the Sum of Squares Error (SSE) and the Relative Error. Additionally, the normalized importance of each independent variable was analyzed to interpret the model and identify the most influential predictors of the SAPS-I score.

Result

Model Development and Dataset Characteristics

A MLP artificial neural network was implemented to predict the SAPS-I. The analysis utilized a dataset of 3600 patient records, which was partitioned into a training set (69.9%, n=2515) for

DOI: 10.5455/medscience.2025.12.341

model development and a testing set (30.1%, n=1085) for unbiased evaluation. No holdout sample was used.

The final network architecture, determined automatically, consisted of three layers. The input layer incorporated 31 units, representing one categorical factor (Gender) and 29 continuous covariates, all standardized prior to analysis. A single hidden layer containing 9 units was utilized, employing the hyperbolic tangent activation function to model non-linear relationships. The output layer consisted of a single unit with an identity activation function, suitable for the regression task of predicting the standardized SAPS-I score. The model was trained using the batch method with a scaled conjugate gradient optimizer.

Model Performance and Predictive Accuracy

The model demonstrated efficient learning, with the training process concluding after 0.33 seconds when the stopping rule was met—specifically, one consecutive step with no decrease in error based on the testing sample. The performance metrics for both the training and testing samples are summarized in Table 1. SSE was 661.248 for the training sample and 295.059 for the testing sample. The relative error was 0.526 for training and 0.535 for testing. The close alignment of error metrics between the training and testing datasets indicates that the model generalized well to unseen data without substantial overfitting.

Independent Variable Importance

The analysis of normalized importance (Table 2) revealed the relative contribution of each predictor variable to the model. The SOFA score was the most salient predictor of SAPS-I, with a normalized importance of 100.0%. This value demonstrates its foundational role in assessing overall illness severity. White blood cell count (WBC, 63.3%) and respiratory rate (RespRate, 59.3%) emerged as the next most critical physiological predictors. Other variables of moderate importance included Mean Arterial Pressure (MAP, 33.8%), non-invasive systolic blood pressure (NISysABP, 34.2%), partial pressure of carbon dioxide (PaCO2, 33.5%), and age (31.0%). In contrast, Gender (3.1%) and the fraction of inspired oxygen (FiO2, 3.8%) demonstrated the lowest predictive importance within the model.

Table 1. Performance summary of the multilayer perceptron model

Sample	Size	Sum of Squares Error	Relative Error
Training	2515	661.248	0.526
Testing	1085	295.059	0.535

This table reports the training and testing performance of the model using the ICU outcome dataset. “Sum of Squares Error (SSE)” represents the total squared difference between observed and predicted values, while “Relative Error” indicates the proportion of prediction error relative to the observed range

Table 2. Normalized importance of predictor variables for SAPS-I prediction

Feature	Importance	Normalized Importance
Gender	0.004	3.1%
SOFA	0.121	100.0%
Age	0.038	31.1%
Height	0.019	15.8%
Weight	0.033	27.0%
DiasABP	0.018	14.9%
HR	0.028	22.8%
MAP	0.041	33.8%
NIDiasABP	0.011	9.4%
NISysABP	0.041	34.2%
RespRate	0.072	59.3%
SpO2	0.027	22.1%
ALT	0.048	39.3%
AST	0.036	30.0%
Albumin	0.025	21.0%
BUN	0.023	19.1%
Bilirubin	0.028	22.9%
Creatinine	0.020	16.7%
FiO2	0.005	3.8%
HCO3	0.035	29.0%
HCT	0.028	22.7%
K	0.028	23.0%
Lactate	0.037	30.2%
Na	0.030	24.4%
PaCO2	0.041	33.5%
PaO2	0.035	28.9%
Platelets	0.020	16.5%
WBC	0.077	63.3%
pH	0.016	13.5%
MechVentDuration	0.016	13.3%

DiasABP: Diastolik arterial pressure, HR: Heart rate, MAP: Median arterial pressure, NIDiasABP: Noninvaziv diastolic arterial blood pressure, NISysABP: Noninvaziv systolic arterial blood pressure, RespRate: Respiratory rate, SpO2: Saturation, ALT: Alanin aminotransferaz, AST: Aspartat transferaz, BUN: Blood urea nitrojen, FiO2: Fraction of inspired oxygen, HCO3: Bicarbonate, HCT: Hematocrit, K: Potassium, Na: Sodium, PaCO2: Arterial carbon dioxide pressure, PaO2: Arterial oxygen pressure, WBC: Whıte blood cell, pH: Blood acidity level, Mechventduration: Mechanical ventilation duration. This table summarizes the relative contribution of each predictor variable to the multilayer perceptron model trained on the ICU outcome dataset. “Importance” reflects each variable’s raw contribution, whereas “Normalized Importance” expresses this contribution as a percentage relative to the most influential variable (set to 100%).

DOI: 10.5455/medscience.2025.12.341

Discussion

The present study successfully developed and evaluated a MLP artificial neural network to predict the SAPS-I in a cohort of ICU patients. The primary findings indicate that the MLP model demonstrated stable learning and generalization capability, with closely aligned performance metrics between the training and testing subsets. Furthermore, the analysis of variable importance provided clinically intuitive insights, revealing the SOFA score as the paramount predictor and uncovering the relative weights of other physiological and laboratory parameters in determining illness severity. This discussion will interpret these key results in the context of existing literature, consider the clinical and methodological implications, acknowledge the study's limitations, and suggest directions for future research.

Interpretation of Model Performance and Generalizability

The core objective of any predictive model is not merely to fit the training data; but to perform robustly on unseen data. The performance of our MLP model is promising in this regard. The SSE and Relative Error for the testing set (295.059 and 0.535, respectively) were only marginally higher than those for the training set (661.248 and 0.526). This close alignment is a strong indicator that the model generalized effectively without succumbing to significant overfitting, a common pitfall in complex neural networks with numerous parameters [7]. The architectural choice of a single, compact hidden layer with 9 units, determined automatically by the algorithm, likely contributed to this generalizability by preventing an excessively complex model that could have memorized the training data noise. The remarkably short training time of 0.33 seconds, facilitated by the efficient scaled conjugate gradient optimizer, underscores the computational feasibility of applying such models even with a substantial number of input variables.

When considering the application of ANNs in critical care, our results align with a growing body of evidence. For instance, studies predicting outcomes like mortality or length of stay have consistently shown that ANNs can capture complex, non-linear relationships between patient covariates that may be missed by traditional logistic or linear regression [8]. Our model's ability to map 31 diverse inputs to a continuous severity score exemplifies this strength. The low relative error suggests that the model learned the underlying functional mapping of the SAPS-I score with reasonable accuracy. However, it is crucial to note that while the relative error provides a useful metric for model development, its clinical meaning is not immediately intuitive. Future work should complement this with metrics like the Mean Absolute Error (MAE) on the original SAPS-I scale to give clinicians a clearer understanding of the average prediction.

Clinical Relevance of Variable Importance

Perhaps the most insightful finding from this analysis is the normalized importance of the predictor variables, which offers a data-driven hierarchy of factors contributing to SAPS-I. The dominance of the SOFA score, with a normalized importance of 100%, is both expected and validating. The SAPS-I and SOFA scores, while distinct, are both comprehensive severity scores that assess dysfunction across multiple organ systems [2,11]. The strong predictive power of SOFA for SAPS-I underscores the centrality of multi-organ failure in determining the overall acuity of ICU patients. It confirms that our model is leveraging a key, clinically established construct to inform its predictions.

Beyond SOFA, the model identified white blood cell count (WBC) and respiratory rate as the second and third most important predictors (63.3% and 59.3%, respectively). This highlights the critical role of two fundamental physiological systems: inflammation and respiration. An elevated WBC is a classic marker of systemic inflammatory response, often seen in sepsis, which is a leading cause of ICU admission and mortality [12]. Similarly, respiratory rate is a vital sign that is highly sensitive to physiological deterioration, and its importance is reflected in its inclusion in early warning scores outside the ICU [13]. The high ranking of these variables suggests that the model has independently identified their paramount importance, reinforcing their value in bedside assessment.

The moderate importance assigned to hemodynamic parameters like MAP and Non-Invasive Systolic Blood Pressure (NISysABP), along with age and PaCO2, further aligns with physiological principles. Adequate perfusion pressure, as represented by MAP, is crucial for end-organ function, and its derangement is a key indicator of shock. Age is a well-known prognostic factor in ICU scores, including the original SAPS-I, due to the association between advanced age and reduced physiological reserve [2]. The inclusion of PaCO2 reflects the importance of ventilatory adequacy and acid-base balance.

Conversely, the very low importance of Gender (3.1%) and FiO2 (3.8%) is equally noteworthy. The minimal role of Gender suggests that, after accounting for the other physiological and clinical variables in the model, it adds little incremental predictive power for the SAPS-I score in this specific cohort. The low importance of FiO2 might be initially surprising, as it is a key component in assessing respiratory failure. However, its low ranking could indicate that its information is already captured by other, more directly consequential variables like SpO2, PaO2, and respiratory rate. The model may be discerning that it is the outcome of oxygen exchange (e.g., SpO2) and ventilation (e.g. PaCO2) that is more directly influential on the score, rather than the therapeutic input (FiO2) itself.

Methodological Considerations and the "Black Box" Dilemma

The use of an MLP model presents a classic trade-off between predictive power and interpretability. While traditional regression models provide easily interpretable coefficients, the MLP's strength lies in its ability to model complex, non-linear interactions between variables through its hidden layers [6]. The parameter estimates table, with its matrices of synaptic weights, is a testament to this complexity. While these weights are essential for the model's function, they are not independently interpretable in the way a regression coefficient is. This "black box" nature is a significant limitation of ANN models in clinical

DOI: 10.5455/medscience.2025.12.341

settings, where understanding the rationale behind a prediction can be as important as the prediction itself.

Our study took a step towards mitigating this issue by employing variable importance analysis. This technique helps illuminate which inputs the model found most useful overall, providing a crucial layer of interpretability. However, it does not reveal the nature of the relationships (e.g. whether high WBC is positively or negatively associated with SAPS-I) or complex interaction effects. Future research could employ more advanced interpretability techniques, such as SHAP (SHapley Additive exPlanations) values, which can quantify the contribution of each variable to individual predictions, thereby opening the black box further [14].

Limitations and Future Directions

Several limitations of this study must be acknowledged. First, the use of a single, retrospective dataset from one source risks introducing spectrum bias and limits the external validity of our findings. The model requires validation on prospective, multi-center cohorts to ensure its generalizability across different patient populations and ICU practices. Second, illness severity in the ICU is dynamic, and a model incorporating time-series data or trends in variables over the first 24 hours could potentially yield more accurate and clinically relevant predictions [7]. Third, as previously mentioned, the black-box nature of the MLP, even with variable importance measures, remains a barrier to clinical adoption. Comparing the performance of our MLP against more interpretable models, such as logistic regression or decision trees, on the same dataset would be a valuable future step to determine if the gain in accuracy justifies the loss of interpretability. Finally, our model predicts the SAPS-I score itself. A more direct clinical application would be to use these same predictors to forecast hard patient outcomes, such as in-hospital mortality or the development of specific organ failures, and to compare the MLP's performance for that task against the standard SAPS-I score.

Conclusion

In conclusion, this study demonstrates that a MLP neural network can effectively learn the complex mapping from a set of routine clinical variables to the SAPS-I severity score. The model generalized well to unseen data, and its internal variable importance rankings aligned strongly with established clinical knowledge, with the SOFA score, inflammatory markers, and respiratory parameters being the most influential predictors. While challenges regarding interpretability and external validation remain, this work adds to the evidence supporting the use of advanced, non-linear modeling techniques in critical care. By leveraging these methods, we can move towards more sophisticated and potentially more accurate tools for risk stratification, which could ultimately support clinical decision-making and improve resource allocation in the intensive care unit. Furthermore, current research promotes the incorporation of MLP-based models into hospital information systems to allow earlier risk classification and expedite critical care procedures. Using real-time data streams, such models can detect clinical deterioration faster and give doctors with additional decision help at the bedside [15]. As a result, modifying and expanding the model created in this work for use in institutional databases might improve the operational accuracy of diagnostic evaluation and risk assessment in the critical care unit.

Conflict of Interests

The authors declare that there is no conflict of interest in the study.

Financial Disclosure

The authors declare that they have received no financial support for the study.

Ethical Approval

Due to the public availability of the data, ethical committee approval was unnecessary for this study.

References

1. Vincent JL. Critical care--where have we been and where are we going? Crit Care. 2013;17:S2.
2. Le Gall JR, Philippe L, Annick A, et al. A simplified acute physiology score for ICU patients. Crit Care Med. 1984;12:975-7.
3. Castella X, Artigas A, Bion J, Kari A. A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. The European/North American Severity Study Group. Crit Care Med. 1995;23:1327-35.
4. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (saps ii) based on a European/North American multicenter study. JAMA. 1993;270:2957-63.
5. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med. 2019;25:24-9.
6. Géron A, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 2nd Edition. O'Reilly Media Inc. 2019.
7. Moor M, Rieck B, Horn M, et al. Early prediction of sepsis in the ICU using machine learning: A systematic review. Front Med (Lausanne). 2021;8:607952.
8. Shillan D, Sterne JAC, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. Crit Care. 2019;23:284.
9. De Moraes Bajolin, F. Death Classification ICU 2023 [cited 2025 2025]; Available from: https://www.kaggle.com/datasets/fdemoribajolin/death-classification-icu.
10. IBM Corp. IBM SPSS Statistics for Windows, version 25.0. Armonk, NY: IBM Corp. 2017.
11. Vincent JL, Moreno R, Takala J, et al. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. on behalf of the working group on sepsis-related problems of the european society of intensive care medicine. Intensive Care Med. 1996 Jul;22:707-10.
12. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315:801-10.
13. Smith GB, Prytherch DR, Meredith P, et al. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. Resuscitation. 2013;84:465-70.
14. Singh R, Lanchantin J, Sekhon A, Qi Y. Attend and predict: understanding gene regulation by selective attention on chromatin. Adv Neural Inf Process Syst. 2017;30:6785-95.
15. Thorsen-Meyer HC, Nielsen AB, Nielsen AP, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digit Health. 2020;2:e179-e191.