JOURNAL OF ENGINEERING MANAGEMENT AND COMPETITIVENESS (JEMC)
VOL. 15, NO. 1, 2025, 57-64

# A LOGISTIC REGRESSION-BASED MODEL FOR PREDICTING HEART FAILURE MORTALITY

DOI: 10.5937/JEMC2501057K

UDC: 005.521:004.6
Original Scientific Paper

**Marija KRSTIĆ[1], Lazar KRSTIĆ[2]**

[1]Academy of Applied Studies Southern Serbia, Department of Higher Business School Leskovac, 16000 Leskovac, Partizanska 7, Republic of Serbia
Corresponding Author, Email: krstic.marija@vpsle.edu.rs
ORCID ID (https://orcid.org/0000-0003-3009-8400)
[2]Academy of Applied Studies Southern Serbia, Department of Higher Business School Leskovac, 16000 Leskovac, Partizanska 7, Republic of Serbia
ORCID ID (https://orcid.org/0000-0001-9131-6876)

Paper received: 17.10.2024.; Paper accepted: 28.02.2025.

**Recent trends in evaluating World Wide Web data include the use of traditional data mining techniques, such as regression, clustering, and classification. This paper aims to develop a model for predicting heart failure mortality based on a publicly available online dataset containing medical records of 299 patients. Since the prediction outcome can have only one of two possible values, the binary logistic regression technique was applied. Research shows that the predictive model created using logistic regression can accurately predict patient mortality based on their clinical characteristics and identify the most significant attributes among those included in their medical records. In addition, applying logistic regression ensures the simplicity and interoperability of the developed model, which was a major drawback of previous studies. The prediction model was created using the RapidMiner software tool. Its contribution lies in incorporating a broader range of clinical attributes, leading to a more comprehensive approach that enhances accuracy and prediction efficiency. The accuracy, precision, and sensitivity values of the developed predictive model are approximately 80%, confirming the model's high quality. The Area Under the Curve (AUC), which provides a graphical overview of the model's overall performance, is 86.7%, reflecting its effectiveness. The indicators of the developed model exhibit strong overall performance, creating the potential for its application to assist healthcare institutions in assessing the clinical status of patients with cardiovascular diseases.**

**Keywords**: Predictive model; Logistic regression; Heart failure; Clinical characteristics; Patient mortality.

## INTRODUCTION

Heart failure is a leading cause of morbidity and mortality worldwide (Savarese et al., 2023). The daily mortality statistics for these patients are devastating, and there is a real need to develop effective prognostic models to assess patient outcomes (Shahim et al., 2023). Predictive models can help identify patients at high risk of adverse outcomes, facilitating timely intervention and improving overall survival rates.

Previous research dealing with the prediction of death in patients with heart failure is based on the application of various machine-learning techniques. Although the high accuracy of these models has been confirmed, their main drawbacks are simplicity and interoperability. To meet the requirements of simplicity and interoperability, a machine learning technique was selected, which offers the following main advantages. In addition, the contribution of the research is reflected in the inclusion of a wider range of clinical attributes, which leads to a more comprehensive approach that improves the accuracy and efficiency of predictions.

The subject of the research is the development and evaluation of a predictive model for assessing the death outcome of patients with heart failure using binary logistic regression. The research is based on the analysis of a publicly available medical data set

ISSN 2217-8147 (Online)

Available online at http://www.tfzr.uns.ac.rs/jemc

with 299 patients and the identification of the most significant clinical factors influencing mortality.

The aim of this paper is the development and evaluation of a predictive model using binary logistic regression to assess its effectiveness in predicting the mortality of patients with heart failure, as well as to improve existing methodologies through a broader analysis of patient's clinical characteristics. The paper is divided into six sections. The first part of the paper includes the theoretical background on logistic regression, the second part presents research dealing with the same topic, and the third part covers the research methodology. In the fourth part of the paper, the created model and explained research results are presented, followed by a discussion and concluding considerations with summarised results and suggestions for future research.

In this prognostic study, a logistic regression-based prediction model was developed that can predict patient mortality based on their clinical attributes. Unlike other studies dealing with the same or similar topics, this model's advantage is reflected in including a more significant number of clinical characteristics, contributing to its greater accuracy and predictive efficiency.

## BACKGROUND

Regression is mainly used in the case of predictive analyses, i.e., to predict the value of a dependent variable based on one or more independent predictor variables (Zapf et al., 2024). There are different forms of regression, such as linear, multiple, logistic, polynomial, and non-parametric. In single and multiple linear regressions, the criterion and predictor(s) are numerical variables (Maulud et al., 2020). On the other hand, binary logistic regression is used when there is binary data on the dependent categorical variable (for example, some yes/no outcome is predicted by other categorical or numerical variables). This method predicts the probability of one or another binary outcome. The two aforementioned regression techniques are most widely used in practice (Schober et al., 2021; Zaidi et al., 2023). Regression provides a hypothetical model of the relationship between the criterion and the predictor (Gregorich et al., 2024). Logistic regression is used to predict the probability of one or another binary outcome, i.e., when there is binary data on the dependent categorical variable (Harris, 2021).

There are three types of logistic regression (Ranganathan et al., 2017):

- Binary logistic regression - used when the answer is binary, i.e., there are two possible outcomes (for example, pass or fail the test) (Wilson et al., 2024);
- Nominal logistic regression - used when there are three or more categories without natural comparison by level (for example, departments in the company - marketing, sales, human resources) (Dinges et al., 2023);
- Ordinal logistic regression - used when there are three or more categories with a natural comparison by levels, but the ranking of the levels does not necessarily mean that the intervals between them are equal (for example, answers of how students evaluate the effectiveness of the faculty - good, medium, bad) (Wurm et al., 2021).

Logistic regression is used in medical research for predictive modelling, especially in estimating binary outcomes such as survival or mortality. This technique allows for the estimation of the probability of an event based on multiple prognostic variables.

This study evaluates the effectiveness of logistic regression in predicting mortality among heart failure patients. The study results could help doctors make clinical decisions regarding the early identification of high-risk patients. The creative model can identify patients at high risk of death by assessing several key clinical characteristics (age, ejection fraction, serum creatinine, and time). All previous studies involve building a predictive model based on a maximum of three clinical characteristics of the patients.

## LITERATURE REVIEW

The paper uses the method of a systematic search of electronic databases. Namely, three e-repositories, ScienceDirect, Google Scholar, and ResearchGate, were selected, and the relevant literature was searched. During the search, the criteria were set for the works to be published in the period from 2018 to 2024. The total number of studies on heart failure prediction amounted to 27. However, additional criteria were applied to filter the studies, requiring that the papers be based on the application of machine learning techniques and emphasise their importance in predicting heart disease. Furthermore, only studies where the model

accuracy was greater than 70% were included. After analysing the identified studies and removing duplicates, five papers were selected as relevant for this research. Each paper emphasises the importance of clinical characteristics, feature selection, and model evaluation to achieve high accuracy and practical applicability. Together, these studies contribute to the growing body of knowledge in predictive medicine, offering insights into best practices for building reliable and efficient models.

Selected studies deal with exploring the application of machine learning techniques in predicting heart failure outcomes, focusing on different clinical characteristics and modelling approaches. One study highlights the significance of serum creatinine and ejection fraction as key predictive factors, using a stratified logistic regression model that achieved an accuracy of 83.8%, with an actual negative rate of 86.0%, a valid positive rate of 78.5%, and a Receiver Operating Characteristic - Area Under the Curve (ROC AUC) of 82.2%. The differences in results compared to the present research can be attributed to variations in the number of independent variables used, the proportion of data allocated for training, and the choice of modelling techniques (Chicco et al., 2020). Another study applied machine learning algorithms to predict heart disease, identifying the Support Vector Machine (SVM) algorithm as the most effective, achieving an accuracy of 85.2% (Sahoo et al., 2020). Similarly, research investigating heart attack prediction through feature selection techniques found that the SVM algorithm outperformed other methods, ensuring a model accuracy of 84.81%. The study also identified the relief method as the optimal technique for selecting the most influential clinical features (Takci, 2018). Further research focused on predicting heart disease based on patient's medical history using machine learning algorithms. In this case, the K-Nearest Neighbors (KNN) algorithm demonstrated the highest effectiveness, achieving an accuracy of 88.52% (Jinda et al., 2021). Another study explored machine learning techniques for coronary heart disease prediction, finding that the SVM algorithm produced the best results with an accuracy of 73.8% (Khdair et al., 2021). These studies collectively demonstrate the potential of machine learning in predicting cardiovascular outcomes, with different algorithms yielding varying levels of accuracy depending on the clinical features considered and the methodologies employed.

## RESEARCH METHODOLOGY

### Research aims and research questions.

The research aims to develop a model with good overall performance for predicting mortality outcomes due to heart failure based on the clinical characteristics of patients from the selected dataset. The result of the research will be given through answers to the following research questions:
**RQ1**: Is the accuracy of the created prediction model satisfactory?
**RQ2**: What does the Area Under the Curve show, is the overall performance of the model good?

### Research sample

The dataset contains the medical records of 299 patients with heart failure, collected over a one-year follow-up period, with each patient profile having twelve clinical characteristics. The data were collected from patients who were under medical supervision for heart failure in a hospital in France. The dataset contains information about their clinical characteristics and treatment outcomes (MVD, n.d.). Clinical features observed during the follow-up period include:

- age of the patient,
- anaemia (if the patient has anaemia 1, otherwise 0),
- high blood pressure (if the patient has high blood pressure 1, otherwise 0),
- creatinine phosphokinase (CPK enzyme level in the blood),
- diabetes (if the patient has diabetes 1, otherwise 0),
- ejection fraction (percentage of blood that leaves the heart with each contraction),
- platelets,
- gender (if the patient is female 0 if the patient is male 1),
- serum creatinine (level of serum creatinine in the blood),
- serum sodium (sodium level in the blood),
- smoking (if the patient is a smoker 1, if the patient is not a smoker 0),
- time (patient monitoring period in days),
- death event (patient died during follow-up period 1, the patient did not die during follow-up period 0).

## Research method

### *Data pre-processing*

In the pre-processing stage, the data is prepared to create a model in the RapidMiner software tool (RapidMiner, n.d.). The dataset was downloaded as a .csv file. Since there were no incorrect records or missing values in the data set, only the appropriate data types were selected during its loading. After loading the data set, the training and testing data were split. The ratio that resulted in the highest accuracy for testing model performance was a 50:50 data split. The 50:50 division gave the best accuracy because it works with a small data set, and the model has a small number of parameters and is not prone to overloading (situations in which this way of data division is justified). The stated reasons led to a deviation from the usual practice, which implies a division of 70:30 or 80:20, where the model's accuracy was lower (slightly more than 70%).

### *Selection of dependent and independent variables*

After the data pre-processing phase, attribute selection was performed. The event to be predicted is death, which, in this particular case, represents the dependent variable. The clinical characteristics that provide the model with the highest predictive power were chosen as the independent variables: age, ejection fraction, serum creatinine, and time. Dependent variables were selected by testing the significance of individual predictors (for each independent variable, a separate logistic regression was performed with the dependent variable). Based on the p-value, statistically significant variables were identified ($p < 0.05$). The binary logistic regression method was applied as the binary criterion's dependent variable, which has only two values, 0 or 1.

## RESULTS

The software tool RapidMiner was used to build the model. The process of model creation includes loading the dataset and adjusting data types (numerical and binary variables), data pre-processing, splitting the dataset into training and test sets, model creation (applying a machine learning technique, i.e., logistic regression), model evaluation (using the Apply Model operator to test the model and Performance to analyse model accuracy with metrics such as AUC, accuracy, precision, and recall), and visualisation and analysis of results. The model is presented in the Figure 1**.**

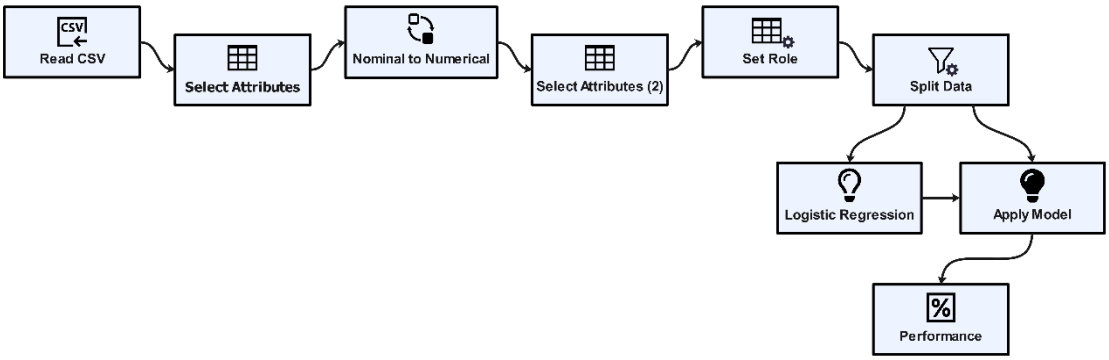


*Figure 1: A model for predicting heart failure mortality based on clinical characteristics of patients*

Table 1 provides an overview of the created model's results. In addition to the values of the dependent variable (death event) and the selected independent variables (age, ejection fraction, serum creatinine, and time), the predicted outcome for each case is also presented. The prediction result is added in a new column named "prediction." The following two columns contain confidence values, indicating the model's confidence in predicting the outcome for each case.

The Confusion Matrix describes the performance of a classification model on a set of test data for which the actual values are known (Krstinić et al., 2020). This matrix shows the distribution of correct and incorrect classifications for each class (Heydarian et al., 2022).

The Confusion Matrix displays four parameters: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). In this specific case, these parameters are presented in Table 2 with the following values: true positives 32 (the number of cases where the model correctly predicts a positive outcome), false positives 16 (the number of cases where the model incorrectly

classifies a positive outcome as negative), true negatives 86 (the number of cases where the model correctly predicts a negative result), and false negatives 15 (the number of cases where the model incorrectly classifies a negative outcome as positive).

*Table 1: Classification of the predicted outcome of the event (extract from the table)*

| Row No. | Death event | prediction | confidence(1) | confidence(0) | age | ejection fraction | serum creatinine | time |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0.971 | 0.029 | 65 | 0.200 | 1.300 | 7 |
| 2 | 1 | 1 | 0.980 | 0.020 | 90 | 0.400 | 2.100 | 8 |
| 3 | 1 | 1 | 0.989 | 0.011 | 75 | 0.150 | 1.200 | 10 |
| 4 | 1 | 1 | 1.000 | 0.000 | 80 | 0.350 | 9.400 | 10 |
| 5 | 1 | 1 | 0.989 | 0.011 | 75 | 0.380 | 4 | 10 |
| 6 | 1 | 1 | 0.919 | 0.081 | 62 | 0.250 | 0.900 | 10 |
| 7 | 0 | 1 | 0.751 | 0.249 | 49 | 0.300 | 1 | 12 |
| 8 | 1 | 1 | 0.836 | 0.164 | 82 | 0.500 | 1.300 | 13 |
| 9 | 1 | 1 | 0.897 | 0.103 | 45 | 0.140 | 0.800 | 14 |
| 10 | 1 | 1 | 0.949 | 0.051 | 70 | 0.250 | 1 | 15 |
| 11 | 1 | 0 | 0.326 | 0.674 | 48 | 0.550 | 1.900 | 15 |
| 12 | 1 | 1 | 0.838 | 0.162 | 68 | 0.350 | 0.900 | 20 |
| 13 | 1 | 1 | 0.963 | 0.037 | 75 | 0.300 | 1.830 | 23 |
| 14 | 1 | 1 | 0.959 | 0.041 | 82 | 0.300 | 1.200 | 26 |
| 15 | 1 | 1 | 0.975 | 0.025 | 94 | 0.380 | 1.830 | 27 |

*Table 2: Confusion Matrix for the created model*

| | true 1 | true 0 | class precision |
|---|---|---|---|
| pred. 1 | 32 | 15 | 68.09% |
| pred. 0 | 16 | 86 | 84.31% |
| class recall | 66.67% | 85.15% | |

Table 3 presents the results of PerformanceVector for the created model, which, in addition to the previously described parameters, also enables the display of values for Precision and Recall parameters. Precision is a parameter that shows the ratio of positive examples correctly identified in the number of positively predicted classes and is 84.31%. In comparison, sensitivity is a parameter that shows the ratio of positive examples correctly identified in the number of really positive classes and is 85.15%. Also, the Kappa coefficient of the model is presented, which is 52.1%. This coefficient shows a moderate agreement between the predicted and the actual state, as it is in the range of 41% to 60% (full interpretation scale: 20% Poor, 21 - 40% Fair, 41 - 60% Moderate, 61 - 80% Good, 81 - 100% Very Good) (McHugh, 2012). The described parameters are used to test the performance of the classification model.

The Accuracy, Precision, and Recall parameters of the prediction model created are around 80%, which is much higher than the satisfactory value, proving the model's quality.

AUC is the area under the ROC curve (Verbakel et al., 2020). This curve is obtained by plotting the values representing the ratio of true positive and false positive results (Polo & Miot, 2020) (Figure 2).

*Table 3: Results of PerformanceVector for the created model*

| | |
|---|---|
| accuracy | 79.19% |
| kappa | 52.1% |
| AUC | 86.7% |
| precision | 84.31% |
| recall | 85.15% |

AUC is a graphical representation of the overall performance of a model (Carrington et al., 2022). Ideally, its value would be 1, meaning there are no false positives or negatives (Peng et al., 2020). Suppose its value exceeds 80%, a specific example is 86.7%. In that case, the model is considered to have very good performance and would be helpful in practice, as evidenced by its comparison with the thresholds for interpreting AUC values (90 - 100% excellent, 80 - 89% very good, 70 - 79% good, 60 - 69% poor, 50 - 59% failure, 50% chance) (Koo et al., 2016).

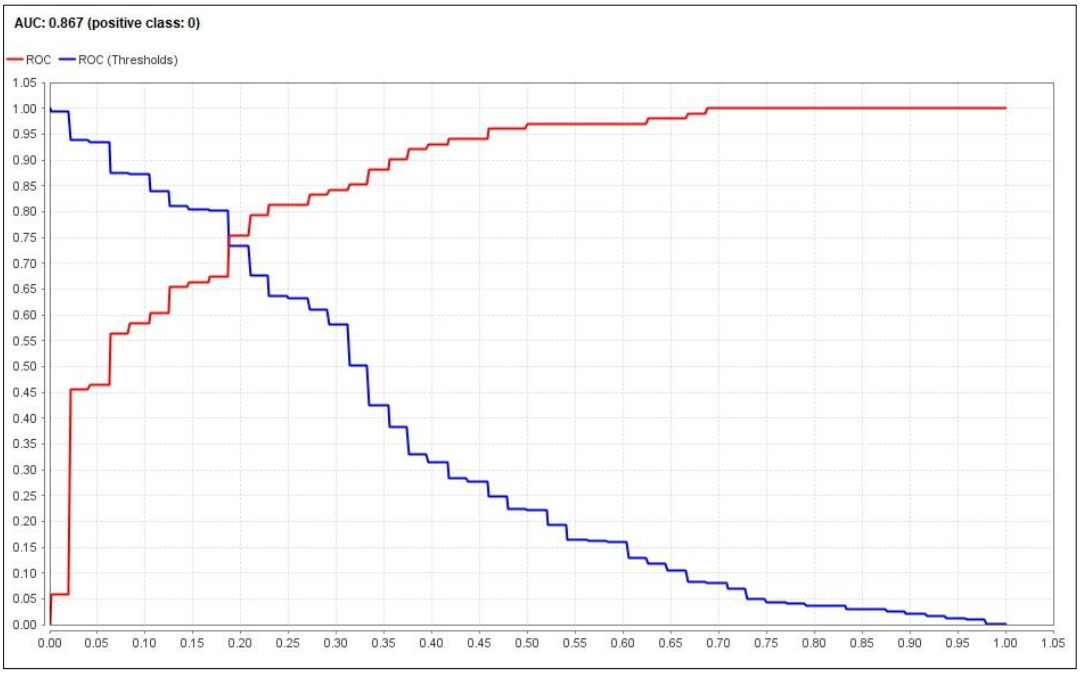


*Figure 2: Graphic representation of Area Under the Curve (AUC)*

## DISCUSSION

The model's accuracy is a parameter that identifies a part of correctly classified examples on a given test set, and the value of this parameter is 79.19%. The model's misclassification error rate is 20.81%. In case the outcome of the event is death, the model classified 32 cases, and this prediction is correct. Fifteen fatal cases were classified as incorrect predictions. The model predicted 16 cases with a fatal outcome, which was not true, while 86 cases were predicted to be non-fatal, and this prediction was correct.

The answer to the first research question is affirmative, meaning that the accuracy of the created predictive model is satisfactory and amounts to 79.19%, which means that the model correctly classifies patient outcomes in approximately 79 out of 100 cases. This value indicates solid model performance and its ability to identify patients at high mortality risk. However, considering this field of medicine requires high prediction accuracy, efforts should be made to improve the model's performance. Since accuracy does not consider class balance, it is necessary to analyse additional metrics to gain a more comprehensive understanding of the model's quality.

The resulting AUC value of 86.7% indicates that the model can distinguish between patients who survived and those who did not. AUC considers different cutoff values and provides a more reliable estimate of the overall performance. The answer to the second research question would be that based on the obtained AUC value, we can speak about the efficiency of the created model in predicting the risk of mortality in patients with heart failure, i.e., the good overall performance of the model has been confirmed.

## CONCLUSION

Predictive models can play a significant role in medical practice. Their primary role is to enable faster and more accurate decision-making. Predictive models can improve patient diagnostics and therapy, improving treatment outcomes.

The study's results confirm that logistic regression is a reliable method for predicting mortality due to heart failure based on clinical data. The developed predictive model shows strong overall performance, with accuracy, precision, and recall values of around 80%, while the AUC of 86.7% highlights its effectiveness in distinguishing between survival and mortality outcomes. The values of the evaluation metrics indicate that logistic regression can successfully identify key clinical attributes that influence patient outcomes, offering a data-driven approach to risk assessment. The unique value of the created model is reflected in the inclusion of a wide range of clinical attributes, which allows for a more precise analysis of risk factors and contributes to improved prediction accuracy. The created model

for predicting the death outcome due to heart failure based on the clinical characteristics of the patients can be used as an aid to health institutions in the assessment of the clinical picture of patients with cardiovascular diseases.

Given that this is a medical field, any improvement in the performance of the created model is valuable. Future research could focus on refining the model with larger datasets and additional machine learning techniques to further improve its predictive capabilities.

## REFERENCES


Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., & Holzinger, A. (2022). Deep ROC analysis and AUC as balanced average accuracy for improved classifier selection, audit, and explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(1), 329–341. https://doi.org/10.1109/TPAMI.2022.3145392

Chicco, D., & Jurman, G. (2020). Machine learning can predict the survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making, 20*, 16. https://doi.org/10.1186/s12911-020-1023-5

Dinges, H. C., Hoeft, J., Cornelius, V. M., Steinfeldt, T., Wiesmann, T., Wulf, H., & Schubert, A. K. (2023). Nominal logistic regression analysis of variables determining needle visibility in ultrasound images: A full factorial cadaver study. *BMC Anesthesiology, 23*, 369. https://doi.org/10.1186/s12871-023-02339-y

Gregorich, M., Strohmaier, S., Dunkler, D., & Heinze, G. (2021). Regression with highly correlated predictors: Variable omission is not the solution. *International Journal of Environmental Research and Public Health, 18*(8), 4259. https://doi.org/10.3390/ijerph18084259

Harris, J. K. (2021). Primer on binary logistic regression. *Family Medicine and Community Health, 9*(Suppl 1), e001290. https://doi.org/10.1136/fmch-2021-001290

Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-label confusion matrix. *IEEE Access, 10*, 19083–19095. https://doi.org/10.1109/ACCESS.2022.3151048

Jindal, H. A., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *IOP Conference Series: Materials Science and Engineering, 1022*(1), 012072. https://doi.org/10.1088/1757-899X/1022/1/012072

Khdair, H., & Dasari, N. M. (2021). Exploring machine learning techniques for coronary heart disease prediction. *International Journal of Advanced Computer Science and Applications, 12*(5), 28–36. https://doi.org/10.14569/IJACSA.2021.0120505

Koo, T. K., & Li, M. Y. (2016). A guideline for selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Krstinić, D., Braović, M., Šerić, L., & Božić-Štulić, D. (2020). Multi-label classifier performance evaluation with a confusion matrix. *Computer Science & Information Technology, 10*(8), 1–14. https://doi.org/10.5121/csit.2020.100801

Maulud, D., & Abdulazeez, A. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends, 1*(2), 140–147. https://doi.org/10.38094/jastt1457

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

MVD, A. (n.d.). *Heart failure clinical data*. Kaggle. https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data

Peng, G., Tang, Y., Cowan, T. M., Enns, G. M., Zhao, H., & Scharfe, C. (2020). Reducing false-positive results in newborn screening using machine learning. *International Journal of Neonatal Screening, 6*(1), 16. https://doi.org/10.3390/ijns6010016

Polo, T. C. F., & Miot, H. A. (2020). Use of ROC curves in clinical and experimental studies. *Jornal Vascular Brasileiro, 19*, e20200186. https://doi.org/10.1590/1677-5449.200186

Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Logistic regression. *Perspectives in Clinical Research, 8*(3), 148–151. https://doi.org/10.4103/picr.PICR_87_17

RapidMiner. (n.d.). *RapidMiner Studio*. RapidMiner Marketplace. https://marketplace.rapidminer.com/UpdateServer/faces/product_details.xhtml?productId=rapidminer-studio-6

Sahoo, P. K., & Jeripothula, P. (2020). Heart failure prediction using machine learning techniques. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3759562

Savarese, G., Becher, P. M., Lund, L. H., Seferovic, P., Rosano, G. M. C., & Coats, A. J. S. (2023). Global burden of heart failure: A comprehensive and updated review of epidemiology. *Cardiovascular Research, 118*(17), 3272–3287. https://doi.org/10.1093/cvr/cvac013

Schober, P., & Vetter, T. R. (2021). Logistic regression in medical research. *Anesthesia & Analgesia, 132*(2), 365–366. https://doi.org/10.1213/ANE.0000000000005247

Shahim, B., Kapelios, C. J., Savarese, G., & Lund, L. H. (2023). Global public health burden of heart failure:

An updated review. *Cardiac Failure Review*, *9*, e11. https://doi.org/10.15420/cfr.2023.05

Takci, H. (2018). Improvement of heart attack prediction by the feature selection methods. *Turkish Journal of Electrical Engineering & Computer Sciences*, *26*(1), 1–10. https://doi.org/10.3906/elk-1611-235

Verbakel, J. Y., Steyerberg, E. W., Uno, H., De Cock, B., Wynants, L., Collins, G. S., & Van Calster, B. (2020). ROC curves for clinical prediction models part 1: ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. *Journal of Clinical Epidemiology*, *126*, 207–216. https://doi.org/10.1016/j.jclinepi.2020.01.028

Wilson, J. R., Lorenz, K. A., & Selby, L. P. (2024). Introduction to binary logistic regression. In *Modeling Binary Correlated Responses: Using SAS, SPSS, R and STATA* (pp. 3–18). Springer. https://doi.org/10.1007/978-3-031-62427-8_1

Wurm, M. J., Rathouz, P. J., & Hanlon, B. M. (2021). Regularized ordinal regression and the ordinalNet R package. *Journal of Statistical Software*, *99*(6), 1–42. https://doi.org/10.18637/jss.v099.i06

Zaidi, A., & Al Luhayb, A. S. M. (2023). Two statistical approaches to justify the use of the logistic function in binary logistic regression. *Mathematical Problems in Engineering*, *2023*, 5525675. https://doi.org/10.1155/2023/5525675

Zapf, A., Wiessner, C., & König, I. R. (2024). Regression analyses and their particularities in observational studies. *Deutsches Ärzteblatt International*, *121*(4), 128–134. https://doi.org/10.3238/arztebl.m2023.0278

# MODEL ZA PREDVIĐANJE SMRTNOSTI OD SRČANE INSUFICIJENCIJE ZASNOVAN NA LOGISTIČKOJ REGRESIJI

**Nedavni trendovi u evaluaciji podataka sa World Wide Web-a uključuju korišćenje tradicionalnih tehnika rudarenja podataka, kao što su regresija, klasterovanje i klasifikacija. Rad ima za cilj razvijanje modela za predviđanje smrtnosti od srčane insuficijencije na osnovu javno dostupnog onlajn skupa podataka koji sadrži medicinske kartone za 299 pacijenata. Pošto ishod predviđanja može imati samo jednu od dve moguće vrednosti, primenjena je tehnika binarne logističke regresije. Istraživanje pokazuje da prediktivni model kreiran logističkom regresijom može precizno predvideti smrtnost pacijenata na osnovu njihovih kliničkih karakteristika, i identifikovati najznačajnije atribute među onima koji su uključeni u njihovu medicinsku dokumentaciju. Pored toga, primena logističke regresije obezbeđuje jednostavnost i interoperabilnost kreiranog modela, što predstavlja glavni nedostatak dosadašnjih istraživanja. Model predviđanja kreiran je upotrebom softverskog alata RapidMiner. Njegov doprinos leži u uključivanju šireg spektra kliničkih atributa, što dovodi do sveobuhvatnijeg pristupa koji poboljšava tačnost i efikasnost predviđanja. Vrednosti tačnosti, preciznosti i osetljivost razvijenog prediktivnog modela su približno 80%, što potvrđuje visok kvalitet modela. Površina ispod krive pruža grafički pregled ukupnih performansi modela i iznosi 86,7%, što odražava njegovu efikasnost. Pokazatelji razvijenog modela pokazuju snažne ukupne performanse, čime se stvara mogućnost za njegovu primenu u vidu pomoći zdravstvenim ustanovama u proceni kliničke slike pacijenata sa kardiovaskularnim bolestima.**

**Ključne reči**: Prediktivni model; Logistička regresija; Srčana insuficijencija; Kliničke karakteristike; Smrtnost pacijenata.