Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 13 [4] March 2024: 185-188
©2024 Academy for Environment and Life Sciences, India
Online ISSN 2277-1808
Journal's URL: http://www.bepls.com
CODEN: BEPLAD

ORIGINAL ARTICLE **OPEN ACCESS**

Predicting Heart Failure Risk using Binary Logistic Regression: Implications for Personalized Interventions and Patient Care

Sivaranjani P and M. Muthukumar[2]
Department of Statistics, PSG College of Arts& Science, Coimbatore, Tamilnadu, India.

ABSTRACT

Millions of people throughout the world suffer from heart failure, which is a serious health risk. Early identification of individuals at risk of heart failure is crucial for implementing preventive measures and improving patient outcomes. In this study, we employed logistic regression, a widely used statistical method for binary classification, to predict heart failure risk based on various clinical and demographic features. A comprehensive dataset comprising medical records of patients with and without heart failure was used for model training and evaluation. Our findings indicate that logistic regression demonstrates promising accuracy in predicting heart failure risk, suggesting its potential as an effective risk assessment and decision-making tool in clinical practice. Further validation and refinement of the model could pave the way for personalized interventions and improved patient care in heart failure management.

***Keywords**: Heart failure, Logistic regression, Risk prediction, Accuracy, Binary classification*

Received 21.12.2023 Revised 21.02.2024 Accepted 03.03.2024

INTRODUCTION

Heart failure continues to be a serious problem for world health, impacting a large number of people and providing difficult problems for healthcare systems everywhere. It is a complex clinical syndrome characterized by the heart's inability to pump blood efficiently, leading to various symptoms and reduced quality of life for affected individuals. Early identification of individuals at risk of heart failure is essential to implement timely interventions and preventive strategies, thereby potentially mitigating the burden of this condition [1-6]. In recent years, advancements in data science and machine learning techniques have opened up new avenues for improving risk assessment and prediction models in healthcare. Logistic regression, a well-established statistical method, has proven to be a valuable tool for binary classification tasks, making it a suitable candidate for predicting heart failure risk based on patient data. In this study, we aimed to harness the power of logistic regression to create a predictive model for heart failure risk using a diverse set of clinical and demographic features [7-10]. By utilizing a comprehensive dataset comprising medical records of patients with and without heart failure, we sought to evaluate the model's accuracy and potential for clinical application. The goal of this study is to advance our understanding of how to manage and assess the risk of heart failure [11-16]. Clinicians and healthcare professionals can create customised therapies and put preventive measures into place by identifying those who are more likely to develop heart failure. This will improve patient outcomes and boost overall healthcare efficiency.

MATERIAL AND METHODS

Information was gathered on the presence of various chronic conditions and the duration of patient observation or observation leading up to their passing. The collected data encompassed serum creatinine and serum sodium measurements recorded in MD/dL units. To interpret the findings, a logistic regression analysis was conducted and to estimate the overall survival time, the study delved into the probability of survival curves. The analysis was facilitated using JMP software, allowing for a comprehensive exploration of the data's insights and implications.

STATISTICAL ANALYSIS

Covariates such as anemia, diabetes, high blood pressure, smoking habits, and gender were included for analysis. The study investigated how these covariates impact the time until the onset of heart failure using a logistic regression model. The primary focus was on the effect of anemia, categorized as 0 for no anemia and 1 for the presence of anemia. The event of death was regarded as 1, indicating the occurrence of death, while being alive was represented as 0.

Logistic Regression model: [1]

Logistic Regression is a powerful statistical method used for modeling the relationship between one or more predictor variables (independent variables) and a binary outcome variable (dependent variable). Unlike linear regression, which predicts continuous outcomes, logistic regression is designed for predicting probabilities and making classifications. The logistic function (also known as the sigmoid function) is a key component of the logistic regression model. It maps any input value to an output value between 0 and 1, which can be interpreted as a probability. The logistic function is defined as:

$$P(y = 1|x) = \frac{1}{1+e^{-(\beta_0 + \beta_1 x)}}$$

Where

- $P(y = 1|x)$ is the probability of the dependent variable y being 1 given input x
- β_0 is the intercept term.
- β_1 is the coefficient of the input feature x
- e is the base of the natural logarithm.

The logistic sigmoid function maps the output of the linear to a value between 0 and 1, representing the probability that the input x belongs to the positive class in a binary classification problem.

RESULT AND DISCUSSION

We have used the clinical dataset for heart failure to conduct predictive analysis. This dataset is a collection of 300 instances with 13 attributes. The predictive model used here is the logistic regression.

Table 1: Description of the attributes

S.No	Attribute	Description	Permissible Values
1	age	Age	age in years
2	Ane	Anaemia	Yes, no
3	cre_pho	Creatinine phosphokinase	Iu/L
4	dieb	Diabetes	Yes, no
5	ejec_fra	Ejection fraction	EF %
6	Hbp	High blood pressure	Yes, no
7	Pl	Platelets	g/dL
8	Cr	Creatinine	mg/dL
9	Sod	Sodium	mmol/L
10	Sex	Sex	Male, female
11	smk	Smoker	Yes, no
12	Tm	Time	(0,1,2,3….)
13	Stat	Status	Death, alive

Table 2: Predictors of the outcome identified by logistic regression

Variables	Coefficient	χ^2	p	Odds ratio	95% CI
Age	0.0532	9.686	0.0003*	1.055	1.018-1.092
Gender	0.0532	1.551	0.217	0.568	0.231-1.394
Anaemia	-0.0148	0.001	0.971	0.985	0.444-2.187
Creatinine phosphokinase	0.0003	2.789	0.128	1.000	1.000-1.000
Diabetes	0.2070	0.278	0.598	1.230	0.570-2.654
Ejection fraction	0.4030	52.60	1.000	1.496	0.000-inf
High blood pressure	-0.2124	0.272	0.604	0.809	0.363-1.804
Platelets	0.007	0.139	0.711	1.000	1.000-1.000
Creatinine	0.4119	2.789	0.089	1.510	0.948-2.405
Sodium	-0.0788	2.809	0.090	0.924	0.844-2.405
Smoker	0.0324	0.005	0.942	1.033	0.428-2.491
Time	-0.0211	60.03	0.001*	0.979	0.973-0.986

Table 2 presents coefficients, chi-square statistics, p-values, and odds ratios for each covariate, accompanied by a 95% confidence interval. In logistic regression, the first category of a categorical variable automatically receives a value of 0, with the model estimating coefficients solely for the remaining categories. The effect of each covariate is captured through parameter estimates, represented as odds ratios, derived from the coefficient values. For the continuous variable 'age,' the odds ratio is 1.055 (95% CI 1.018-1.092), signifying a 1.055-fold increase in the risk of death per unit change in age, similar to other continuous variables like HBP, diabetes, and smoking. As for the categorical variable 'gender,' the odds ratio is 0.217 (95% CI 0.231-1.394), indicating that the risk of death for female patients is 0.217 times that of male patients. This pattern holds for all categorical predictor variables.

Table 3: Confusion Matrix

	Classified as alive	**Classified as not alive**
Alive	182	21
Not alive	25	71

The provided confusion matrix in Table 3 is a common tool for evaluating the performance of a classification model when applied to test data. In this case, the sum of 182 and 71 represents the correctly predicted observations, both true positives and true negatives. Conversely, the sum of 25 and 21 indicates instances where the predictions contradicted the actual outcomes, constituting false positives and false negatives. These values serve as the basis for assessing the prediction model's performance.

Table 4: Evaluation result for the prediction model

Logistic regression accuracy	84.6%
Logistic regression AUC value	92.2%
Logistic regression classification report:	
Precision	0.846
F1-Score	0.846
Support	167

The assessment of the prediction model in Table 4 reveals an accuracy of approximately 84.6%, indicating an 84% accuracy in our model's predictions. Furthermore, metrics such as precision, F1-Score, and support all demonstrated favorable outcomes for this model.

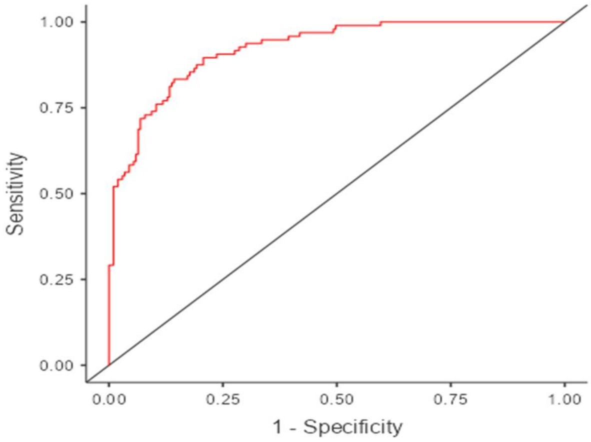

Figure 1. Roc Curve for Prediciton Model

ROC analysis was also conducted using predictions derived from the testing set. Fig 1 was created by plotting sensitivity against specificity, considering a probability threshold ranging from 0 to 1. Curves that are higher on the Y-axis and extend towards the top-right corner signify higher true positive values and lower false negative values. The area under the ROC curve was determined to be 0.92, indicating a reasonably accurate prediction performance.

CONCLUSION

This study endeavors to identify significant covariates influencing patient outcomes. Logistic regression's performance evaluation utilizes the ROC curve, yielding an area under the curve of 0.92 and a prediction accuracy of approximately 85%. Key risk factors for heightened mortality among heart failure patients encompass advancing age, elevated blood pressure (above the normal range), increased anemia levels, and reduced ejection fraction (EF) values. Conversely, elevated serum sodium levels are associated with a reduced risk of mortality. Consequently, from these assessments, logistic regression emerges as the most suitable survival regression model for predicting clinical data with binary outcomes.

CONFLICT OF INTEREST

All authors declare that no conflict of interest.

REFERENCES

1. Abbott, R.D. (1985). Logistic Regression in Survival Analysis. American Journal of Epidemiology, 121(3), 465-471.
2. Allison, P., 2001. Survival analysis using the SAS system: a practical guide. Cary, NC: SAS Publishing.
3. Anderson, J. A. and Senthilselvan, A. 1982. "A Two-step Regression Model for Hazard Functions,". Applied Statistics, 31: 44–51.
4. Bewick, V., Cheek, L., & Ball, J. (2004). Statistics Review 12: Survival Analysis. Critical Care, 8, 389-394.
5. Blagoev, K.B., Wilkerson, J., Fojo, T. (2012). Hazard Ratios in Cancer Clinical Trials – A Primer. Nat Rev Clin Oncol, 9, 178-183.
6. Cox, D. R. and Oakes, D. 1984. Analysis of Survival Data, London: Chapman & Hall.
7. Fine, J.P., & Gray, R.J. (2012). A Proportional Hazards Model for the Sub distribution of a Competing Risk. Journal of the American Statistical Association, 94(446), 496-509.
8. Goldfarb-Rumyantzev, A.S., Scandling, J.D., Pappas, L., Smout, R.J., & Horn, S. (2003). Prediction of 3-yr Cadaveric Graft Survival based on Pre-transplant Variables in a Large National Dataset. Clinical Transplantation, 17, 485-497.
9. He, C., Zhang, Y., Cai, Z., Duan, F., Lin, X., & Li, S. (2018). Nomogram to Predict Cancer-Specific Survival in Patients with Pancreatic Acinar Cell Carcinoma: A Competing Risk Analysis. Journal of Cancer, 9(22), 4117-4127.
10. Hosmer, D.W., Lemeshow, S., & May, S. (2008). Applied Survival Analysis: Regression Modeling of Time-to-Event Data, Second Edition, John Wiley & Sons, 92-131.
11. King, G. and Zeng, L., 2001. Logistic regression in rare events data. Political Analysis, 9, 137.
12. Oztekin, A., Delen, D., & Kong, Z. (2009). Predicting the Graft Survival for Heart-Lung Transplantation Patients: An Integrated Data Mining Methodology. International Journal of Medical Informatics, 78, e84-e96.
13. Prentice, R. and Gloeckler, L. 1978. "Regression Analysis of Grouped Survival Data With Application to Breast Cancer Data,". Biometrics, 34: 57–68.
14. Schober, P., & Vetter, T.R. (2018). Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. International Anesthesia Research Society, 127(3), 792-798.
15. Wannamethee, S. G., Shaper, A. G., Walker, M., & Ebrahim, S. (1998). Lifestyle and 15-year survival Free of Heart Attack, Stroke, and Diabetes in Middle-aged British Men. Archives of Internal Medicine, 158(22), 2433-2440.
16. Xiong, Z., Deng, G., Huang, X., Li, X., Xie, X., Wang, J., Shuang, Z., & Wang, X. (2018). Score for the Survival Probability in Metastasis Breast Cancer: A Nomogram- Based Risk Assessment Model. Cancer Res Treat, 50(4), 1260-1269.

CITATION OF THIS ARTICLE

Sivaranjani P, M. Muthukumar. Predicting Heart Failure Risk using Binary Logistic Regression: Implications for Personalized Interventions and Patient Care. Bull. Env.Pharmacol. Life Sci., Vol 13 [4] March 2024: 185-188