



Center for Machine Learning and Intelligent Systems

[About](#) [Citation Policy](#) [Donate a Data Set](#)
[Contact](#)

☒ Repository ☐ Web

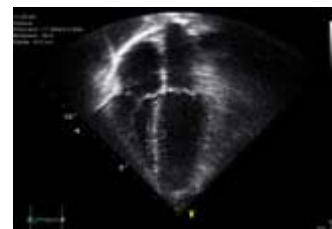


[View ALL Data Sets](#)

Echocardiogram Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Data for classifying if patients will survive for at least one year after a heart attack



Data Set Characteristics:	Multivariate	Number of Instances:	132	Area:	Life
Attribute Characteristics:	Categorical, Integer, Real	Number of Attributes:	12	Date Donated	1989-02-28
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	2451

Source:

Donor:

Steven Salzberg (salzberg '@' cs.jhu.edu)

Collector:

Dr. Evlin Kinney
The Reed Institute
P.O. Box 402603
Miami, FL 33140-0603

Data Set Information:

All the patients suffered heart attacks at some point in the past. Some are still alive and some are not. The survival and still-alive variables, when taken together, indicate whether a patient survived for at least one year following the heart attack.

The problem addressed by past researchers was to predict from the other variables whether or not the patient will survive at least one year. The most difficult part of this problem is correctly predicting that the patient will NOT survive. (Part of the difficulty seems to be the size of the data set.)

Attribute Information:

1. survival -- the number of months patient survived (has survived, if patient is still alive). Because all the patients had their heart attacks at different times, it is possible that some patients have survived less than one year but they are still alive. Check the second variable to confirm this. Such patients cannot be used for the prediction task mentioned above.
2. still-alive -- a binary variable. 0=dead at end of survival period, 1 means still alive
3. age-at-heart-attack -- age in years when heart attack occurred
4. pericardial-effusion -- binary. Pericardial effusion is fluid around the heart. 0=no fluid, 1=fluid
5. fractional-shortening -- a measure of contractility around the heart lower numbers are increasingly abnormal
6. epss -- E-point septal separation, another measure of contractility. Larger numbers are increasingly abnormal.
7. lvdd -- left ventricular end-diastolic dimension. This is a measure of the size of the heart at end-diastole. Large hearts tend to be sick hearts.
8. wall-motion-score -- a measure of how the segments of the left ventricle are moving
9. wall-motion-index -- equals wall-motion-score divided by number of segments seen. Usually 12-13 segments are seen in an echocardiogram. Use this variable INSTEAD of the wall motion score.
10. mult -- a derivate var which can be ignored
11. name -- the name of the patient (I have replaced them with "name")
12. group -- meaningless, ignore it
13. alive-at-1 -- Boolean-valued. Derived from the first two attributes. 0 means patient was either dead after 1 year or had been followed for less than 1 year. 1 means patient was alive at 1 year.

Relevant Papers:

Salzberg, S. (1988). Exemplar-based learning: Theory and implementation (Technical Report TR-10-88). Harvard University, Center for Research in Computing Technology, Aiken Computation Laboratory (33 Oxford Street; Cambridge, MA 02138).
[\[Web Link\]](#)

Kan, G., Visser, C., Kooler, J., & Dunning, A. (1986). Short and long term predictive value of wall motion score in acute myocardial infarction. British Heart Journal, 56, 422-427.

Papers That Cite This Data Set¹:



Marc Sebban and Richard Nock and Stéphane Lallich. Stopping Criterion for Boosting-Based Data Reduction Techniques: from Binary to Multiclass Problem. Journal of Machine Learning Research, 3. 2002. [\[View Context\]](#).

Gabor Melli. A Lazy Model-Based Approach to On-Line Classification. University of British Columbia. 1989. [\[View Context\]](#).

Zhi-Hua Zhou and Xu-Ying Liu. Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem. [\[View Context\]](#).

Federico Divina and Elena Marchiori. Handling Continuous Attributes in an Evolutionary Inductive Learner. Department of Computer Science Vrije Universiteit. [\[View Context\]](#).

D. Randall Wilson and Roel Martinez. Improved Center Point Selection for Probabilistic Neural Networks. Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithms. [\[View Context\]](#).

Citation Request:

Please refer to the Machine Learning Repository's [citation policy](#)

[1] Papers were automatically harvested and associated with this data set, in collaboration with [Rexa.info](#)

