|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Hepatitis Data Set**  *Download*: [Data Folder](https://archive.ics.uci.edu/ml/machine-learning-databases/hepatitis/), [Data Set Description](https://archive.ics.uci.edu/ml/machine-learning-databases/hepatitis/hepatitis.names)  **Abstract**: From G.Gong: CMU; Mostly Boolean or numeric-valued attribute types; Includes cost data (donated by Peter Turney) | https://archive.ics.uci.edu/ml/assets/MLimages/Large46.jpg |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 155 | **Area:** | Life | | **Attribute Characteristics:** | Categorical, Integer, Real | **Number of Attributes:** | 19 | **Date Donated** | 1988-11-01 | | **Associated Tasks:** | Classification | **Missing Values?** | Yes | **Number of Web Hits:** | 209076 |   **Source:**  Creator:   unknown   Donor:   G.Gong (Carnegie-Mellon University) via  Bojan Cestnik  Jozef Stefan Institute  Jamova 39  61000 Ljubljana  Yugoslavia (tel.: (38)(+61) 214-399 ext.287) }  **Data Set Information:**  Please ask Gail Gong for further information on this database.  **Attribute Information:**  1. Class: DIE, LIVE  2. AGE: 10, 20, 30, 40, 50, 60, 70, 80  3. SEX: male, female  4. STEROID: no, yes  5. ANTIVIRALS: no, yes  6. FATIGUE: no, yes  7. MALAISE: no, yes  8. ANOREXIA: no, yes  9. LIVER BIG: no, yes  10. LIVER FIRM: no, yes  11. SPLEEN PALPABLE: no, yes  12. SPIDERS: no, yes  13. ASCITES: no, yes  14. VARICES: no, yes  15. BILIRUBIN: 0.39, 0.80, 1.20, 2.00, 3.00, 4.00  -- see the note below  16. ALK PHOSPHATE: 33, 80, 120, 160, 200, 250  17. SGOT: 13, 100, 200, 300, 400, 500,  18. ALBUMIN: 2.1, 3.0, 3.8, 4.5, 5.0, 6.0  19. PROTIME: 10, 20, 30, 40, 50, 60, 70, 80, 90  20. HISTOLOGY: no, yes   The BILIRUBIN attribute appears to be continuously-valued. I checked this with the donater, Bojan Cestnik, who replied:   About the hepatitis database and BILIRUBIN problem I would like to say the following: BILIRUBIN is continuous attribute (= the number of it's "values" in the ASDOHEPA.DAT file is negative!!!); "values" are quoted because when speaking about the continuous attribute there is no such thing as all possible values. However, they represent so called "boundary" values; according to these "boundary" values the attribute can be discretized. At the same time, because of the continious attribute, one can perform some other test since the continuous information is preserved. I hope that these lines have at least roughly answered your question.  **Relevant Papers:**  Diaconis,P. & Efron,B. (1983). Computer-Intensive Methods in Statistics. Scientific American, Volume 248.  [[Web Link]](http://rexa.info/paper/0b6c9ae768733e4bef704033b5102db85008cfbc)   Cestnik,G., Konenenko,I, & Bratko,I. (1987). Assistant-86: A Knowledge-Elicitation Tool for Sophisticated Users. In I.Bratko & N.Lavrac (Eds.) Progress in Machine Learning, 31-45, Sigma Press.  [[Web Link]](http://rexa.info/paper/929f34b1fbef4c2335cebebf22840bcf78577ae4)  **Papers That Cite This Data Set1:**  https://archive.ics.uci.edu/ml/assets/rexa.jpg  Amaury Habrard and Marc Bernard and Marc Sebban. [IOS Press Detecting Irrelevant Subtrees to Improve Probabilistic Learning from Tree-structured Data](http://rexa.info/paper/61172cddb9dbfd4d26caeeaa3a8a8deb58a26bb5). Fundamenta Informaticae. 2004. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#61172cddb9dbfd4d26caeeaa3a8a8deb58a26bb5)].  Zhi-Hua Zhou and Yuan Jiang and Shifu Chen. [Extracting symbolic rules from trained neural network ensembles](http://rexa.info/paper/faf03ce4427609aa6db25e3d0e6479fb2ae153a7). AI Commun, 16. 2003. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#faf03ce4427609aa6db25e3d0e6479fb2ae153a7)].  Xiaoli Z. Fern and Carla Brodley. [Boosting Lazy Decision Trees](http://rexa.info/paper/d891d25459c4422e9cf2262272938d1ada438a9e). ICML. 2003. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#d891d25459c4422e9cf2262272938d1ada438a9e)].  Jinyan Li and Limsoon Wong. [Using Rules to Analyse Bio-medical Data: A Comparison between C4.5 and PCL](http://rexa.info/paper/4f29b49d0d83edf9e0a21d102467df5ffdf64fcb). WAIM. 2003. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#4f29b49d0d83edf9e0a21d102467df5ffdf64fcb)].  Michael L. Raymer and Travis E. Doom and Leslie A. Kuhn and William F. Punch. [Knowledge discovery in medical and biological datasets using a hybrid Bayes classifier/evolutionary algorithm](http://rexa.info/paper/076c5b4e7bf69148bcde8c83621ecf1140a7d55c). IEEE Transactions on Systems, Man, and Cybernetics, Part B, 33. 2003. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#076c5b4e7bf69148bcde8c83621ecf1140a7d55c)].  Takashi Matsuda and Hiroshi Motoda and Tetsuya Yoshida and Takashi Washio. [Mining Patterns from Structured Data by Beam-Wise Graph-Based Induction](http://rexa.info/paper/1c7b9b0abcac674ef129893e65f67ae6458161fb). Discovery Science. 2002. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#1c7b9b0abcac674ef129893e65f67ae6458161fb)].  Wl/odzisl/aw Duch and Karol Grudzinski. [Ensembles of Similarity-based Models](http://rexa.info/paper/7372c1c34417795a4d0752a07c77b3595dba16c4). Intelligent Information Systems. 2001. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#7372c1c34417795a4d0752a07c77b3595dba16c4)].  Petri Kontkanen and Petri Myllym and Tomi Silander and Henry Tirri and Peter Gr. [On predictive distributions and Bayesian networks](http://rexa.info/paper/084aa2a0b1ffd67537222e2c439d2fadce6090ca). Department of Computer Science, Stanford University. 2000. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#084aa2a0b1ffd67537222e2c439d2fadce6090ca)].  Gary M. Weiss and Haym Hirsh. [A Quantitative Study of Small Disjuncts: Experiments and Results](http://rexa.info/paper/55a7f45a6b42ceb9097c7260032694bcbd8a2fd2). Department of Computer Science Rutgers University. 2000. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#55a7f45a6b42ceb9097c7260032694bcbd8a2fd2)].  David W. Opitz and Richard Maclin. [Popular Ensemble Methods: An Empirical Study](http://rexa.info/paper/2d05e6777bcfde449fc35a3a95dde0697a5c49ac). J. Artif. Intell. Res. (JAIR, 11. 1999. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#2d05e6777bcfde449fc35a3a95dde0697a5c49ac)].  Yk Huhtala and Juha Kärkkäinen and Pasi Porkka and Hannu Toivonen. [Efficient Discovery of Functional and Approximate Dependencies Using Partitions](http://rexa.info/paper/ac8fe867e1d16d4d09f9bd759ba46699055c7ca6). ICDE. 1998. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#ac8fe867e1d16d4d09f9bd759ba46699055c7ca6)].  Floriana Esposito and Donato Malerba and Giovanni Semeraro. [A Comparative Analysis of Methods for Pruning Decision Trees](http://rexa.info/paper/8cf017f4c8d633c1694c2ea3fa2a165e0d8d35f1). IEEE Trans. Pattern Anal. Mach. Intell, 19. 1997. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#8cf017f4c8d633c1694c2ea3fa2a165e0d8d35f1)].  . [Prototype Selection for Composite Nearest Neighbor Classifiers](http://rexa.info/paper/b0009a0081cc5fbfbae758def55cfd5b3256623b). Department of Computer Science University of Massachusetts. 1997. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#b0009a0081cc5fbfbae758def55cfd5b3256623b)].  Ron Kohavi. [The Power of Decision Tables](http://rexa.info/paper/dd3f32548422fd3db3846c2fba689a4406d9cf0c). ECML. 1995. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#dd3f32548422fd3db3846c2fba689a4406d9cf0c)].  Peter D. Turney. [Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm](http://rexa.info/paper/f6f302674e1188614eaa6f23c782d8d1c0ea2320). CoRR, csAI/9503102. 1995. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#f6f302674e1188614eaa6f23c782d8d1c0ea2320)].  Christophe Giraud and Tony Martinez and Christophe G. Giraud-Carrier. [University of Bristol Department of Computer Science ILA: Combining Inductive Learning with Prior Knowledge and Reasoning](http://rexa.info/paper/01ede1d6464c9533cd6ec1a4492d134729ce99bd). 1995. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#01ede1d6464c9533cd6ec1a4492d134729ce99bd)].  Gabor Melli. [A Lazy Model-Based Approach to On-Line Classification](http://rexa.info/paper/287f8092a743949a6e0151893b9e3bc4d03466ed). University of British Columbia. 1989. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#287f8092a743949a6e0151893b9e3bc4d03466ed)].  Rafael S. Parpinelli and Heitor S. Lopes and Alex Alves Freitas. [An Ant Colony Based System for Data Mining: Applications to Medical Data](http://rexa.info/paper/899bdb470e48c308144216cc22048c88816ee035). CEFET-PR, CPGEI Av. Sete de Setembro, 3165. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#899bdb470e48c308144216cc22048c88816ee035)].  Suresh K. Choubey and Jitender S. Deogun and Vijay V. Raghavan and Hayri Sever. [A comparison of feature selection algorithms in the context of rough classifiers](http://rexa.info/paper/341c37581773a2a38cecc290e6272ca752b9df84). [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#341c37581773a2a38cecc290e6272ca752b9df84)].  Takao Mohri and Hidehiko Tanaka. [An Optimal Weighting Criterion of Case Indexing for Both Numeric and Symbolic Attributes](http://rexa.info/paper/177158d5e743e0f71b640c20a162ee8af5c79dcf). Information Engineering Course, Faculty of Engineering The University of Tokyo. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#177158d5e743e0f71b640c20a162ee8af5c79dcf)].  Wl/odzisl/aw Duch and Rafal/ Adamczak Email:duchraad@phys. uni. torun. pl. [Statistical methods for construction of neural networks](http://rexa.info/paper/53ac23f963b3607aae9580b356e6b236d2955314). Department of Computer Methods, Nicholas Copernicus University. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#53ac23f963b3607aae9580b356e6b236d2955314)].  Chris Drummond and Robert C. Holte. [C4.5, Class Imbalance, and Cost Sensitivity: Why Under-Sampling beats Over-Sampling](http://rexa.info/paper/de9b2ac4ef9c6449914cdc375972f722190fee7a). Institute for Information Technology, National Research Council Canada. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#de9b2ac4ef9c6449914cdc375972f722190fee7a)].  Alexander K. Seewald. [Dissertation Towards Understanding Stacking Studies of a General Ensemble Learning Scheme ausgefuhrt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Naturwissenschaften](http://rexa.info/paper/e2b2b723df700c90e69a31a4403b740c2d2a7b2f). [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#e2b2b723df700c90e69a31a4403b740c2d2a7b2f)].  Ida G. Sprinkhuizen-Kuyper and Elena Smirnova and I. Nalbantis. [Reliability yields Information Gain](http://rexa.info/paper/fac79a1ce7fa29979fe50466d471cdac753f26d5). IKAT, Universiteit Maastricht. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#fac79a1ce7fa29979fe50466d471cdac753f26d5)].  Christophe Giraud and Tony Martinez. [ADYNAMIC INCREMENTAL NETWORK THAT LEARNS BY DISCRIMINATION](http://rexa.info/paper/ecfe10b847878abd1208e57bbbc4040fb87bebbc). AA. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#ecfe10b847878abd1208e57bbbc4040fb87bebbc)].  Federico Divina and Elena Marchiori. [Handling Continuous Attributes in an Evolutionary Inductive Learner](http://rexa.info/paper/954de642cab661d060a2dbc68d3023ba3a9763e1). Department of Computer Science Vrije Universiteit. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#954de642cab661d060a2dbc68d3023ba3a9763e1)].  Zhi-Hua Zhou and Xu-Ying Liu. [Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem](http://rexa.info/paper/c042581c25e66281bb5ce382f70738b0233e5f5a). [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#c042581c25e66281bb5ce382f70738b0233e5f5a)].  Rafael S. Parpinelli and Heitor S. Lopes and Alex Alves Freitas. [PART FOUR: ANT COLONY OPTIMIZATION AND IMMUNE SYSTEMS Chapter X An Ant Colony Algorithm for Classification Rule Discovery](http://rexa.info/paper/f4405e32dbb5dea3ece303e2a5b3edb6b413271e). CEFET-PR, Curitiba. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#f4405e32dbb5dea3ece303e2a5b3edb6b413271e)].  Wl/odzisl/aw Duch and Rafal Adamczak and Geerd H. F Diercksen. [Neural Networks from Similarity Based Perspective](http://rexa.info/paper/118fe8f48c22375803e81b47d91c0bfd96c19999). Department of Computer Methods, Nicholas Copernicus University. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#118fe8f48c22375803e81b47d91c0bfd96c19999)].  Wl/odzisl/aw Duch and Karol Grudzinski and Geerd H. F Diercksen. [Minimal distance neural methods](http://rexa.info/paper/0f1456aba1cd88fe698ef134fa761a3ee2fc03f2). Department of Computer Methods, Nicholas Copernicus University. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#0f1456aba1cd88fe698ef134fa761a3ee2fc03f2)].  Wl odzisl and Rafal Adamczak and Krzysztof Grabczewski. [Optimization of Logical Rules Derived by Neural Procedures](http://rexa.info/paper/15b7c2d67edb6f722c80cf48928bc53c17ad7bef). Department of Computer Methods, Nicholas Copernicus University. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#15b7c2d67edb6f722c80cf48928bc53c17ad7bef)].  Wl/odzisl/aw Duch and Rafal Adamczak and Geerd H. F Diercksen. [Classification, Association and Pattern Completion using Neural Similarity Based Methods](http://rexa.info/paper/4967c873f90995ec2e7fd690a0ec6118f1adc807). Department of Computer Methods, Nicholas Copernicus University. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#4967c873f90995ec2e7fd690a0ec6118f1adc807)].  Elena Smirnova and Ida G. Sprinkhuizen-Kuyper and I. Nalbantis and b. ERIM and Universiteit Rotterdam. [Unanimous Voting using Support Vector Machines](http://rexa.info/paper/e5d994d772cfe5ec4d0f3e6d669f0bc28180a3ae). IKAT, Universiteit Maastricht. [[View Context](https://archive.ics.uci.edu/ml/support/Hepatitis#e5d994d772cfe5ec4d0f3e6d669f0bc28180a3ae)].  **Citation Request:**  Please refer to the Machine Learning Repository's [citation policy](https://archive.ics.uci.edu/ml/citation_policy.html) |