

## 1. Title: Hepatitis Domain

## 2. Sources:

- (a) unknown
- (b) Donor: G.Gong (Carnegie-Mellon University) via  
Bojan Cestnik  
Jozef Stefan Institute  
Jamova 39  
61000 Ljubljana  
Yugoslavia (tel.: (38)(+61) 214-399 ext.287) }
- (c) Date: November, 1988

## 3. Past Usage:

- 1. Diaconis,P. & Efron,B. (1983). Computer-Intensive Methods in Statistics. Scientific American, Volume 248.  
-- Gail Gong reported a 80% classification accuracy
- 2. Cestnik,G., Kononenko,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.  
-- Assistant-86: 83% accuracy

## 4. Relevant Information:

Please ask Gail Gong for further information on this database.

## 5. Number of Instances: 155

## 6. Number of Attributes: 20 (including the class attribute)

## 7. 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 continuous 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.

## 8. Missing Attribute Values: (indicated by "?")

Attribute Number:	Number of Missing Values:
1:	0
2:	0
3:	0
4:	1
5:	0
6:	1
7:	1
8:	1
9:	10
10:	11
11:	5
12:	5
13:	5
14:	5
15:	6
16:	29
17:	4
18:	16
19:	67
20:	0

## 9. Class Distribution:

DIE: 32  
LIVE: 123