HCV data
Donated on 6/9/2020

The data set contains laboratory values of blood donors and Hepatitis C patients and demographic values like age.

Dataset Characteristics
Multivariate

Subject Area
Life Science

Associated Tasks
Classification, Clustering

Feature Type
Integer, Real

 # Instances
615

 # Features
11

Dataset Information

What do the instances in this dataset represent?
Instances are patients

Additional Information
The target attribute for classification is Category (blood donors vs. Hepatitis C, including its progress: 'just' Hepatitis C, Fibrosis, Cirrhosis).

Has Missing Values?
Yes

Introductory Paper

Using machine learning techniques to generate laboratory diagnostic pathways—a case study
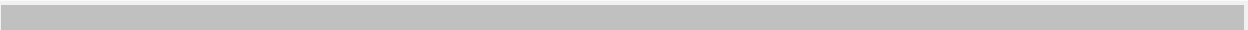
By Georg F. Hoffmann, A. Bietenbeck, R. Lichtinghagen, F. Klawonn. 2018
Published in Journal of Laboratory and Precision Medicine

Variables Table

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
ID	ID	Integer		Patient ID		no
Age	Feature	Integer	Age		years	no
Sex	Feature	Binary	Sex			no
ALB	Feature	Continuous				yes
ALP	Feature	Continuous				yes
AST	Feature	Continuous				yes
BIL	Feature	Continuous				no
CHE	Feature	Continuous				no
CHOL	Feature	Continuous				yes
CREA	Feature	Continuous				no
CGT	Feature	Continuous				no
PROT	Feature	Continuous				yes
Category	Target	Categorical		values: '0=Blood Donor', '0s=suspect Blood Donor', '1=Hepatitis', '2=Fibrosis', '3=Cirrhosis'		no

Rows per page 15 ▾   0 to 13 of 13   ‹ ›

Additional Variable Information ⌃

All attributes except Category and Sex are numerical. The laboratory data are the attributes 5-14.
    1) X (Patient ID/No.)
    2) Category (diagnosis) (values: '0=Blood Donor', '0s=suspect Blood Donor', '1=Hepatitis', ...

SHOW MORE ⌄

Creators

Ralf Lichtinghagen
👤 lichtinghagen.ralf@mh-hannover.de
Institute of Clinical Chemistry; Medical University Hannover (MHH)

Frank Klawonn
👤 frank.klawonn@helmholtz-hzi.de
Helmholtz Centre for Infection Research

Georg Hoffmann
👤 georg.hoffmann@trillium.de
Trillium GmbH

DOI

10.24432/C5D612

License

This dataset is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given.