Breast Cancer Wisconsin (Diagnostic)

Donated on 10/31/1995

Diagnostic Wisconsin Breast Cancer Database.

Dataset Characteristics
Multivariate

Subject Area
Life Science

Associated Tasks
Classification

Feature Type
Real

 # Instances
569

 # Features
30

Dataset Information

Additional Information
Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. A few of the images can be found at http://www.cs.wisc.edu/~street/images/...

SHOW MORE

Has Missing Values?
No

Introductory Paper

Nuclear feature extraction for breast tumor diagnosis
By W. Street, W. Wolberg, O. Mangasarian. 1993
Published in Electronic imaging

Variables Table

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
symmetry3	Feature	Continuous				false
fractal_dimension3	Feature	Continuous				false

Rows per page 10 30 to 32 of 32

Additional Variable Information

Additional Information
1) ID number
2) Diagnosis (M = malignant, B = benign)
3-32)

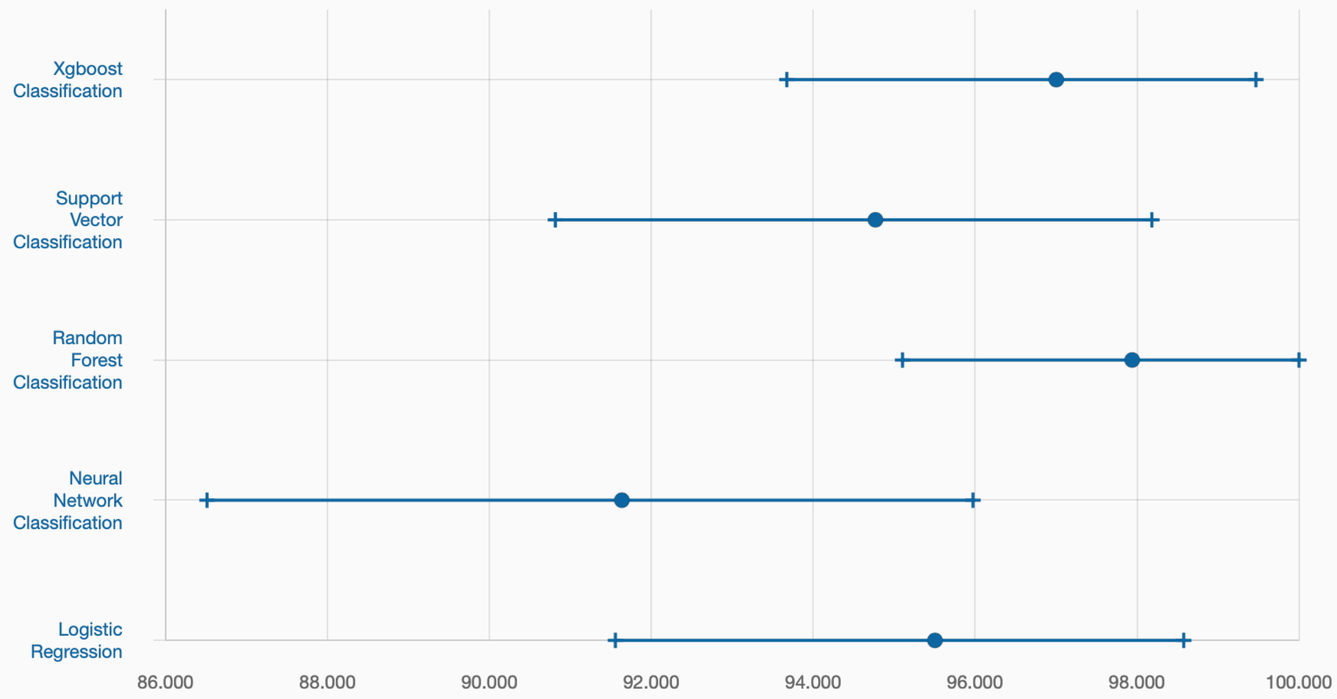
Ten real-valued features are computed for each cell nucleus:

a) radius (mean of distances from center to points on the perimeter)
b) texture (standard deviation of gray-scale values)
c) perimeter
d) area
e) smoothness (local variation in radius lengths)
f) compactness (perimeter^2 / area - 1.0)
g) concavity (severity of concave portions of the contour)
h) concave points (number of concave portions of the contour)
i) symmetry
j) fractal dimension ("coastline approximation" - 1)

SHOW LESS

Baseline Model Performance

Accuracy Precision



Papers Citing this Dataset

SORT BY YEAR, DESC

Machine learning in medicine: a practical introduction
By Jenni Sidey-Gibbons, Chris Sidey-Gibbons. 2019
Published in BMC medical research methodology.

Data Augmentation Using GANs
By Fabio Tanaka, Claus Aranha. 2019
Published in ArXiv.

Empirical study towards understanding line search approximations for training neural networks
By Younghwan Chae, Daniel Wilke. 2019
Published in

Interpretable Counterfactual Explanations Guided by Prototypes
By Arnaud Looveren, Janis Klaise. 2019
Published in ArXiv.

PIDT: A Novel Decision Tree Algorithm Based on Parameterised Impurities and Statistical Pruning Approaches
By Daniel Stamate, Wajdi Alghamdi, Daniel Stahl, Doina Logofatu, Alexander Zamyatin. 2018
Published in AIAI.

Rows per page 5 0 to 5 of 37

DOWNLOAD

CITE

37 citations
138157 views

Keywords

health cancer

Creators

William Wolberg
Olvi Mangasarian
Nick Street
W. Street

DOI

10.24432/C5DW2B

License