


☒ Repository ☐ Web [View ALL Data Sets](#)

Optical Recognition of Handwritten Digits Data Set

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

Abstract: Two versions of this database available; see folder

Data Set Characteristics:	Multivariate	Number of Instances:	5620	Area:	Computer
Attribute Characteristics:	Integer	Number of Attributes:	64	Date Donated	1998-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	112857

Source:

E. Alpaydin, C. Kaynak
Department of Computer Engineering
Bogazici University, 80815 Istanbul Turkey
alpaydin '@' boun.edu.tr

Data Set Information:

We used preprocessing programs made available by NIST to extract normalized bitmaps of handwritten digits from a preprinted form. From a total of 43 people, 30 contributed to the training set and different 13 to the test set. 32x32 bitmaps are divided into nonoverlapping blocks of 4x4 and the number of on pixels are counted in each block. This generates an input matrix of 8x8 where each element is an integer in the range 0..16. This reduces dimensionality and gives invariance to small distortions.

For info on NIST preprocessing routines, see M. D. Garris, J. L. Blue, G. T. Candela, D. L. Dimmick, J. Geist, P. J. Grother, S. A. Janet, and C. L. Wilson, NIST Form-Based Handprint Recognition System, NISTIR 5469, 1994.

Attribute Information:

All input attributes are integers in the range 0..16.
The last attribute is the class code 0..9

Relevant Papers:

C. Kaynak (1995) Methods of Combining Multiple Classifiers and Their Applications to Handwritten Digit Recognition, MSc Thesis, Institute of Graduate Studies in Science and Engineering, Bogazici University.

[\[Web Link\]](#)

E. Alpaydin, C. Kaynak (1998) Cascading Classifiers, Kybernetika. [\[Web Link\]](#)

[\[Web Link\]](#)

Papers That Cite This Data Set¹:

Ken Tang and Ponnuthurai N. Suganthan and Xi Yao and A. Kai Qin. [Linear dimensionality reduction using relevance weighted LDA](#). School of Electrical and Electronic Engineering Nanyang Technological University. 2005. [\[View Context\]](#).

Claudio Gentile. [A New Approximate Maximal Margin Classification Algorithm](#). NIPS. 2000. [\[View Context\]](#).

Stephen D. Bay. [Nearest neighbor classification from multiple feature subsets](#). Intell. Data Anal, 3. 1999. [\[View Context\]](#).

Ethem Alpaydin. [Combined 5 x 2 cv F Test for Comparing Supervised Classification Learning Algorithms](#). Neural Computation, 11. 1999. [\[View Context\]](#).

Erick Cantú-Paz and Chandrika Kamath. [Using Evolutionary Algorithms to Induce Oblique Decision Trees](#). Center for Applied Scientific Computing Lawrence Livermore National Laboratory. [\[View Context\]](#).

Ayhan Demiriz and Kristin P. Bennett and John Shawe and I. Nourtdinov V.. [Linear Programming Boosting via Column Generation](#). Dept. of Decision Sciences and Eng. Systems, Rensselaer Polytechnic Institute. [\[View Context\]](#).

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](#)

Supported By:



In Collaboration With:

[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)