The **MNIST database** (*Modified*[*National Institute of Standards and Technology*](https://en.wikipedia.org/wiki/National_Institute_of_Standards_and_Technology)*database*[[1]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-1)) is a large [database](https://en.wikipedia.org/wiki/Database) of handwritten digits that is commonly used for [training](https://en.wikipedia.org/wiki/Training_set) various [image processing](https://en.wikipedia.org/wiki/Image_processing) systems.[[2]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-2)[[3]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-3) The database is also widely used for training and testing in the field of [machine learning](https://en.wikipedia.org/wiki/Machine_learning).[[4]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-4)[[5]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-5) It was created by "re-mixing" the samples from NIST's original datasets.[[6]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-6) The creators felt that since NIST's training dataset was taken from American [Census Bureau](https://en.wikipedia.org/wiki/United_States_Census_Bureau) employees, while the testing dataset was taken from [American](https://en.wikipedia.org/wiki/Americans) [high school](https://en.wikipedia.org/wiki/High_school) students, it was not well-suited for machine learning experiments.[[7]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-LeCun-7) Furthermore, the black and white images from NIST were [normalized](https://en.wikipedia.org/wiki/Normalization_(image_processing)) to fit into a 28x28 pixel bounding box and [anti-aliased](https://en.wikipedia.org/wiki/Spatial_anti-aliasing), which introduced grayscale levels.[[7]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-LeCun-7)

The MNIST database contains 60,000 training images and 10,000 testing images.[[8]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-8) Half of the training set and half of the test set were taken from NIST's training dataset, while the other half of the training set and the other half of the test set were taken from NIST's testing dataset.[[9]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-9) The original creators of the database keep a list of some of the methods tested on it.[[7]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-LeCun-7) In their original paper, they use a [support-vector machine](https://en.wikipedia.org/wiki/Support-vector_machine) to get an error rate of 0.8%.[[10]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-Gradient-10)

Extended MNIST (EMNIST) is a newer dataset developed and released by NIST to be the (final) successor to MNIST.[[11]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-11)[[12]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-12) MNIST included images only of handwritten digits. EMNIST includes all the images from NIST Special Database 19, which is a large database of handwritten uppercase and lower case letters as well as digits.[[13]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-13)[[14]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-14) The images in EMNIST were converted into the same 28x28 pixel format, by the same process, as were the MNIST images. Accordingly, tools which work with the older, smaller, MNIST dataset will likely work unmodified with EMNIST.

The original MNIST dataset contains at least 4 wrong labels.[[15]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-15)

**History**

[[edit](https://en.wikipedia.org/w/index.php?title=MNIST_database&action=edit&section=1)]

The set of images in the MNIST database was created in 1994 as a combination of two of [NIST](https://en.wikipedia.org/wiki/NIST)'s databases: Special Database 1; and Special Database 3.[[16]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-bottou1994-16)

Special Database 1 and Special Database 3 consist of digits written by high school students and employees of the [United States Census Bureau](https://en.wikipedia.org/wiki/United_States_Census_Bureau), respectively.[[7]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-LeCun-7)

The original dataset was a set of 128x128 binary images, processed into 28x28 grayscale images. The training set and the testing set both originally had 60k samples, but 50k of the testing set samples were discarded.[[17]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-yadav2019-17)

[Fashion MNIST](https://en.wikipedia.org/wiki/Fashion_MNIST) was created in 2017 as a more challenging replacement for MNIST. The dataset consists of 70,000 28x28 grayscale images of fashion products from 10 categories.[[18]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-18)

**Performance**

[[edit](https://en.wikipedia.org/w/index.php?title=MNIST_database&action=edit&section=2)]

Some researchers have achieved "near-human performance" on the MNIST database, using a committee of [neural networks](https://en.wikipedia.org/wiki/Neural_networks); in the same paper, the authors achieve performance double that of humans on other recognition tasks.[[19]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-Multideep-19) The highest error rate listed[[7]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-LeCun-7) on the original website of the database is 12 percent, which is achieved using a simple [linear classifier](https://en.wikipedia.org/wiki/Linear_classifier) with no preprocessing.[[10]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-Gradient-10)

In 2004, a best-case error rate of 0.42 percent was achieved on the database by researchers using a new classifier called the LIRA, which is a neural classifier with three neuron layers based on [Rosenblatt's perceptron](https://en.wikipedia.org/wiki/Perceptron) principles.[[20]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-20)

Some researchers have tested [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) systems using the database put under random distortions. The systems in these cases are usually neural networks and the distortions used tend to be either [affine distortions](https://en.wikipedia.org/wiki/Affine_transformation) or [elastic distortions](https://en.wikipedia.org/wiki/Elastic_deformation).[[7]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-LeCun-7) Sometimes, these systems can be very successful; one such system achieved an error rate on the database of 0.39 percent.[[21]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-21)

In 2011, an error rate of 0.27 percent, improving on the previous best result, was reported by researchers using a similar system of neural networks.[[22]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-22) In 2013, an approach based on [regularization](https://en.wikipedia.org/wiki/Regularization_(mathematics)) of neural networks using [DropConnect](https://en.wikipedia.org/wiki/DropConnect" \o "DropConnect) has been claimed to achieve a 0.21 percent error rate.[[23]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-23) In 2016, the single [convolutional neural network](https://en.wikipedia.org/wiki/Convolutional_neural_network) best performance was 0.25 percent error rate.[[24]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-:0-24) As of August 2018, the best performance of a single convolutional neural network trained on MNIST training data using no [data augmentation](https://en.wikipedia.org/wiki/Data_augmentation) is 0.25 percent error rate.[[24]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-:0-24)[[25]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-25) Also, the Parallel Computing Center (Khmelnytskyi, Ukraine) obtained an ensemble of only 5 convolutional neural networks which performs on MNIST at 0.21 percent error rate.[[26]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-Romanuke3-26)[[27]](https://en.wikipedia.org/wiki/MNIST_database#cite_note-Romanuke4-27)