

## Supplemental Digital Content

The Supplemental Digital Content was not copyedited by Archives of Pathology & Laboratory Medicine.

### Machine learning glossary

Term	Definition
Adversarial attack	Purposely creating instances that resemble members of a class but that a machine learning model will classify incorrectly. In image classification, such images are called “adversarial images.”
Annotation	Manually adding information or interpretation to an image or other instance; may include marking regions, identifying image elements, or defining diagnoses.
Augmented intelligence	Use of artificial intelligence tools to improve expert human performance rather than replace experts.
Backpropagation	A method for training neural networks in which input and transfer function weights are adjusted in waves propagating backward through the layers of a neural network, from the output toward the input.
Bagging	Short for “bootstrap aggregating;” producing diversity in an ensemble of models by training each model against a different random subset of the training data.
Bias	A systematic inaccuracy in classification or regression, suggesting either that the model does not accurately represent the underlying data relationships or that there is an unanticipated difference between training and operational data.
Black box problem	Lack of operational transparency and explainability in machine learning models. Refers to the difficulty in relating patterns of model parameters, and weights to relevant patterns in training data and to model output, especially for models with many parameters.
Boosting	Weighting classification errors in sequential ensembles of weak models, such that later models in the chain focus on and correct the errors of the earlier models.
Bootstrapping	Short for “bootstrap resampling.” A technique for creating multiple preliminary training data sets by random sampling with replacement from the main training data set. Model performance is reported as the mean and standard deviation across the preliminary sets.
Calibration	In classification, the ability to assign class membership probability accurately across a range of probabilities (see Fig. 6).
Class	A name for similar instances, i.e., type of instance. May be positive/negative for a diagnosis or prognosis, or one of a set of diagnoses.
Classification	Distinguishing instances by class. Binary classification distinguishes between two classes. Multi-class classification determines a single class from a set of more than two classes. Multi-label classification allows more than one class label from a set of classes (non-exclusive class membership).
Class imbalance	A difference in incidence between the classes in a data set.
Cleansing data	The overall process of preparing data for use in machine learning. Includes aggregation, dealing with missing values and errors, standardization including mapping different representations, normalization, and formatting.

Cross-validation	Dividing training data randomly into multiple segments (folds) where one fold is held out as test data and the rest are used as training data. Preliminary training and testing is carried out repetitively with each fold being used for testing once. Model performance is reported as the mean and standard deviation of the individual fold performances. Stratified cross validation holds the class incidence constant across the folds.
Deep learning	Machine learning using deep neural networks, which have multiple hidden layers (see Fig. 4).
Dimensions	Also called "features." Dimensions are the data elements that are associated with instances. Instances with only two or three dimensions can be plotted using the dimension values, to show clusters of similar instances.
Discrimination	The ability to distinguish between classes.
Distance metric	A method for calculating the degree of difference between two instances from values of their dimensions.
Drift	A gradual change in the distribution of values in a dimension/feature, or the incidence of classes. May reduce model performance.
Epoch	One pass through all the training data during model training. Training may require many epochs.
Feature	Also called a "dimension." Features are the data elements associated with each instance
Feature map	A compact representation of an image, produced by sliding a "mask" or window across an image and extracting a value (e.g., maximum or average) for each location. Feature maps are produced in the pooling layers of convolutional neural networks (see Fig. 4).
Feature selection	The art and science of choosing instance features most likely to be useful in a machine learning or analytics task. Most instances have many features and correctly selecting the ones most contributory to the task at hand speeds up machine learning and improves its accuracy.
FROC	Free response receiver operating characteristic curve, plots sensitivity vs. number of false positives per instance. Used in image classification (see Fig. 6).
Generalizability	The ability of a model to be used outside of its specific development environment, or for tasks that are similar to but not identical to its trained purpose.
Generative adversarial network (GAN)	A neural network designed to create or modify images that elicit specific responses from other machine learning models. Some GANs create images that purposely cause classification errors.
Gradient descent	An optimization method that drives a cost function progressively to a minimum value.
Hyperparameter	A model parameter whose value controls or defines the training process and is set by model developers rather than learned.
Instance	A single example of the item being classified, for example a slide or patient. Instances typically have multiple features or dimensions that contain the data used to train and test the model, and by which they are classified. Training and test data are populations of instances.

Instance-based method	A machine learning method that does not use a model, but instead incorporates training instances into a map that is used directly to classify new instances. See k-Nearest Neighbors for an example (Fig. 4).
Label	Class identifier.
Loss function	A function that calculates the degree of difference between a model's output and the true answer during supervised learning. This calculation determines the amount of adjustment that is made in the model's learned parameters in response to the error. Objective function and cost function are related terms.
Model	A function or algorithm that is in training or has been trained and includes learned weights and parameters specific to a particular training data set.
Multiple instance learning	Labeling instances with relatively broad categories and allowing algorithms to learn detailed features associated with those categories automatically, rather than extensively annotating detail (e.g., in images). Can reduce the work of annotation at the cost of requiring large numbers of instances (i.e., large image libraries).
Overfit	A model that does not accurately represent the underlying relationships between the data elements because it incorporates too much specific detail from the training set; an overfit model performs well against the training data but poorly against the held-out test data.
Oversampling	Increasing the representation of a class in the training data by randomly duplicating class members or creating additional synthetic but plausible class members.
Receiver operating characteristic (ROC) curve	A plot of sensitivity (recall) vs. false positive rate for a machine learning model or other classifier. Used to evaluate discrimination (see Fig. 6).
Regression	A machine learning model that returns a continuous value.
Regularization	A group of techniques that penalize increasing model complexity to avoid overfitting.
Reinforcement learning	Awards points for progress toward a goal and pursues maximization of points, rather than a defined answer or complete solution. Particularly useful for optimizing action sequences in domains such as game playing and robotics.
Saliency map	A visual representation highlighting the portions of an image important for defining a class or distinguishing between classes.
SaMD	Software as a medical device; software used on its own rather than being a supporting part of another device.
Semi-supervised learning	A supervised learning method that trains a model with a small annotated data set initially, uses that model to label additional unlabeled data, and then re-trains on the combined larger data set.
Shift	A relatively sudden change in the distribution of values of a dimension/feature (also called "covariate shift"), or in the incidence of classes. May reduce model performance.
Stability (of models)	The ability of a model to produce similar output over a range of similar inputs, including inputs not seen previously but not substantially different from those that were seen. Stable models are referred to as "robust" and unstable models that are likely to produce unanticipated output are "brittle."

Supervised learning	A machine learning strategy that uses pre-labeled instances to train a model that can then predict labels for unlabeled instances.
Test data	Data randomly selected from a correctly labeled developmental data set that is sequestered prior to training to provide an unbiased estimate of model performance after training.
Training	Optimizing a machine learning model by using it to process known instances and adjusting weights on input data and model parameters to reduce errors in output.
Training data	Data randomly selected from a correctly labeled developmental data set that is used for training a machine learning model.
Transfer function	A mathematical function, often simple, that processes the inputs to a neural network node to create the output.
Transfer learning	Training a model initially using a large, generic data set and then finishing training with a smaller, domain-specific data set. Frequently used in image classification.
Underfit	A model that does not accurately represent the underlying relationships in the data elements because it has been incompletely trained; an underfit model performs poorly with both the training and test data sets.
Undersampling	Reducing the representation of a class in the training data by including only a random subset of the class.
Unsupervised learning	A machine learning strategy that uses unlabeled instances and identifies data patterns or instance groups solely based on the characteristics of the data. Often used to investigate the types of structure present in data.
Validation	Optimizing model settings and reliability during training using a subset of training data that is temporarily held out for evaluation purposes (e.g., cross-validation).