Construction of the model

Random forest

Random forest [45] (RF) is parallel integrated learning based on decision trees. It is trained by randomly selecting samples from the original data through a sampling technique with put-back. Finally, the results are filtered by voting [46].

Naive Bayesian classification

Naive Bayesian classification (NBC) is a simple method for constructing classifiers. Bayes' theorem is to assume that X is a data tuple, usually described by the measurements of a set of n attributes; H denotes the case of some hypotheses; $P(H|X)$ is the posterior probability, which indicates the probability that hypothesis H holds under some condition X; $P(H)$ is the prior probability of hypothesis H, i.e., the probability that hypothesis H holds; $P(X)$ is the probability that hypothesis X holds; $P(X|H)$ is the class conditional probability, in the case of H, X probability of holding. Bayes' theorem is:

$$P(X|H) = \frac{P(X|H)P(H)}{P(X)}$$

Since the screened key genes are normally expressed in normal individuals and abnormally expressed in gastric cancer patients, the calculation of the frequency of each sample and each category in the training samples in the target dataset can be performed separately according to their expressions, and the final output of the classification results and generation of classifiers according to the probability of each dataset sample are classified into the corresponding category [47].

K-nearest neighbor algorithm

K-nearest neighbor (KNN) [48] was proposed by Cover and Hart in 1968. The basic idea of the KNN algorithm is to calculate the distance between the samples to be classified and each training sample, according to the distance function, selects the k samples with the smallest distance from the samples to be classified as the KNN of x and finally determines the class of x according to the KNN of x. The KNN algorithm is simple, effective, and easy to understand [49].