Brief introduction to classifiers used in the research work

3.8.1 Naïve Bayes: Naïve Bayes classifier [31] is the simplest probabilistic classifier based on the Bayes theorem and assumes that features are independent. The posterior probability of the class for a given feature vector is computed as shown in equation (S1)

$$P(c|x) = \frac{P(x|c)*P(c)}{P(x)} \tag{S1}$$

where $P(x|c)$ is the likelihood probability, $P(c)$ is prior probability of the class, $P(x)$ is the prior probability of the feature and $P(c|x)$ is the posterior probability.

3.8.2 Logistic Regression: Logistic regression [32] uses the logit function to model the output of the linear equation between 0 and 1, and the logit function is shown in equation (S2).

$$logistic(\eta) = \frac{1}{1+\exp(-\eta)} \tag{S2}$$

where η is the linear equation that models the output based on the input features.

3.8.3 Linear Discriminant Analysis (LDA): LDA [33] assumes that features are normally distributed and the covariance of the features are equal, and the discriminant function of LDA is shown in equation (S3).

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2}\mu_k^T \Sigma^{-1} - \mu_k + \log(\pi_k) \tag{S3}$$

where Σ is the covariance matrix that is common to all classes, μ_k is the mean of the kth class and π_k is the prior probability that the instance belongs to the class k.

3.8.4 Quadratic Discriminant Analysis (QDA): QDA [33] is a statistical analysis technique that assumes that features are normally distributed and does not assume that the covariance of the features are same. The quadratic discrimination function is shown in equation (S4).

$$\delta_k(x) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) + \log \pi_k \tag{S4}$$

where Σ_k is the covariance matrix of the kth class, μ_k is the mean of the kth class and π_k is the prior probability that the instance belongs to the class k.

3.8.5 Logit Boost: Logit boost [34] is an additive logistic regression that tries to minimize the logistic loss shown in equation (S5).

$$\sum_i \log(1 + e^{-y_i f(x_i)}) \tag{S5}$$

where y_i is the actual class and $f(x_i)$ is the predicted class.

3.8.6 Stacking: Stacking [35], also known as stacked generalization, has base-classifiers which learn to predict the output for the given set of feature vectors. The meta-classifier of the stacking algorithm takes the output of the base-classifiers as input and estimates the ensemble output.