Chapter 5

An Ensemble Learning based method for Classification of Ground Exercises of Sattriya Dance

Many of the machine learning algorithms suffers from overfitting or underfitting problems like Decision Trees, SVM, and KNN. To reduce high bias and high variance ensemble learning algorithms can be used.

From the experimental results presented in the previous chapter it is observed that classification accuracies for group 2 (FA) ground exercises are not satisfactory. In this chapter, ensemble classifiers are used to improve the classification rates of ground exercises of Sattriya dance. The output of each single classifier within the two groups FF and FA, presented in previous chapter are combined and to make final decision, four well known ensemble methods are used: Random Forest, AdaBoost, Gradient Boost and XGBoost. These ensemble learning methods are based on Bagging and Boosting algorithm which are very powerful methods. Random Forest is based on Bagging algorithm. AdaBoost, Gradient Boost and XGBoost are Boosting based ensemble learning methods.

The rest of the chapter is organized as follows. In Section 5.1, a brief description of different ensemble classifiers chosen for this work is presented. The proposed method is illustrated in Section 5.2. In Section 5.3 experimental results are presented. Section 5.4 concludes the chapter with possible future direction.

5.1 Ensemble Learning Methods

The process of combining multiple classifiers for solving a problem is known as ensemble learning. Basically, it is used to improve the classification, prediction, and function approximation. Ensemble is a set of individual component classifiers whose predictions are combined to predict new incoming instances. Ensembles are also called a committee or multiple classifiers. It is an efficient way to improve predictive accuracy or breaking down a perplexing, troublesome difficult learning problem into easier sub-problems.

Wolpert [66] presented the no free lunch theorem that is the main motivation for using classifier ensembles. By this theorem, no individual classifier is sufficient for all the tasks, since each classifier has its own area of ability. Instead of taking single classifier if we take a set of multiple classifiers the performance can be improved.

A generic framework of ensemble classifier-based method is depicted in Figure 5.1.

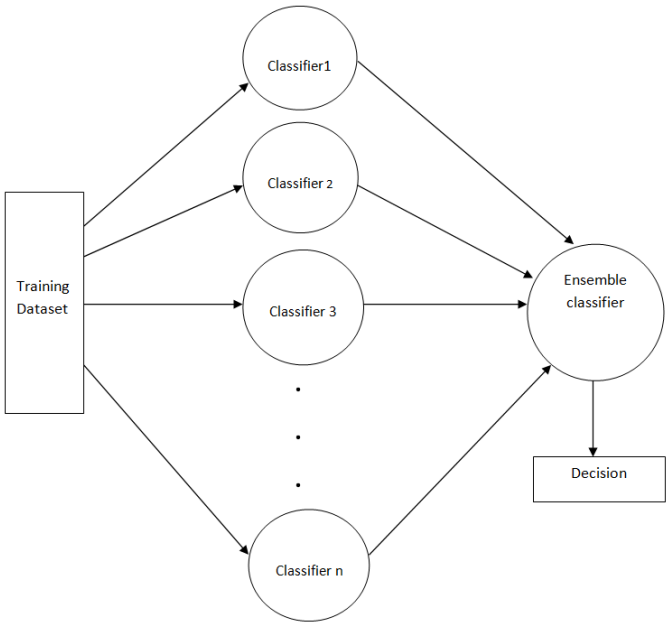

Figure 5.1 A generic framework of ensemble classifier

Bayesian averaging is the original ensemble method; however various methods of Bagging and Boosting are available. In this work we have selected ensemble classifiers from Bagging as well as Boosting algorithms. Random Forest, AdaBoost, Gradient Boost and XGBoost ensemble classifiers are used for classification of ground exercises of Sattriya dance.

5.1.1 Random Forest

Random forest is a bagging algorithm. It randomly chooses a bunch of feature elements that are used to decide the best split at each node of the decision tree. Decision tree is the base estimators in random forest. In a random forest model, arbitrary subsets are made from the initial dataset. At each node in the decision tree, only a random set of features are considered to decide the best split. A decision tree model is fitted on each of the subsets. The final prediction is calculated by averaging the predictions from all decision trees.

5.1.2 AdaBoost

AdaBoost or Adaptive boosting is one of the simplest boosting algorithms. Using boosting algorithm, the complexity of the high bias models can be boosted up. In AdaBoost algorithm, decision trees are used as base classifier. Here, multiple models are created sequentially, in which the errors in the last model are corrected at each step. At first, AdaBoost algorithm assigns equal weights to all observations and a model is built based on these observations. With this model, prediction is carried out on the whole dataset. By comparing the actual value with the predicted value errors are calculated. Now, for the incorrectly predicted data points, higher weights should be assigned for creating the next model. From the error values weights can be found out. Thus, AdaBoost algorithm creates models sequentially by assigning weights to the incorrectly predicted observations. Here, higher weights are assigned for the higher error observations. When a stable error function is obtained or the maximum limit of the number of estimators is reached, this process is completed.

5.1.3 Gradient Boosting

Gradient Boosting is another boosting ensemble learning algorithm. It works for regression, classification and ranking. In this work, the task is classification. Gradient boosting algorithm uses the boosting technique, combining a number of weak learners to form a strong learner. Here regression trees are used as a base learner in which each subsequent tree in series is built on the errors calculated by the previous tree.

5.1.4 XGBoost

XGBoost (extreme Gradient Boosting) is an advanced implementation of the gradient boosting algorithm. XGBoost has proved to be a highly effective ML algorithm. XGBoost has high predictive power and is faster than the other gradient boosting techniques. It also includes a variety of regularization which reduces over fitting and improves overall performance. Hence it is also known as regularized boosting technique.

5.2 The Proposed Method

In this work we have selected ensemble classifiers from Bagging as well as Boosting algorithms. Random Forest, AdaBoost, Gradient Boost and XGBoost ensemble classifiers are used for classification of ground exercises of Sattriya dance. This method classifies all the 28 ground exercises into two groups.

We have applied ensemble classifiers on the results presented in the previous chapter of single classifier within two groups of ground exercises, FF and FA.

The framework of the proposed method is depicted in Figure 5.2. Here we have taken five well known classifiers, KNN, SVM, Bayesian Network, Decision Tree and HMM that are discussed in the previous chapter. In this method the individual results of each single classifier are fed to the Random Forest ensemble classifier. Similarly, the outputs of each single classifier are fed to the ensemble classifiers AdaBoost, GradientBoost and XGBoost.

Finally, the decision is given by the ensemble classifiers. The experimental results are presented in the next section.

5.3 Experimental Results:

The experimental description and the obtained outcome for the ensemble learning based method on Sattriya dance ground exercise dataset is discussed in the following sub-sections.

5.3.1 Dataset Description

In this experiment, the feature dataset containing the three features- EN, EP and HW is used from the SDGE dataset which was described in chapter 3.

5.3.2 Results and Discussion

The average classification accuracy of the different ground exercises that are achieved are shown in Table 5.1. A comparative analysis is shown among the four states of the art ensemble classifiers for classification of the ground exercises of Sattriya dance. Here also classification is done based on the two groups of ground exercises: FF and FFBS.

Table 5.1: Classification accuracy of two groups of ground exercises

Ensemble Method	**Classification Rate**	
	Group 1 (FF)	**Group 2 (FA)**
Random Forest	94.73	80.32
AdaBoost	94.72	82.56
Gradient Boost	97.36	85.57
XGBoost	96.49	84.23

5.4 Summary

Here Gradient Boost ensemble learning method outperforms in both group of classes. For group 1, classification rate is 97.36% and for group 2, it is 85.57%. We hypothesized that the data are generated by multiple models. These different models reflect the fact that the different classes may have different characteristics accounting for intra-class variability within the data. Here ensemble classifiers are used to improve the classification accuracy of ground exercises of Sattriya dance. To combine the output of single classifiers presented in previous chapter, and make final decision, four well known ensemble methods are used: Random Forest, AdaBoost, Gradient Boost and XGBoost. These ensemble learning methods are based on Bagging and Boosting algorithm which are very powerful methods.