

Evaluating Heterogeneous Ensembles with Boosting Meta-Learner



Seba Susan , Aishwary Kumar, and Anmol Jain

Abstract In this paper, heterogeneous ensemble of classifiers is evaluated and the outputs are integrated by a boosting meta-learner. Both ADABOOST and XGBOOST are tried for the meta-learning stage, and XGBOOST performed best. The heterogeneous ensemble consists of a diverse set of base classifiers—k-nearest neighbors, logistic regression, Support Vector Machines (linear, Gaussian kernels), Random forest of decision trees, and Naïve Bayes classifier. Smaller ensembles are also hierarchically formed by removing the weak learner in every stage. The Entropy of base classifier predictions is computed to identify the presence of weak learners. The predictions of the base classifiers are learnt by the boosting meta-learner using a 9:1 split of the training data, where 9 parts are used for training the base classifiers and 1 part for obtaining the ensemble predictions and training the meta-learner. A 10-fold cross-validation is introduced to avoid bias. Experimental results show higher scores on evaluating the Human Action Recognition (HAR) smartphone dataset using our ensemble model as compared to the other state-of-the-art models.

Keywords Meta-learning · Boosting · XGBOOST · Heterogeneous ensemble

1 Introduction

Recognizing human actions from mobile or smartphone information is evolving research with significant outcomes. Mobile phones contain the accelerometer sensor that tracks motion over time. Potential applications include tracking the motion of the elderly [1], detecting suspicious motion in crowded places [2], and analysis of sports activities [3]. The limitations of most of the current datasets based on smartphone sensors are that they are specific and relatively smaller in size. Hence complex models like deep neural networks do not yield good performance. Using multiple classifiers

S. Susan (✉) · A. Kumar · A. Jain

Department of Information Technology, Delhi Technological University (DTU), Bawana Road, Delhi 110042, India

e-mail: seba_406@yahoo.in

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 145,

https://doi.org/10.1007/978-981-15-7345-3_60

is a reliable technique of performance enhancement in machine learning. For smaller datasets, stacking combinations of classifiers is more useful rather than relying on a single model [4]. It was observed in [5] that heterogenous models give an improved performance than homogenous models for small datasets. Integrating the outputs of multiple classifiers is achieved by bagging, boosting, or stacking [6]. Various strategies have been proposed to integrate the outputs of multiple classifiers in a classifier ensemble, ranging from majority voting to advanced meta-learners [7]. In the meta-learning scheme, the predictions of the base classifiers in the ensemble are learnt by a meta-learner classifier in a non-linear manner. A combination of stacking and boosting is investigated in this work. Stacking of heterogenous classifiers is implemented for the formation of a diverse ensemble whose predictions are learnt by a boosting meta-learner. An Entropy score is assigned to the ensemble based on the feature weights computed by the boosting algorithm that serves to measure the efficiency of the ensemble. The organization of this paper is as follows: a review of meta-learners and base learners in a stacked ensemble classifier model is given in Sect. 2, the proposed heterogeneous ensemble model based on boosting meta-learner is presented in Sect. 3, the results are analyzed in Sect. 4 and the final conclusions are drawn in Sect. 5.

2 Ensemble Learning—A Review of Basic Concepts

2.1 *Meta-Learning with Stacked Ensemble of Classifiers*

A meta-learner is an additional classification layer in an ensemble model for fine-tuning the base classifier predictions. The term was introduced by Wolpert in [8], though several indicative works pre-exist [9]. Various learning strategies for stacked classifier ensemble were investigated by Chan and Stolfo in [10] to improve the performance of the ensemble as compared to that of the base classifiers. Binary classifier predictions and fusion of attributes with base classifier predictions were found to be successful in this regard. The precursor of meta-learning is classifier ensembles in which a pre-defined number of trained classifiers individually predict the class label of the test sample and the decisions are compiled, usually, by majority voting [7]. Majority voting finds its genesis in the history of social sciences, wherein decisions are taken by social committees by a popular vote among the committee members [11]. A variety of meta-learners have been tested and tried over the years. Table 1 summarizes the base and meta-classifiers in a few of these works.

Approximate Ranking Tree Forests are proposed as the meta-learner in [12], which involves comparing base classifiers with each other to generate meta-features. Decision trees are observed to be used by many researchers as the base classifier. In [13], a meta-learning approach is proposed with the goal of interpreting the hidden structures (having H neurons) in Convolutional Neural Networks (CNN). Here, a $H \times N$ feature matrix of cluster IDs is fed to the meta-learner (Decision Tree) along with

Table 1 Examples of a few distinctive ensemble models and Meta-classifier combinations in literature

Method	List of base classifiers	Meta-learner
Aydin (2018) [1]	LDA, SVM, neural networks	Fuzzy Integral
Rodriguez et al. [26]	Rotation forest: Decision trees (with PCA)	Majority voting
Nguyen et al. (2018) [27]	LDA, Naïve Bayes, k-Nearest Neighbors	Fuzzy IF-THEN rules
Alexandropoulos et al. (2019) [28]	Decision trees	ACO-stacking
Melville and Mooney (2004) [29]	Decision trees (with artificial data)	Meta-learning algorithm ‘ACTIVE-decorate’

N training labels for training. The test vector comprising of cluster IDs is classified by the meta-learner into the actual label of the test sample.

2.2 Learning with Heterogenous Ensemble of Classifiers

Most of the ensemble methods involve homogeneous classifiers due to ease of programming and parameter selection [14]. The ensemble output is compiled in some form of average error performance across the classifiers in the ensemble. The common ensemble methods that involve homogeneous classifiers are bagging, boosting, and random forest, depending on the integration technique [15]. However, lack of diversity in classifier predictions due to similar type of errors in predictions limits the performance of homogeneous ensembles. Classifiers such as neural networks are subject to local entrapping of solutions due to the derivative optimization process [16]. The network is hence, not sensitive to the variety present in the dataset. Corrective procedures do exist, such as, Network Architecture Search (NAS) for optimal parameters [17] and dynamic neural networks that evolve as training progresses [18]. Locality of solutions and lack of variety is the motivation behind using a heterogeneous mix of classifiers in an ensemble. For the homogenous classifiers in an ensemble, variety can be induced by carefully sampling the training data and the feature subspace for custom-training of each member in the ensemble [19]. Lack of variety in predictions promoted research in heterogenous ensembles where a variety of popular classifiers like logistic regression, k-Nearest Neighbors, Naïve Bayes, Support Vector Machines, and decision trees are trained on the same data. The outputs of individual classifiers are integrated by committee voting or weighted average [20]. In a rather unconventional approach in [1], the predictions of the base learners, comprising of Linear Discriminant Analysis classifier (LDA), Support Vector Machines (SVM), and neural networks, are combined using the fuzzy

integral. Prior to the integration, the confidence-scores of the base learners are optimized using cuckoo search. The idea is to efficiently utilize the diversity of the classifiers in the ensemble.

3 Proposed Meta-Learning for Heterogeneous Ensemble

3.1 *The Problem Statement*

The focus in this research is on heterogeneous ensemble. Its characteristic is the diversity of the base classifiers that are the constituents of the ensemble. Due to the diversity, the predictions made by the base classifiers will have a variety, which if integrated by a carefully planned meta-learning classification stage, would yield high accuracies of prediction. All the base classifiers are trained on the same data. The meta-learner for most heterogeneous ensembles in literature is majority voting. The maximum vote amongst the base classifier predictions indicates the class label of the test sample. In case of an unclear majority, an ambiguity exists, that is primarily ignored by most researchers. In our work, this problem is addressed by introducing a boosting classifier in the meta-learning stage that learns the entire set of predictions made by the base classifiers. Boosting algorithms are by themselves, based on ensembles of decision trees. These are known to concentrate on ambiguous data points for decision-making. This addresses the issue of those predictions that do not form a clear majority and are difficult to summarize. The subset of training data used in the meta-learning stage is different from that used for training the base classifiers.

3.2 *Proposed Meta-Learning for Heterogeneous Ensemble*

In our proposed method, the boosting algorithm is employed as the meta-learner in the second stage of classification where it learns the predictions from the base learners in the first stage. The meta-classifier integrates the predictions of all the base classifiers in a decision-making module. The base learners are a heterogeneous mix of classifiers that are trained on the same training data. Ensembles of six, five, four, and three classifiers are tested, with the heterogeneous classifier models being k -nearest neighbors, logistic regression, Support Vector Machines (linear, Gaussian kernels), Random forest of decision trees and Naïve Bayes classifier. The diversity of the base classifiers aids in unbiased prediction. The training set is divided into 10 parts out of which 9 parts are used for training the base classifiers and 1 part out of 10 is used for training the meta-learner. Ten-fold cross-validation is used to avoid bias. The algorithm for our meta-learning model incorporating boosting in the latter stage is given below.

Algorithm Meta-learning by boosting for heterogeneous ensemble

Input: *Training set_9parts, Training set_1part, Test set with unlabeled data*

Parameter: Base-learner specific

Output: Class label of *Test* sample

- 1: For each base-learner C_i say, Support Vector Machines
- 2: perform learning on *Training set_9parts*
- 3: obtain predicted class label $C_i(\mathbf{T})$, for $\mathbf{T} = \text{Training set}_1\text{part}$
- 4: From all the n base-learners, extract:
 $\text{Meta-features} = \{C_1(\mathbf{T}), C_2(\mathbf{T}), \dots, C_n(\mathbf{T})\}$
- 5: Learn the Meta-features using boosting based meta-learner
- 6: Extract the Meta-features from *Test* sample and give as input to the trained meta-learner for predicting the class label of *Test*
- 7: **return** solution

The choice of the boosting algorithm is decided next. Freund et al. (1999) proposed ADABOOST [21] as a novel boosting ensemble technique that combines the outputs of a number of weak learners or estimators that constitute the boosting ensemble. Since the method intrinsically assigns higher weights to noisy samples that are least correlated with the output label, it generally yields high accuracies and is known to be a competitor to ensemble models. In [22], a new meta-learning algorithm was devised on ADABOOST principles called meta-boosting that combines the outputs of weak learners by applying a strong learner on them. XGBOOST (Chen and Guestrin, 2016) [23] stands for eXtreme Gradient Boost that enumerates splitting points in a tree in a greedy manner. It starts with a leaf and enumerates the gradient of the dynamically changing loss function and adds branches in an additive manner. While in ADABOOST, the focus is on the misclassified samples that are assigned higher weight, in XGBOOST, the focus is on the gradient with faster optimization algorithms. XGBOOST employs L1 and L2 regularization that prevents overfitting. Since the conventional approach to integrate ensemble decision is by majority voting [20], we explore the use of boosting as a meta-learner that focuses on the difficult-to-learn-examples in the base classifier predictions. Both ADABOOST and XGBOOST are applied, within our heterogeneous ensemble framework, as the meta-learner that receives the meta-features or predictions of individual base classifiers. In our case, the difficult-to-learn-examples in the meta-features are the prediction vectors that indicate no clear majority amongst the base classifiers in the ensemble. We do not partition the dataset into smaller subsets for training the base classifiers, following the observation by Chan and Stolfo (1995) [24], that partitioning smaller datasets may have an overall negative impact on the learning performance if do not have remedies at hand to reduce the bias that ensues because of the partitioning. XGBOOST which is the gradient boosting algorithm has won several Kaggle competitions due to its unique features such as scalability and fast execution, its tendency to form deeper trees with less variance, and computing similarities with data points in an adaptive neighborhood. To identify the presence of weak learners in an ensemble, an Entropy grade is now computed, as

$$H = - \sum_p p \log p \tag{1}$$

The Entropy measures the randomness among a set of points [3] and is computed here from the normalized feature weights ($=p$) assigned by the boosting algorithm; the features being the base classifier predictions and weights being the frequencies of occurrence of the features in all the tree splits. A higher Entropy indicates more equal importance among the features which are the predictions of the base classifiers. A lower Entropy indicates the presence of one or more weak learners whose predictions are not contributive to the final decision taken. The ensemble with a high normalized Entropy (Entropy/Maximum value of Entropy) is considered the optimal choice of an ensemble for the given dataset.

The functional block diagram of the proposed model is shown in Fig. 1. The six heterogeneous classifiers are shown along with the meta-learner. The overall process flow for the learning model shown in Fig. 1, and summarized in the following steps:

Step 1: Split the training set in a 9:1 ratio (with ten-fold cross-validation) labeled *Training set_9parts* and *Training set_1part*.

Step 2: Train the base classifiers with *Training set_9parts* (shown by blue (solid line) arrows in Fig. 1).

Step 3: Train the boosting meta-learner on the predictions made by the trained base classifiers on the data *Training set_1part* (shown by red dotted arrows in Fig. 1).

Step 4: In the testing phase, apply the predictions made by the trained base classifiers on the test data as input to the trained boosting meta-learner (red dotted arrows only).

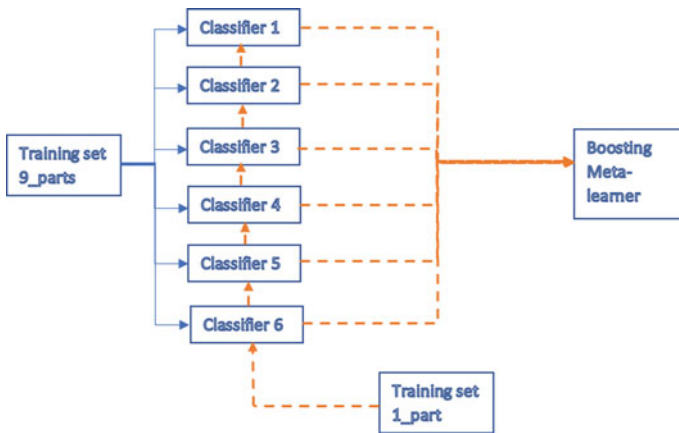


Fig. 1 Functional block diagram of the proposed model

4 Experimental Results and Discussions

The experiments are conducted in Python 3.7 version software on an Intel Pentium processor. The benchmark dataset on Human Activity Recognition (HAR) based on smartphone data [25] is used for the experiments. It is segregated at the source into 7352 samples (from 21 subjects) and 2947 samples (from 9 subjects) for training and testing, respectively. There are 561 features in all, representing 2.56 s of human activity. The activities themselves are divided into six categories—*walking, upstairs, downstairs, sitting, standing, lying*.

4.1 Results of Meta-Learning by Boosting for Heterogeneous Ensemble

The implementation of our meta-learning model is done as per the guidelines in Sect. 3. Experiments are conducted, in hierarchical order, for ensembles of 6, 5, 4 and 3 heterogeneous classifiers with the heterogeneous models being— k -nearest neighbors ($k = 8$), logistic regression, support vector machine (linear, Gaussian kernels), Random forest of decision trees (300 trees) and Naïve Bayes classifiers. The compositions of the heterogeneous ensembles are shown in Tables 2 and 3 respectively. The worst performer is removed at each stage to form the smaller ensemble. Two different boosting meta-learners—ADABOOST, XGBOOST are investigated for our scheme. A Grid Search strategy is used for tuning the hyperparameters (learning rate, number of estimators, etc.) of the boosting algorithms. A stratified 10-fold cross-validation scheme is used for the training data in our ensemble learning. As per this scheme, the training data is split into ten parts, nine parts are given as aggregate input to the

Table 2 Our Meta-learning strategy for 6- and 5-classifier ensembles (Test accuracy in %) with comparison to majority voting scheme

Type of classifier	Ensemble # (5 heterogeneous models)	Ensemble # (5 heterogeneous models)	Ensemble # (6 heterogeneous models)
Base classifiers	*Naïve Bayes—77.02% *Logistic Regression—96.19% *SVM (Gaussian)—94.02% *Random forest (100 trees)—90.19% *k-NN (k = 8)—90.73%	*SVM (linear)—96.40% *Logistic Regression—96.19% *SVM (Gaussian)—94.02% *Random forest (300 trees)—92.63% *k-NN (k = 8)—90.73%	*Naïve Bayes—77.02% *SVM (linear)—96.40% *Logistic Regression—96.19% *SVM (Gaussian)—94.02% *Random forest (300 trees)—92.63% *k-NN (k = 8)—90.73%
Meta-learner	Majority voting—95.12%	ADABOOST— 96.64% XGBOOST—96.64%	ADABOOST— 96.26% XGBOOST—96.77%

Table 3 Our Meta-learning strategy for 4- and 3- classifier ensembles (Test accuracy in %)

Type of classifier	Ensemble # (4 heterogenous models)	Ensemble # (3 heterogenous models)
Base classifiers	*SVM (linear)—96.40% *Logistic Regression—96.19% *SVM (Gaussian)—94.02% *Random forest (300 trees)—92.63%	*SVM (linear)—96.40% *Logistic Regression—96.19% *SVM (Gaussian)—94.02%
Meta-learner	XGBOOST—96.40% ADABOOST—96.40%	XGBOOST—96.36% ADABOOST—96.33%

ensemble of classifiers. The tenth part is used to evaluate the predictions of the base classifiers that are given as input to the meta-learner. The results are shown in Table 2 for six and five ensemble combinations for both boosting and the majority voting scheme. The results in Table 2 indicate that XGBOOST (*base-estimator = Random forest, number of estimators = 300*) gives the best results, much higher than majority voting.

Among the base learners, SVM with linear kernel gives the individual best accuracy followed by logistic regression. The Naïve Bayes classifier proves to be a weak learner in Table 2 with an accuracy of 77.02%. The weak learner affects the results of ADABOOST for which the best results are observed for the 5-classifier ensemble. The gradient boosting scheme is found to override the weak learner as observed from the highest results of 96.77% in Table 2 for the six-classifier ensemble. A similar observation is made for the results 4- and 3-classifier ensembles in Table 3.

Overall, the best accuracy is observed for the XGBOOST meta-learner for the 6-classifier ensemble. The Entropies of various ensembles are computed in Table 4 from the relative feature importance plots shown in Fig. 2. The feature weights are shown normalized so that they form a complete probability distribution (sum of all p is 1).

From Table 4, a higher normalized Entropy of features is observed for XGBOOST as compared to that of ADABOOST. The heterogeneous ensemble with the highest Entropy ($=0.978$ for 6-classifier ensemble), in the training phase, is determined to be most suitable for the given dataset. Overall, gradient boosting has proved to be the best meta-learner for our ensemble. It is generally observed that due to meta-learning there is improvement over the performance of all the individual classifiers in the ensemble.

Table 4 Entropy of ensemble (Highlighted values indicate the normalized Entropies)

Number of base classifiers in ensemble = n	3	4	5	6
$Max Entropy = \log(n)$	1.0986	1.3863	1.6094	1.6094
Entropy—ADABOOST	0.7430	0.8741	1.1143	1.4847
(Entropy—ADABOOST/Max Entropy)	0.676	0.6305	0.692	0.9225
Entropy—XGBOOST	1.0954	1.1571	1.1143	1.5740
(Entropy—XGBOOST/Max Entropy)	0.997	0.8347	0.6924	0.978

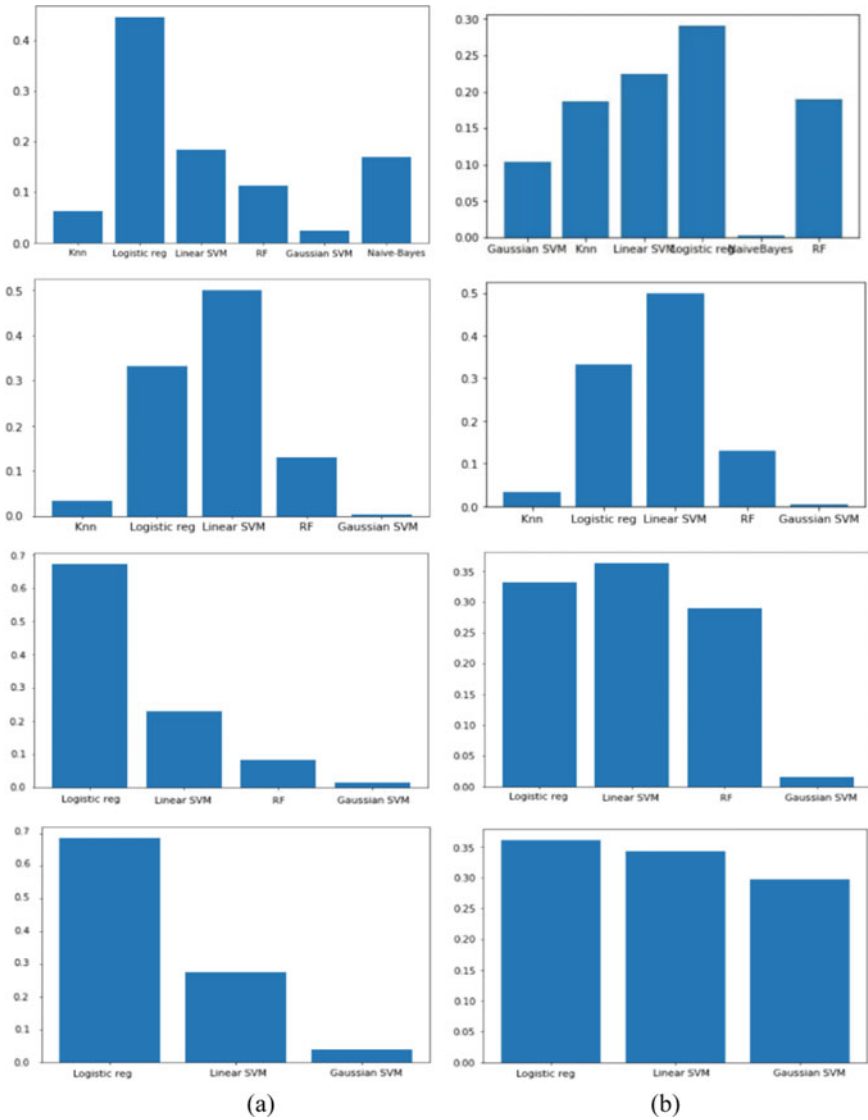


Fig. 2 The relative feature importance plot for base classifiers in the ensemble **a** ADABOOST meta-learner **b** XGBOOST meta-learner [top to bottom: 6-, 5-, 4-, 3-classifier ensembles]

4.2 Comparison to the State-of-the-Art Techniques

Comparison of our results to some state-of-the-art methods listed in Table 5, highlights the efficiency of our approach from the higher accuracies achieved by gradient boosting (=96.77% for XGBOOST as meta-learner for the 6-classifier ensemble).

Table 5 A comparison of the recent state-of-the-art results: test accuracy (in %)

Recent state-of-the-art results for smartphone HAR	Accuracy (%)
Probabilistic-First-Take-It-All (PFTA) [30] (Ye et al. 2018)	92
Deep ConvNet [31] (Jiang and Yin 2015)	95.18
CNN + Attention model [32] (Wang et al. 2019)	93.4
Deep Neural Net [33] (Ronao et al. 2016)	94.61
Proposed 6-classifier heterogeneous ensemble with XGBOOST meta-learner	96.77

5 Conclusion

A heterogenous ensemble of diverse classifier models with a boosting meta-learner is proposed in this work for recognizing human activities from smartphone data. In this paper, the boosting meta-learner is employed to take advantage of the diversity of predictions made by the ensemble. A heterogenous mix of classifiers namely— k -nearest neighbors, logistic regression, Support Vector Machines (linear, Gaussian kernels), Random forest of decision trees, and Naïve Bayes classifier forms the six-classifier ensemble. XGBOOST, that is a tree-based gradient boosting algorithm, is the meta-learner that performs the final decision regarding the test label based on the individual predictions of the ensemble classifiers. The ADABOOST boosting algorithm is also used for comparison. Higher classification scores especially with XGBOOST validate the efficiency of our meta-learning model despite the presence of a weak learner in the ensemble. Entropy of the feature weights assigned during boosting is computed for grading the ensemble based on participation of the constituents in the decision-making. A high Entropy coupled with high accuracy is observed for the six-classifier ensemble, in the case of XGBOOST, as compared to the smaller classifier ensembles. Application of our ensemble on large datasets with federated learning forms the future scope of our work.

References

1. Aydin I (2018) Fuzzy integral and cuckoo search based classifier fusion for human action recognition. *Adv Electr Comput Eng* 18(1):3–11
2. Susan S, Hanmandlu Ma (2015) Unsupervised detection of nonlinearity in motion using weighted average of non-extensive entropies. *SIViP* 9(3):511–525
3. Susan S, Chaurawat S, Nishad V, Sharma M, Sahay S (2016) Speed and trajectory based sports event categorization from videos. In: 2016 international conference on signal processing and communication (ICSC). IEEE, pp 496–501
4. Yu C, Skillicorn DB (2001) Parallelizing boosting and bagging. Queen's University, Kingston, Canada, Tech. Rep
5. Merkwirth C, Wichard J, Ogorzałek MJ (2009) Ensemble modeling for bio-medical applications. In: *Modelling dynamics in processes and systems*. Springer, Berlin, Heidelberg, pp 119–135

6. Matan O (1996) On voting ensembles of classifiers. In: Proceedings of AAAI-96 workshop on integrating multiple learned models, pp 84–88
7. Sagi O, Rokach L (2018) Ensemble learning: a survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 8(4):e1249
8. Wolpert David H (1992) Stacked generalization. *Neural Netw* 5(2):241–259
9. Mitchell TM (1980) The need for biases in learning generalizations. Department of Computer Science, Laboratory for Computer Science Research, Rutgers University, New Jersey
10. Chan PK, Stolfo SJ (1993) Experiments on multistrategy learning by meta-learning. In: Proceedings of the second international conference on information and knowledge management. ACM, pp 314–323
11. Brams SJ, Fishburn PC (2002) Voting procedures. *Handb Soc Choice Welfare* 1:173–236
12. Sun Q, Pfahringer B (2013) Pairwise meta-rules for better meta-learning-based algorithm ranking. *Mach Learn* 93(1):141–161
13. Liu X, Wang X, Matwin S (2018) Interpretable deep convolutional neural networks via meta-learning. In: 2018 international joint conference on neural networks (IJCNN). IEEE, pp 1–9
14. Wang Y, Wang D, Geng N, Wang Y, Yin Y, Jin Y (2019) Stacking-based ensemble learning of decision trees for interpretable prostate cancer detection. *Appl Soft Comput* 77:188–204
15. Seni G, Elder JF (2010) Ensemble methods in data mining: improving accuracy through combining predictions. *Synth Lect Data Min Knowl Discov* 2(1):1–126
16. Susan S, Rohit R, Udyant T, Shivang R, Pranav A (2019) Neural net optimization by weight-entropy monitoring. In: Computational intelligence: theories, applications and future directions. Springer, Singapore, vol II, pp 201–213
17. Verma M, Sinha P, Goyal K, Verma A, Susan S (2019) A novel framework for neural architecture search in the Hill Climbing Domain. In: 2019 IEEE second international conference on artificial intelligence and knowledge engineering (AIKE). IEEE, pp 1–8
18. Susan S, Dwivedi M (2014) Dynamic growth of hidden-layer neurons using the non-extensive entropy. In: 2014 fourth international conference on communication systems and network technologies. IEEE, pp 491–495
19. Queipo NV, Nava E (2019) A gradient boosting approach with diversity promoting measures for the ensemble of surrogates in engineering. *Struct Multi Optim* 60(4):1289–1311
20. Prodromidis AL, Stolfo SJ (1999) A comparative evaluation of meta-learning strategies over large and distributed data sets. In: Workshop on meta-learning, sixteenth international conference on machine learning, pp 18–27
21. Freund Y, Schapire R, Abe N (1999) A short introduction to boosting. *J Jap Soc Artif Intell* 14(771–780):1612
22. Liu X, Wang X, Japkowicz N, Matwin S (2013) An ensemble method based on adaboost and meta-learning. In: Canadian conference on artificial intelligence. Springer, Berlin, Heidelberg, pp 278–285
23. Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, pp 785–794
24. Chan PK, Stolfo SJ (1995) A comparative evaluation of voting and meta-learning on partitioned data. In: Machine learning proceedings 1995. Morgan Kaufmann, pp 90–98
25. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL (2012) Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In: International workshop on ambient assisted living. Springer, Berlin, Heidelberg, pp 216–223
26. Rodriguez JJ, Kuncheva LI, Alonso CJ (2006) Rotation forest: a new classifier ensemble method. *IEEE Trans Pattern Anal Mach Intell* 28(10):1619–1630
27. Nguyen TT, Nguyen MP, Pham XC, Liew AW-C (2018) Heterogeneous classifier ensemble with fuzzy rule-based meta learner. *Inf Sci* 422:144–160 (2018)
28. Alexandropoulos S-AN, Aridas CK, Kotsiantis SB, Vrahatis MN (2019) Stacking strong ensembles of classifiers. In: IFIP international conference on artificial intelligence applications and innovations. Springer, Cham, pp 545–556

29. Melville P, Mooney RJ (2004) Diverse ensembles for active learning. In: Proceedings of the twenty-first international conference on machine learning. ACM, p 74
30. Ye J, Qi G, Zhuang N, Hao Hu, Hua KA (2018) Learning compact features for human activity recognition via probabilistic first-take-all. *IEEE Trans Pattern Anal Mach Intell* (2018)
31. Jiang W, Yin Z (2015) Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proceedings of the 23rd ACM international conference on Multimedia. ACM, pp 1307–1310
32. Wang K, He J, Zhang L (2019) Attention-based convolutional neural network for weakly labeled human activities recognition with wearable sensors. *IEEE Sens J*
33. Ronao CA, Cho S-B (2016) Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst Appl* 59:235–244