Enhancing Decision Combination in Classifier Committee via Positional Voting

Jacek Trelinski and Bogdan Kwolek

AGH University of Science and Technology
30 Mickiewicza, 30-059 Krakow, Poland
{tjacek,bkw}@agh.edu.pl

Abstract. In this work, we propose an approach for aggregating classifiers using positional voting techniques. We extend the positional voting by optimizing weights of the preferences to better aggregate the committee classifiers. Staring from initial weights determined by a voting algorithm the aggregating weights are optimized by a differential evolution algorithm. The algorithm has been evaluated on a human action dataset. We demonstrate experimentally that on SYSU 3DHOI dataset the proposed algorithm achieves superior results against recent algorithms including skeleton-based ones.

Keywords: Classifier committee, collective intelligence, voting rules.

1 Introduction

Ensemble techniques combine a number of diverse models to build a composite model that improves generalizability/robustness over each of them alone, either by using many different modeling algorithms or using different training datasets. They involve aggregating multiple models with the aim of decreasing both bias and variance. A classifier committee (or ensemble) is a classifier constructed by combining the predictions of multiple classifiers [1].

Classifier committees tend to yield better results if the individual classifiers work as independently as possible, i.e. when there is a significant diversity among the models [2]. The conventional ensemble methods include bagging, boosting, and stacking-based methods. These methods have been well studied in recent years and applied widely in different applications [3]. More recent approaches for ensemble learning such as XGBoost and LightGBM [4] permit achieving very competitive results on commonly used benchmark datasets. In the last decade, due to availability computational power that permits training large ensembles in a reasonable time, the number of ensemble-based applications has grown increasingly.

In ensemble learning we can distinguish three phases [5]. The first one consists in generating a set of models, and aims at obtaining a pool of models. In the second step a single classifier or a subset of best classifiers is selected. In the last phase a strategy to combine individual models is defined. The combination

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

of classifiers can be achieved using class labels, such as in the majority voting scheme, which calculates the prediction on the basis of the most frequent class assignment, or by utilizing scores of individual classifiers. In [6], a weighted voting ensemble was employed to improve the classification model's performance by combining classification results and selecting a group with the highest vote based on the weights given to the single classifiers. The impact of ensemble size with majority voting and optimal weighted voting aggregation rules has recently been discussed in [7]. Ranked voting approaches are recommended for combining classifiers if and when the classifiers can rank the order of the classes [8]. Borda count is a rank-based combination technique in which each classifier ranks the classes according to their potentiality to be the correct class [9]. Weights are linearly proportional to position in the ordering. It is considered to be one of the simplest scoring rules. Recently, in [10] a feature selection using election methods and ranker clustering for ensembles has been proposed.

Human action recognition is an important component in many applications including but not limited to ambient intelligence [11], human-computer interaction systems, and video surveillance. Little research has been done in the area of human action recognition on raw depth maps [12]. Recently, in [13] an algorithm for human action classification on depth motion images and Laplacian pyramids as structured multi-scale feature maps has been proposed. In a recent work [14], voting rules as an aggregation technique for classifier combination have been utilized to improve human action recognition on raw depth maps. In this work, an approach to aggregating classifiers through positional voting techniques is proposed. The proposed optimized positional voting achieves better results in comparison to results achieved by Borda count, Coombs, Bucklin, and Copeland.

2 Algorithm

The architecture of the classifier committee is based on architecture discussed in [15]. With the feature extraction in mind, the main difference is that in this work a Dynamic Time Warping (DTW) is utilized instead of the shapelets. In Section 3 we compare results achieved by both versions of the algorithm. In Subsection 2.1 we outline main ingredients of classifier committee. Afterwards, in Subsection 2.2 we present a DTW-based extraction of action features. Finally, in Subsection 2.3 we introduce the optimized positional voting.

2.1 Main Ingredients of Classifier Committee.

Having on regard that most frequently used benchmark datasets for human action recognition on depth maps contain limited number of depth map sequences, we utilize a multiple classifier system. The recognition of human actions is performed on raw depth maps only. We learn various features in different domains, like single depth map, time-series of embedded features, time-series represented by DTW-based features. A Siamese neural network as well as a convolutional

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

autoencoder (CAE) are learned on single frames to extract features. The multivariate time-series of Siamese features are fed to the DTW in order to extract the features representing actions. The multivariate time-series of CAE features are fed to a 1D CNN in order to extract another representation of action features. The discussed features are common for all classes. We extract also class-specific action features using TimeDistributed and LSTM layers (TD-LSTM). Multiclass logistic classifiers are trained on concatenated class-specific features and features common for all classes, see Fig. 1. The most discriminative classifiers are selected using a differential evolution (DE) [15]. The final decision is made through aggregating classifiers on the basis of voting.

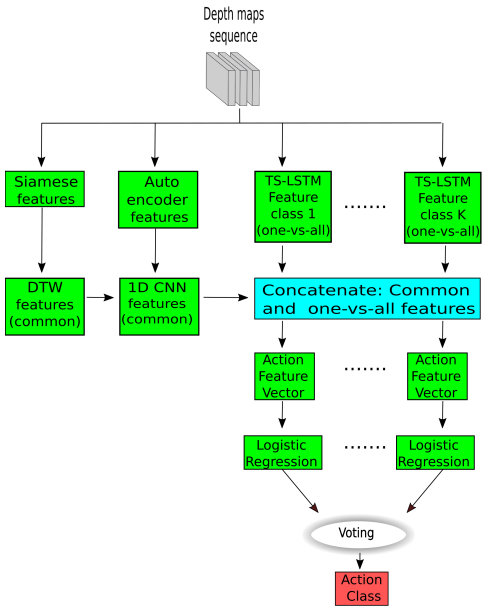

Fig. 1. Flowchart of the classifier committee operating on learned features, and aggregating decisions of selected classifiers via optimized positional voting.

2.2 Dynamic Time Warping-based Action Features

In order to compactly represent actions we learn features on representative depth maps. A Siamese neural network is trained on pairs of single depth maps as a representation of the whole depth map sequences. The network trained on such a compact representation of depth map sequences is used to extract frame-features on depth map sequences. In contrast to the CAE it has been trained only on frontal depth maps. In the current implementation, for each sequence

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

from the training subset a middle depth map as a representation the whole depth map sequence has been selected and then included in a training subset for the Siamese neural network. The Siamese neural network operates on depth maps of size $1 \times 64 \times 128$. It consists of 64 Conv2D filters of size 5×5 followed by max-pooling, 32 Conv2D filters of size 5×5 followed by max-pooling, which in turn are followed by the flattening layer and then a dense layer consisting of 128 neurons. The neural network has been trained using the contrastive loss [16]. The neural network trained in such a way was used to extract features on every depth map from a given input sequence. A human action represented by a number of depth maps is described by a multivariate time-series of length equal to number of frames in the sequence and dimension equal to 128.

In time-series analysis, the DTW can be employed for measuring similarity between two temporal sequences, which may vary in speed. Let us assume that our aim is to measure the distance between two time-series: $\mathbf{a} = \{a_1, a_2, \ldots, a_n\}$ and $\mathbf{b} = \{b_1, b_2, \ldots, b_n\}$. Let us denote by $M(\mathbf{a}, \mathbf{b})$ the $n \times n$ pointwise distance matrix between $\mathbf{a}$ and $\mathbf{b}$, where $M_{i,j} = (a_i - b_i)^2$. A warping path $P = (u_1, v_1), (u_2, v_2), \ldots, (u_s, v_s)$ is a set of pairs of indexes that define a traversal of M. The valid warping path must satisfy: $(u_1, v_1) = (1, 1)$ and $(u_s, v_s) = (n, n)$ as well as $0 \leq u_{i+1} - u_i \leq 1$ and $0 \leq v_{i+1} - v_i \leq 1$ for all $i < n$. Let p_i stand for the distance between elements with indexes u_i and v_i for the i-th pair of points. The distance for a path P is: $D_P(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{s} p_i$. The DTW path P^* has the minimum distance

$$P^* = \min_{P \in \mathcal{P}}(D_P(\mathbf{a}, \mathbf{b})) \tag{1}$$

over all possible paths $\mathcal{P}$, which can be found by dynamic programming (DP). For a given depth map sequence we calculated the DTW distances with respect to all remaining depth maps sequences in the training subset. For each depth map sequence the DTW distances between multivariate time-series have been determined for the discussed above Siamese features. The distances between a given sequence and all remaining sequences from the training set have then been utilized as features. This means that the resulting feature vector is of size equal to $n_t \times 2$, where n_t denotes the number of training depth map sequences.

2.3 Optimized Positional Voting

In a few papers the voting-based aggregation techniques have been utilized to improve the performance of ensembles [10]. Recently, in [14] Borda counts, Bucklin and Coombs have been investigated in terms of improving the performance of action classification using a classifier committee. In this work, we examine positional voting [17], in which points are allocated to the candidates under consideration based on the order they were ranked. In such voting systems, the voters order the candidates from best to worst, and a pool of winners is chosen on the basis of positions of the candidates in the preference order. The rank position of each voter preference has allocated a specific fixed weighting [17]. A candidate with the most points overall wins. Borda counting is an important example of

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

positional voting techniques. Figure 2 illustrates the process of positional voting in an example classifier committee.

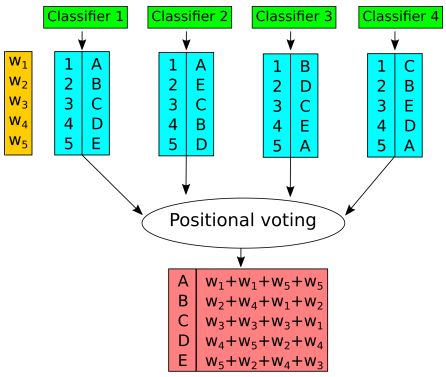

Fig. 2. Positional voting.

In this work, we extend the positional voting by optimizing weights of the preferences. Area Under the ROC (Receiver Operating Characteristics Curve), commonly called AUC has been utilized in the objective function. Direct optimization of AUC can lead to solving an NP-hard problem since it can be cast into a combinatorial optimization problem. Recently, in [18] a fast stochastic AUC optimization with $O(1/n)$ convergence rate has been proposed. In this work, the objective function mentioned above has been optimized using the differential evolution. The differential evolution initiated its search from a population of weights determined by the Borda count algorithm. The convergence is reached when the standard deviation of the fitness function for each individual in the population, normed by the average, is smaller than the given tolerance value. The weights determined in such a way have been utilized in making the final decision by the classifier committee.

Figure 3 presents a schematic diagram of algorithm steps. In the considered toy example, we assume that a dataset consists of five samples, whereas the classifier committee comprises four classifiers. The dataset is split into a training subset on which the classifiers are trained and a validation subset on which the weights are optimized. This means that for training of the classifiers as well as optimization of the weights of the preferences the training data was further divided into the training subset and the validation subset. In the considered example, all four classifiers are trained on the training subset. The validation subsets are fed into the trained classifiers, whose outputs are stored for the subsequent optimization. The optimization of the weights is conducted using the fitness function with AUC components for all validation samples.

The classifiers have been trained on the training subset, the weights of the preferences have been optimized on the validation data, whereas the performance

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

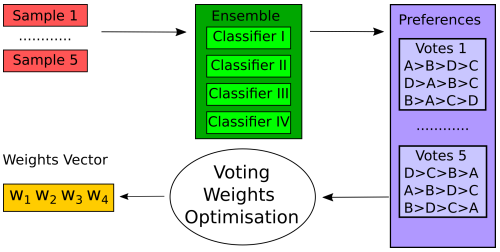

Fig. 3. Optimized positional voting.

of the optimized classifier committee has been judged on the test data. Such an approach is often called holdout validation. Figure 4 depicts data split on a toy example. As illustrated on discussed figure, a 5-fold data split has been employed for the training and optimization of the classifier committee.

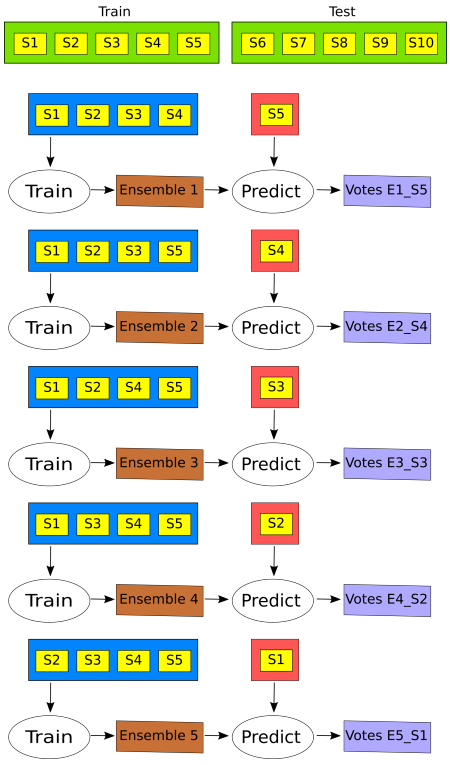

Fig. 4. Data split for training classifiers and optimizing weights of the preferences.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

3 Experimental Results

The performance of proposed algorithm has been determined on publicly available SYSU 3D Human-Object Interaction Set (SYSU 3DHOI) [19]. The dataset consists of 480 RGB-D image sequences with 12 action classes that include calling with a cell phone, playing with a cell phone, drinking, pouring, moving a chair, sitting on a chair, packing a backpack, wearing a backpack, sweeping, mopping, taking something out from the wallet and taking out a wallet. Each activity is a kind of human-object interaction. Actions were performed by forty performers. This dataset is challenging for human action recognition as a number of actions have similar motions or the same operating object at the early temporal stages. The algorithm has been evaluated in setting-1 [19] in which for each activity class, half of the samples is selected for training and the rest samples are used in testing. The evaluations were also performed in cross-subject setting, which is more challenging in comparison to the setting-1 using the same subjects for both training and testing. According to a recommendation in [19], the evaluations were done on thirty training/testing splits. Because the performers are not extracted from the background, they have been extracted by us. A window surrounding the performer has been determined on each frame and then used to crop the raw depth map. It has been then scaled to the required input shape.

Table 1 presents accuracies, precisions, recalls and F1-scores achieved by our algorithm on the SYSU 3D HOI dataset in the setting-1. In discussed configuration of the algorithm the features extracted by the Siamese neural network have been processed by shapelets algorithm to extract features representing actions, same as in [15]. In 3rd and 4th rows there are results achieved by DE optimizing the classification accuracy and DE optimizing the objective function proposed in [20], respectively. We add new results, which were achieved by classifier committee built on voting aggregation schemes: Borda count, Coombs, Bucklin, and Copeland, see rows 5–8 in Tab. 1 as well as proposed in this work: optimized positional voting and optimized positional voting operating on a subset of classifiers selected in advance. As we can observe, the optimized positional voting achieves superior results in comparison to results achieved by Borda count, Coombs, Bucklin and Copeland, c.f. results in rows 5–8 and row 9. The classification performance attained by the optimized positional voting, which operates on outputs of classifiers selected in advance is superior in comparison to results obtained by the classifier committee with classifiers selected in advance, see also experimental results in 4th row.

Table 2 presents accuracies, precisions, recalls and F1-scores achieved by our algorithm on 3D HOI dataset in the setting-1, where features representing human actions have been extracted by the DTW algorithm. Comparing results achieved by shapelets and DTW we can observe that results achieved by DTW are superior in comparison to results achieved by shapelets algorithm. Among election methods the best results have been achieved by the Copeland. The results achieved by classifier committee based on the Copeland are slightly better in comparison to results achieved by the hard voting and the soft voting, which are frequently used as aggregation techniques in the ensembles. The `DE-sel. ens.`

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

Table 1. Recognition performance on SYSU 3DHOI dataset (setting-1) using shapelets.

voting	num. class.	Accuracy	Precision	Recall	F1-score
hard voting	12	0.9167	0.9217	0.9167	0.9171
soft voting	12	0.9079	0.9102	0.9079	0.9071
DE-acc.	12	0.9079	0.9110	0.9079	0.9073
DE-sel. ens.	7	0.9254	0.9271	0.9254	0.9246
Borda count	12	0.9079	0.9141	0.9079	0.9079
Coombs	12	0.9035	0.9113	0.9035	0.9029
Bucklin	12	0.9035	0.9097	0.9035	0.9030
Copeland	12	0.9035	0.9097	0.9035	0.9030
opt. pos. voting	12	0.9167	0.9223	0.9167	0.9166
DE-sel. ens., opt. pos. voting	7	**0.9254**	**0.9300**	**0.9254**	**0.9259**

algorithm selected seven classifiers and the results achieved by classifier committee built on such a pool of the classifiers are worse in comparison to results obtained by discussed methods. The results achieved by the optimized positional voting are better in comparison to results achieved by classifier committee with `DE-acc`-based classifier selection. As we can observe, the best results have been achieved by the optimized positional voting, which operates on outputs of classifiers selected in advance.

Table 2. Recognition performance on SYSU 3DHOI dataset (setting-1) using DTW.

voting	num. class.	Accuracy	Precision	Recall	F1-score
hard voting	12	0.9341	0.9298	0.9341	0.9298
soft voting	12	0.9341	0.9298	0.9341	0.9298
DE-acc.	11	0.9266	0.9211	0.9266	0.9211
DE-sel. ens.	7	0.9340	0.9298	0.9340	0.9298
Borda count	12	0.9304	0.9254	0.9304	0.9254
Coombs	12	0.9230	0.9167	0.9230	0.9167
Bucklin	12	0.9336	0.9298	0.9336	0.9298
Copeland	12	0.9342	0.9374	0.9342	0.9341
opt. pos. voting	12	0.9386	0.9410	0.9386	0.9385
DE-sel. ens., opt. pos. voting	7	**0.9386**	**0.9414**	**0.9386**	**0.9384**

Table 3 presents accuracies, precisions, recalls and F1-scores achieved by our algorithm on SYSU 3D HOI dataset in the cross-subject setting (setting-2). As

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

in the setting-1, the features extracted by the Siamese neural network have been further processed by shapelets algorithm in order to extract features representing actions. As previously, among election methods the best results have been achieved by the Copeland. The optimized positional voting permits achieving better classification performance in comparison to performance obtained by the Copeland. Once again, the best results have been achieved through selecting the most discriminative classifiers by the DE and then executing optimized positional voting on outputs determined by such a pool of best classifiers.

Table 3. Recognition performance on SYSU 3DHOI dataset (setting-2, cross-subject) using shapelets.

voting	num. class.	Accuracy	Precision	Recall	F1-score
hard voting	12	0.8991	0.9079	0.8991	0.8990
soft voting	12	0.9035	0.9098	0.9035	0.9036
DE-acc.	2	0.9123	0.9175	0.9123	0.9119
DE-sel. ens.	2	0.9211	0.9259	0.9211	0.9209
Borda count	12	0.8904	0.8933	0.8904	0.8896
Coombs	12	0.8904	0.8925	0.8904	0.8895
Bucklin	12	0.8947	0.8967	0.8947	0.8941
Copeland	12	0.8991	0.9079	0.8991	0.8990
opt. pos. voting	12	0.9079	0.9107	0.9079	0.9077
DE-sel. ens., opt. pos. voting	2	**0.9211**	**0.9219**	**0.9211**	**0.9201**

Table 4 presents results obtained in the cross-subject setting (setting-2), where features representing actions have been extracted by the DTW algorithm. Amongst the election-based methods the best results have been achieved by the Coombs. A classifier committee built on only three best classifiers selected by the DE algorithm, achieved superior results in comparison to results discussed above. The number of the most discriminative classifiers selected by the DE optimizing the accuracy is far larger and the resulting classification performance is smaller. The classification performance achieved by the optimized positional voting is better than performance achieved by common election methods. It is smaller in comparison to performance achieved by the `DE-sel. ens.` algorithm. As we can observe, the best results have been achieved by the optimized positional voting, operating on outputs of classifiers selected in advance.

In summary, in both settings, both with shapelets and DTW-based algorithms, the best classification performances have been achieved by the optimized positional voting operating on subset of classifiers selected in advance. In all considered cases the optimized positional voting achieved better results than Borda count, Coombs, Bucklin and Copeland. The results achieved by DTW-based algorithm are better in comparison to shapelets-based algorithm [15].

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

Table 4. Recognition performance on SYSU 3DHOI dataset (setting-2, cross-subject) using DTW.

voting	num. class.	Accuracy	Precision	Recall	F1-score
hard voting	12	0.9230	0.9211	0.9230	0.9211
soft voting	12	0.9230	0.9211	0.9230	0.9211
DE-acc.	11	0.9204	0.9167	0.9204	0.9167
DE-sel. ens.	3	0.9291	0.9254	0.9291	0.9254
Borda count	12	0.9159	0.9133	0.9159	0.9123
Coombs	12	0.9202	0.9167	0.9202	0.9167
Bucklin	12	0.9152	0.9123	0.9152	0.9123
Copeland	12	0.9079	0.9098	0.9079	0.9074
opt. pos. voting	12	0.9211	0.9242	0.9211	0.9207
DE-sel. ens., opt. pos. voting	3	**0.9291**	**0.9343**	**0.9291**	**0.9295**

Table 5 presents action recognition accuracies that are achieved by recent algorithms. As we can observe, the proposed algorithm achieves superior results against all recent algorithms on challenging 3D HOI dataset. It outperforms all recent algorithms on both settings. As far as we know, on SYSU 3DHOI dataset the best classification accuracies among skeleton-based algorithms achieves a recently published SGN algorithm [21]. The proposed algorithm achieves far better classification accuracy on the discussed dataset.

Table 5. Comparative recognition performance of the proposed method with recent algorithms on 3D HOI dataset.

Method	Modality	setting	Acc. [%]
LGN [22]	skel.	II	83.33
SGN [21]	skel.	II	86.90
MSRNN [23]	depth+RGB+skel.	II	79.58
LAFF [24]	depth+RGB	II	80.00
PTS [25]	depth+skeleton	II	87.92
bidirect. rank p. [26]	depth	I	76.25
bidirect. rank p. [26]	depth	II	75.83
D3C [14]	depth	I	88.75
D3C [14]	depth	II	92.98
HAR [15]	depth	I	93.54
HAR [15]	depth	II	92.11
Proposed method	depth	I	**93.86**
Proposed method	depth	II	**92.91**

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

4 Conclusions

In this paper, we presented an approach to aggregating classifiers through positional voting techniques. The proposed optimized positional voting achieved better results in comparison to results achieved by Borda count, Coombs, Bucklin and Copeland, which have been previously used in classifier committees to combine decisions. We demonstrated experimentally that significant gains in classification performance can be obtained by executing the proposed optimized positional voting on decisions of classifiers selected in advance by the DE algorithm.

Acknowledgment. This work was supported by Polish National Science Center (NCN) under a research grant 2017/27/B/ST6/01743.

References

1. Kuncheva, L.I.: Combining Pattern Classifiers: Methods and Algorithms. Wiley, USA (2004)
2. Kuncheva, L.I., Whitaker, C.J.: Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Mach. Learn. **51**(2) (2003) 181–207
3. Wozniak, M., Grana, M., Corchado, E.: A survey of multiple classifier systems as hybrid systems. Information Fusion **16** (2014) 3–17
4. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: LightGBM: A highly efficient gradient boosting decision tree. In: Proc. of the 31st Int. Conf. on Neural Information Processing Systems. NIPS'17 (2017) 3149–3157
5. Sagi, O., Rokach, L.: Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **8** (2018)
6. Osamor, V.C., Okezie, A.F.: Enhancing the weighted voting ensemble algorithm for tuberculosis predictive diagnosis. Scientific Reports **11**(1) (Jul 2021) 14806
7. Bonab, H., Can, F.: Less is more: A comprehensive framework for the number of components of ensemble classifiers. IEEE Trans. on Neural Networks and Learning Systems **30**(9) (2019) 2735–2745
8. Polikar, R.: Ensemble based systems in decision making. IEEE Circuits and Systems Magazine **6**(3) (2006) 21–45
9. van Erp, M., Vuurpijl, L., Schomaker, L.: An overview and comparison of voting methods for pattern recognition. In: Proc. Eighth Int. Workshop on Frontiers in Handwriting Recognition. (2002) 195–200
10. Drotár, P., Gazda, M., Vokorokos, L.: Ensemble feature selection using election methods and ranker clustering. Information Sciences **480** (2019) 365–380
11. Haque, A., Milstein, A., Fei-Fei, L.: Illuminating the dark spaces of healthcare with ambient intelligence. Nature **585**(7824) (2020) 193–202
12. Wang, L., Huynh, D.Q., Koniusz, P.: A comparative review of recent Kinect-based action recognition algorithms. IEEE Trans. Image Process. **29** (2020) 15–28
13. Li, C., Huang, Q., Li, X., Wu, Q.: A multi-scale human action recognition method based on Laplacian pyramid depth motion images. In: Proc. the 2nd ACM Int. Conf. on Multimedia in Asia, ACM (2021)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64

14. Treliński, J., Kwolek, B.: Decision combination in classifier committee built on deep embedding features. In Nguyen, N.T., Iliadis, L., Maglogiannis, I., Trawiński, B., eds.: Computational Collective Intelligence, Springer (2021) 480–493
15. Treliński, J., Kwolek, B.: Human action recognition on raw depth maps. In: VCIP, IEEE (2021)
16. Wang, F., Liu, H.: Understanding the behaviour of contrastive loss. In: IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). (2021) 2495–2504
17. Saari, D.G.: Basic geometry of voting. Springer (2015)
18. Liu, M., Zhang, X., Chen, Z., Wang, X., Yang, T.: Fast stochastic AUC maximization with $o(1/n)$-convergence rate. In: Proc. of the 35th Int. Conf. on Machine Learning, PMLR (2018) 3189–3197
19. Hu, J., Zheng, W., Lai, J., Zhang, J.: Jointly learning heterogeneous features for RGB-D activity recognition. In: CVPR. (2015) 5344–5352
20. Zhou, Z.H., Wu, J.X., Jiang, Y., Chen, S.F.: Genetic algorithm based selective neural network ensemble. In: Proc. of the 17th Int. Joint Conf. on Artificial Intelligence - Volume 2. (2001) 797–802
21. Zhang, P., Lan, C., Zeng, W., Xing, J., Xue, J., Zheng, N.: Semantics-guided neural networks for efficient skeleton-based human action recognition. In: IEEE/CVF Conf. on Computer Vision and Pattern Recognition, IEEE 1109–1118
22. Ke, Q., Bennamoun, M., Rahmani, H., An, S., Sohel, F., Boussaid, F.: Learning latent global network for skeleton-based action prediction. IEEE Trans. Img. Proc. **29** (2020) 959–970
23. Hu, J., Zheng, W., Ma, L., Wang, G., Lai, J., Zhang, J.: Early action prediction by soft regression. IEEE Trans. PAMI **41**(11) (2019) 2568–2583
24. Hu, J.F., Zheng, W.S., Ma, L., Wang, G., Lai, J.: Real-time RGB-D activity prediction by soft regression. In: European Conf. on Comp. Vision, Springer (2016) 280–296
25. Wang, X., Hu, J.F., Lai, J.H., Zhang, J., Zheng, W.S.: Progressive teacher-student learning for early action prediction. In: CVPR. (2019) 3551–3560
26. Ren, Z., Zhang, Q., Gao, X., Hao, P., Cheng, J.: Multi-modality learning for human action recognition. Multimedia Tools and Applications **80**(11) (2021) 16185–16203

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_64