



A Flexible Classifier Based on Optimum Curve Fitting Approach

Övünç Polat

To cite this article: Övünç Polat (2018) A Flexible Classifier Based on Optimum Curve Fitting Approach, Applied Artificial Intelligence, 32:7-8, 660-669, DOI: [10.1080/08839514.2018.1501930](https://doi.org/10.1080/08839514.2018.1501930)

To link to this article: <https://doi.org/10.1080/08839514.2018.1501930>



Published online: 26 Jul 2018.



Submit your article to this journal [↗](#)



Article views: 222



View related articles [↗](#)



View Crossmark data [↗](#)



A Flexible Classifier Based on Optimum Curve Fitting Approach

Övünç Polat

Faculty of Engineering, Department of Electrical and Electronics Engineering, Akdeniz University, Antalya, Turkey

ABSTRACT

This study proposes a curve fitting approach for classification problems. The different classification data sets are utilized to test and evaluate the suggested method. For tested classification problems, the Gaussian curve fitting models are used. In the curve fitting stage, the number of curves equals the number of attributes in the related classification problem. For example, there are 4 attributes for iris dataset, thus four Gaussian curves are fitted for this problem. Then, output values of these fitted curves are calculated to average values, and this average value is rounded to the nearest integers. The same procedure is applied to the other dataset with having different number of features. In optimization stage, for each of classification application, the optimum values of constants of Gaussian function are determined by using genetic algorithm. For all used classification dataset, a part of the set is used during the optimization phase, and then the proposed model is validated with the remainder of the dataset. Furthermore, the optimal values of each of the attributes in tested classification application are determined by optimization algorithm. It is a valuable property of the proposed method that the accuracy of high classification can be achieved with a low number of reference data by the stage of determination of optimal feature set. Simulation results show that proposed classification approach with optimum values of constants and optimal feature set based on curve fitting has high accuracy rate. The proposed approach can be used for different classification problems.

Introduction

The pattern classification problem is an important research area due to its wide application areas. In the literature, different classifiers such as support vector machines, artificial neural networks, fuzzy classifiers and K-nearest neighbors (KNN) classifiers are used in different applications. The classifiers are used in various applications such as Electroencephalography signal classification (Zhang, Ji, and Liu et al. 2016), time series classification (Zheng et al. 2016) and text classification (Fu, Qin, and Liu 2015). There are different classifier

CONTACT Övünç Polat  ovuncpolat@akdeniz.edu.tr  Faculty of Engineering, Department of Electrical and Electronics Engineering, Akdeniz University, Antalya, Turkey

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uaai.

systems in the literature such as combining classifier using nearest decision prototypes (Kheradpisheh, Behjati-Ardakani, and Ebrahimpour 2013), an enhanced swarm intelligence clustering-based RBFNN classifier (Feng et al. 2010), nearest-neighbor classifier motivated marginal discriminant projections (Huang et al. 2011), attribute weighted Naive Bayes classifier (Taheri, Yearwood, and Mammadov et al. 2014).

This study presents an approach for classification with determination of optimal feature set and determination of optimum constants of Gaussian function based on curve fitting.

In literature, there are different applications based on curve fitting (Ahsae, Yazdi, and Naghibzadeh 2011; Gálvez and Iglesias 2013; Liu, Wang, and Cai 2010; Ryoo, Lim, and Kim 2001). In reference (Ahsae, Yazdi, and Naghibzadeh 2011), the authors present a curve fitting space approach for classification. In this referred study, the proposed curve fitting space method is based on fitting a hyperplane or curve to the learning data (Ahsae, Yazdi, and Naghibzadeh 2011). In reference (Ryoo, Lim, and Kim 2001), the researchers present a method for classify an unknown material using temperature response curve fitting and fuzzy neural network (Ryoo, Lim, and Kim 2001). In reference (Gálvez and Iglesias 2013), the authors present a new iterative mutually coupled hybrid genetic algorithm-particle swarm optimization approach for curve fitting in manufacturing (Gálvez and Iglesias 2013). In reference (Liu, Wang, and Cai 2010), the researchers present an approach for target detection in ground-penetrating radar images based on image processing and curve fitting (Liu, Wang, and Cai 2010).

In reference (Ramakrishnan and Selvan 2006), the researchers present an approach for image texture classification based on curve fitting using wavelet packet transform and singular value decomposition (Ramakrishnan and Selvan 2006). In reference (Murao, Hirao, and Hashimoto 2011), the researchers present an objective skill-level evaluation approach for Taijiquan based on curve fitting and a logarithmic distribution diagram of curvature (Murao, Hirao, and Hashimoto 2011).

In reference (Baohua, Feifang, and Liu 2001), the authors present a model that fits well learning curves for large data sets (Baohua, Feifang, and Liu 2001). In reference (Gudi and Nagaraj 2009), the researchers present an approach to optimal curve fitting of speech signal for disabled children (Gudi and Nagaraj 2009). In reference (Xue, Zhang, and Browne 2014), the researchers present a feature selection approach based on PSO for selecting a smaller number of features and acquiring even better or similar classification performance than using all features (Xue, Zhang, and Browne 2014). In reference (Polat 2015), a robust regression based classification approach was presented. In robust regression stage, the ordinary least squares analysis used for tested all datasets, and in optimization phase in this referred study,

the each of optimum attributes values in classification problem were obtained using optimization algorithm (Polat 2015).

In this paper, a classification approach with determination of optimal feature set by using curve fitting is presented for three different datasets from UCI dataset archives. The optimum values of the each of features in classification problem are determined by using genetic algorithm. In the curve fitting process, the Gaussian curve fitting model used for these datasets. Furthermore, for each application, the optimum values of constants of Gaussian function are determined by using optimization algorithm. Next section gives a curve fitting procedure. The optimization procedure of values of constants of Gaussian function is given in the third section. The optimization and determination of optimal feature set procedure are given in the fourth section. Simulations and results are given in the last section.

Curve fitting procedure

In curve fitting stage, fitting process is done with Gaussian curve fitting model for all applications, and number of curve is equal to the number of attributes in classification problem. For example, there are four attributes for iris dataset, thus four curves are fitted for this problem. Then, output values of these four fitted curves are calculated to average of arithmetic, and this average value is rounded to nearest integers. The same procedure is applied to the other classification applications with having different number of attributes. The mathematical function for Gaussian curve fitting stage given in the following equation;

$$y_k = \sum_{i=1}^n a_i e^{\left[-\left(\frac{x_k - b_i}{c_i} \right)^2 \right]},$$

where n is the number of peaks to fit, a is the amplitude, b is the centroid, c is related to the peak width (Rodrigues, Marcal and Cunha 2013). In this study, the two-term Gaussian model is used ($n = 2$). x_k is the each of attributes in classification application, where $k = 1, 2, 3, \dots$ number of attributes.

The following equation is the arithmetic mean of function outputs:

$$Y = \frac{1}{k} \sum_{i=1}^k y_i,$$

where Y is the arithmetic mean of function outputs, y_k is the value of each output for each attributes and k is the number of attributes in the related dataset. Then, this calculated Y value is rounded to nearest integers. The calculation of arithmetic mean of the outputs is used in a classifier based on

robust regression with determination of optimal feature set (Polat 2015) and curve fitting based classification (Polat 2015) approaches.

The optimization of values of constants of Gaussian function

The optimum values of a , b and c constants in Gaussian function are determined by using Genetic algorithm (Goldberg 1989). In optimization stage for all examined classification dataset, a part of the dataset is used, and then the optimized structure is validated with the remainder of the dataset. The fitness function for genetic algorithm in proposed approach is classification accuracy rate of the reference set. Figure 1 shows the outline of this stage.

Determination of optimal feature set

The optimum attributes values in related classification application are determined by using genetic algorithm. Thus, new optimal reference feature sets are acquired by using proposed approach. Nine, ten and nine new reference set values are determined for iris plant, Statlog (heart) and balance scale dataset, respectively. Figure 2 shows the outline of stage of determination of optimal feature set.

Simulation results

The classification performance of suggested approach is proved by the heart, iris plant and balance scale dataset from UCI dataset archives (Machine Learning

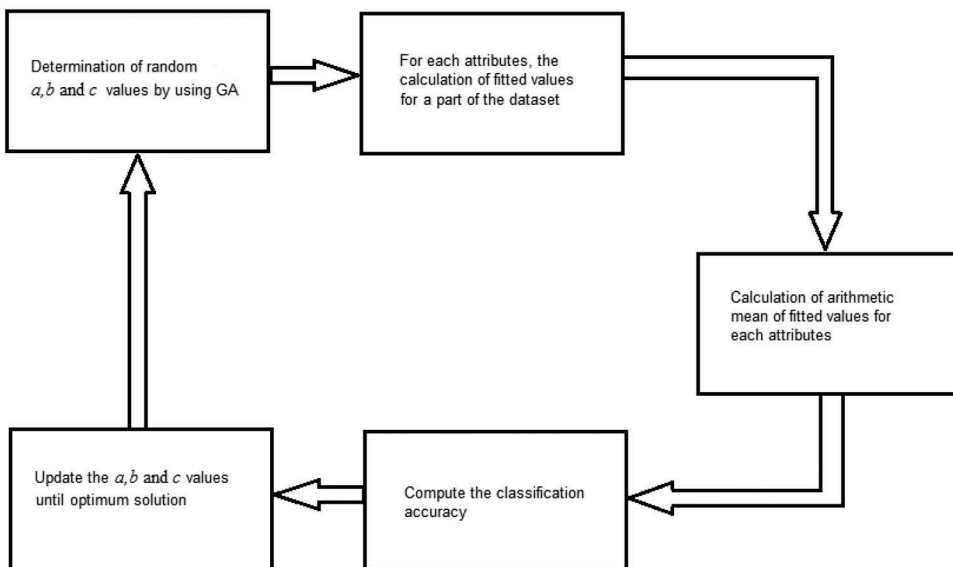


Figure 1. The outline of optimization of values of constants of Gaussian function.

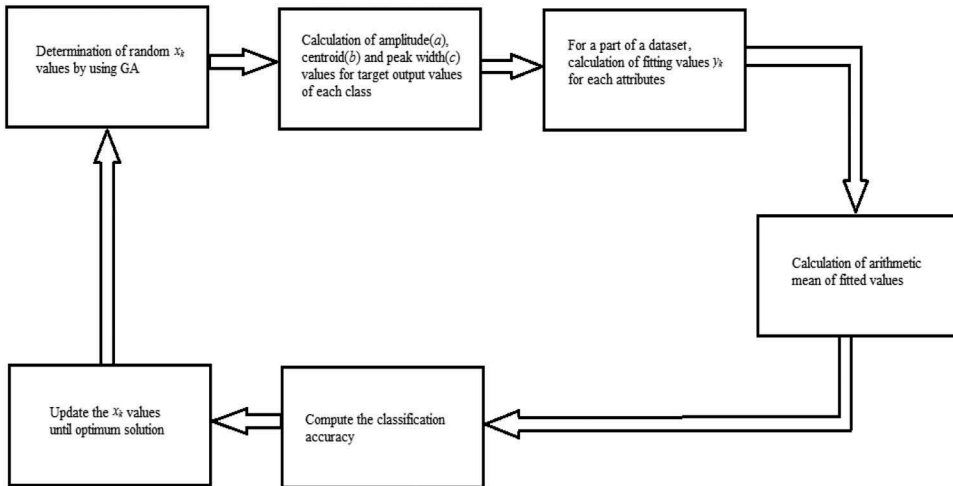


Figure 2. The outline of the determination of optimal feature set.

Repository 2016). For iris dataset, three types of iris plants are classified according to four attributes. There are totally 150 samples divided into 3 classes in this dataset. Totally 75 samples (25 instances from each class) are used in optimization process for iris plant dataset. The remaining seventy-five samples are used for validation of the optimized structure. The presence or absence of heart disease is classified with according to 13 attributes for Statlog (heart) dataset. There are totally 270 samples in this dataset. In optimization stage, totally 135 samples from this dataset are used. The remaining 135 samples are used in the validation process of optimized structure. The tested another dataset is balance scale. There are totally 625 samples from 3 classes. This dataset is classified with according to 4 attributes. In optimization process, 312 samples from this dataset are used. The remaining 313 samples are used in the validation process of optimized structure. The fitness function for optimization algorithm in suggested approach is accuracy rate of the reference set.

Simulation results for the optimization of values of constants of Gaussian function

For the optimization of values of constants of Gaussian function, the optimization variables are a , b and c constants in function. The aim of the proposed method is to obtain maximum classification accuracy. For iris dataset, 24 optimum constant values are determined. Because, the six constant values are obtained for each attributes (There are four attributes). A total of 78 optimum constant values for heart dataset and 24 optimum constant values for balance scale dataset are determined by using genetic algorithm.

Table 1. The average accuracy rates for the optimization of values of constants of Gaussian function by using proposed method and KNN.

	Iris dataset (%)	Heart dataset (%)	Balance scale dataset (%)
For reference data set by using proposed method	100	85.9	84.6
For validation set by using proposed method	94.6	83.7	93.2
KNN (For validation set)	94.6	55.5	79.2

The accuracy results of classification for three different dataset are presented in Table 1. As can be seen from Table 1, the accuracy rate is quite high for all dataset. The same datasets are classified by using KNN. The obtained results showed that proposed method better than KNN algorithm for validation set. For KNN, training set is same with reference data set in proposed method. (The K value is equal to 1 for all dataset) For tested three classification applications, high classification accuracy rate is obtained by using proposed method.

Figure 3 shows the variation of each output ($y_1 \dots y_4$) for each of attributes and the variation of the arithmetic mean of outputs. As can be seen from Figure 3 for variation of rounded output, there are only 4 samples incorrectly classified from 75 validations set samples for iris dataset.

Simulation results for determination of optimal feature

The optimization variables are each of features in classification applications for stage of determination of optimal feature. For iris and balance scale

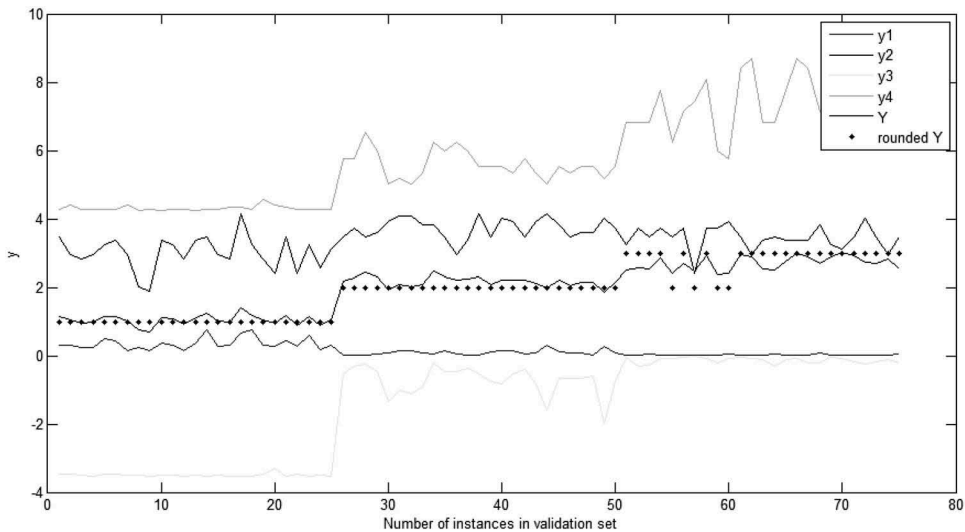


Figure 3. The variation of each individual output for each of attributes and the variation of the arithmetic mean of outputs for iris dataset in stage of the optimization of values of constants of Gaussian function.

dataset, nine optimal reference feature set values are determined (three feature set for each class). For heart dataset, 10 optimal reference feature set values are determined (five feature set for each class).

The aim of the proposed method is to obtain maximum classification accuracy with minimum reference data. The classification accuracy results for tested datasets are presented in Table 2. As can be seen from Table 2, the accuracy rate is quite high for all dataset. For KNN (shown in Table 1), there are 75 reference instances for iris dataset and 135 reference instances for heart dataset. However, nine, ten and nine optimum reference feature set are used for iris plant, heart disease and balance scale dataset, respectively in proposed approach.

Figure 4 shows the variation of each output for each of attributes and the variation of the arithmetic mean of outputs for iris dataset. As can be seen from Figure 4 for variation of rounded output, there are only 2 samples incorrectly classified from 75 validation set.

Table 2. The average classification accuracy rates by using proposed method for stage of determination of optimal feature.

	Iris dataset (%)	Heart dataset (%)	Balance dataset (%)
For a part of the dataset in optimization stage by using proposed method with optimal reference feature set	97.3%	80.7%	63.7
For validation set by using proposed method with optimal reference feature set	97.3	80.0	71.1

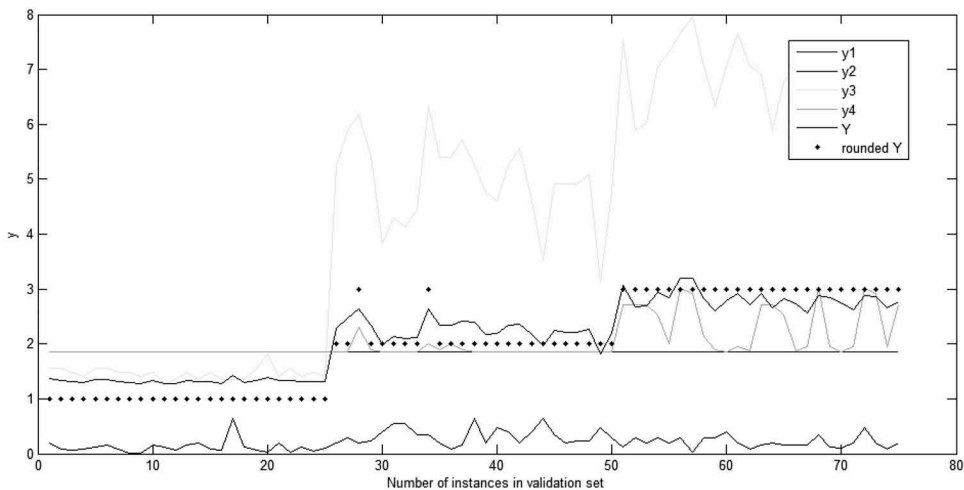


Figure 4. The variation of each individual output for each of attributes and the variation of the arithmetic mean of outputs for iris dataset.

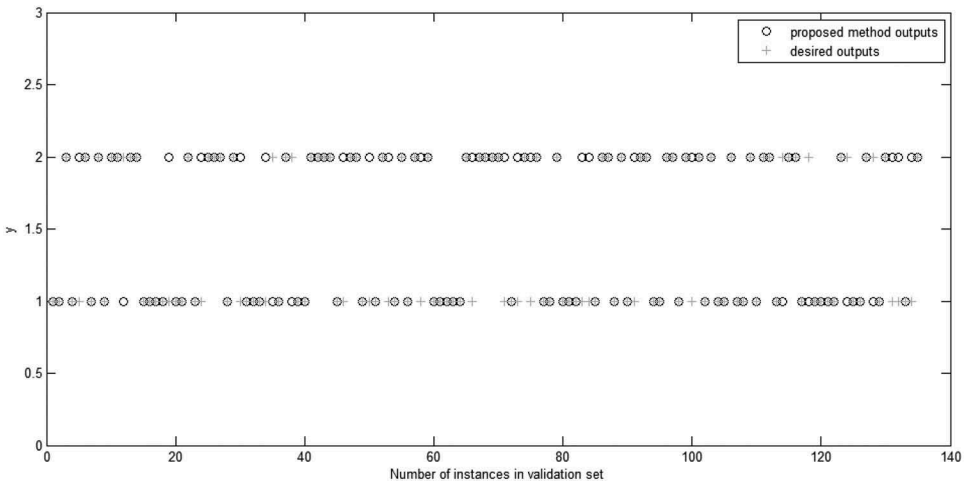


Figure 5. The variation of obtained output values and the variation of desired output values for heart dataset.

Figure 5 shows the variation of obtained outputs (class codes) using proposed method and desired output values (desired class codes). As can be seen from Figure 5, there are only 27 (number of “+” symbol) samples incorrectly classified from 135 validation set samples for heart dataset.

Conclusions

In this study, a classifier is designed based on Gaussian curve fitting model with determination of optimal constant values of curve function and determination of optimal feature set values. The genetic algorithms are used in order to determine optimal values. The proposed model is carried out for three different datasets and high classification accuracy rate is obtained for all applications. The application of classification of heart and balance scale dataset, the higher accuracy is obtained by using determination of optimal constant values of curve function than curve fitting model with determination of optimal feature set values.

Simulation results show that classification by proposed method improves the accuracy rate considerably in comparison to KNN. The proposed model can be used for other classification problems. The different curve fitting models can be used in order to increase the accuracy. The ability of classification with fewer reference data is the valuable property of designed approach with determination of optimal feature set values.

Conflict of interest

I certify that there is no actual or potential conflict of interest in relation to this article.

Funding

The research has been supported by the Research Project Department of Akdeniz University, Antalya, Turkey.

References

- Ahsaee, M. G., H. S. Yazdi, and M. Naghibzadeh. 2011. Curve fitting space for classification. *Neural Computing and Applications* 20:273–85. doi:10.1007/s00521-010-0383-7.
- Baohua, G., H. Feifang, and H. Liu. Modelling classification performance for large data sets (An Empirical Study). In: Proceedings of the Second International Conference on Advances in Web-Age Information Management; 9–11 July 2001; China: pp.317–28.
- Feng, Y., W. Zhongfu, J. Zhong, Y. Chunxiao, and W. Kaigui. 2010. An enhanced swarm intelligence clustering-based RBFNN classifier and its application in deep Web sources classification. *Frontiers of Computer Science in China* 4 (4):560–70. doi:10.1007/s11704-010-0104-5.
- Fu, R., B. Qin, and T. Liu. 2015. Open-categorical text classification based on multi-lda models. *Soft Computation* 190 (1):29–38. doi:10.1007/s00500-014-1374-x.
- Gálvez, A., and A. Iglesias. 2013. A new iterative mutually coupled hybrid GA–PSO approach for curve fitting in manufacturing. *Applied Soft Computing* 13 (3):1491–504. doi:10.1016/j.asoc.2012.05.030.
- Goldberg, D. E. 1989. Genetic algorithm in search, optimization, and machine learning. Addison-Wesley.
- Gudi, A. B., and H. C. Nagaraj. 2009. Optimal curve fitting of speech signal for disabled children. *International Journal of Computer Science & Information Technology* 1 (2):99–107.
- Huang, P., Z. Tang, C. Chen, and X. Cheng. 2011. Nearest-neighbor classifier motivated marginal discriminant projections for face recognition. *Frontiers of Computer Science in China* 5 (4):419–28. doi:10.1007/s11704-011-1012-z.
- Kheradpisheh, S. R., F. Behjati-Ardakani, and R. Ebrahimpour. December 2013. Combining classifiers using nearest decision prototypes. *Applied Soft Computing* 13(12):Pages4570–4578. doi: 10.1016/j.asoc.2013.07.028.
- Liu, Y., M. Wang, and Q. Cai. The target detection for GPR images based on curve fitting. In: 3rd International Congress on Image and Signal Processing; 16–18 Oct. 2010; Yantai: pp. 2876–79.
- Machine Learning Repository (2016). Center for machine learning and intelligent systems. Retrieved from: <http://archive.ics.uci.edu/ml/>
- Murao, T., Y. Hirao, and H. Hashimoto. 2011. Skill level evaluation for Taijiquan based on curve fitting and logarithmic distribution diagram of curvature. *SICE Journal of Control, Measurement, and System Integration* 4 (1):001–005.
- Polat, Ö. 2015. A robust regression based classifier with determination of optimal feature set. *Journal of Applied Research and Technology* 13:443–46. doi:10.1016/j.jart.2015.08.001.
- Polat, Ö., The curve fitting approach for classification problems. 15th Industrial Conference on Data Mining, Poster Proceedings, pp. 44–49, July 15–19, 2015, Hamburg, Germany.
- Ramakrishnan, S., and S. Selvan. Image texture classification using exponential curve fitting of wavelet domain singular values, In: Proceedings of IEE 3rd International Conference on Visual Information Engineering; 26–28 September 2006; Bangalore: pp.505–10.

- Rodrigues, A., Andre, R., Marcal, and Cunha, M. 2013 April. "Monitoring vegetation dynamics inferred by satellite data using the phenosat tool,". In *Ieee Transactions on Geoscience and Remote Sensing* 51(4):2096-2104. doi:[10.1109/TGRS.2012.2223475](https://doi.org/10.1109/TGRS.2012.2223475).
- Ryoo, Y. J., Y. C. Lim, and K. H. Kim. 2001. Classification of materials using temperature response curve fitting and fuzzy neural network. *Sensors and Actuators A: Physical* 94 (1-2):11-18. doi:[10.1016/S0924-4247\(01\)00681-1](https://doi.org/10.1016/S0924-4247(01)00681-1).
- Taheri, S., J. Yearwood, M. Mammadov, et al. 2014. Attribute weighted Naive Bayes classifier using a local optimization. *Neural Computation & Application* 24:995. doi:[10.1007/s00521-012-1329-z](https://doi.org/10.1007/s00521-012-1329-z).
- Xue, B., M. Zhang, and W. N. Browne. 2014. Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Applied Soft Computing* 18:261-76. doi:[10.1016/j.asoc.2013.09.018](https://doi.org/10.1016/j.asoc.2013.09.018).
- Zhang, Y., X. Ji, B. Liu, et al. 2016. Combined feature extraction method for classification of EEG signals. *Neural Computation & Application*. doi:[10.1007/s00521-016-2230-y](https://doi.org/10.1007/s00521-016-2230-y).
- Zheng, Y., Q. Liu, E. Chen, Y. Ge, and J. L. Zhao. 2016. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science* 10 (1):96-112. doi:[10.1007/s11704-015-4478-2](https://doi.org/10.1007/s11704-015-4478-2).