



Multiple classifiers fusion and CNN feature extraction for handwritten digits recognition

Hui-huang Zhao^{1,2} · Han Liu³

Received: 29 December 2018 / Accepted: 16 February 2019 / Published online: 22 February 2019
© The Author(s) 2019

Abstract

Handwritten digits recognition has been treated as a multi-class classification problem in the machine learning context, where each of the ten digits (0–9) is viewed as a class and the machine learning task is essentially to train a classifier that can effectively discriminate the ten classes. In practice, it is very usual that the performance of a single classifier trained using a standard learning algorithm is varied on different datasets, which indicates that the same learning algorithm may train strong classifiers on some datasets but weak classifiers may be trained on other datasets. It is also possible that the same classifier shows different performance on different test sets, especially when considering the case that image instances can be highly diverse due to the different handwriting styles of different people on the same digits. To address the above issue, development of ensemble learning approaches have been very necessary to improve the overall performance and make the performance more stable on different datasets. In this paper, we propose a framework that involves CNN-based feature extraction from the MINST dataset and algebraic fusion of multiple classifiers trained on different feature sets, which are prepared through feature selection applied to the original feature set extracted using CNN. The experimental results show that the classifiers fusion can achieve the classification accuracy of $\geq 98\%$.

Keywords Machine learning · Ensemble learning · Classification · Classifiers fusion · Random forests · Granular computing

1 Introduction

Handwriting digits recognition refers to the process of transforming the ordered trajectory generated by writing on handwriting equipment into the internal code of digits. It is actually a mapping process from the coordinate sequence of handwritten trajectory to the internal code of digits. It is one of the most natural and convenient means of human–computer interaction. With the popularity of mobile information

tools such as smartphones and handheld computers, handwritten digits recognition technology has entered the era of large-scale application. Handwritten digits recognition enables users to input text in the most natural and convenient way. It is easy to learn and use, and can replace keyboards or mouses. There are many kinds of devices for handwriting inputs, such as electromagnetic induction handwriting boards, pressure-sensitive hand-writing boards, touch screens, touch panels, ultrasonic pens, etc. Handwriting digits recognition belongs to the category of digits recognition and pattern recognition. In terms of the recognition process, digits recognition can be divided into two categories: off-line recognition and on-line recognition. In terms of recognition objects, it can also be divided into two categories: handwriting digits recognition and print digits recognition.

Also, it is well known that the handwritten digits recognition is a challenging problem. In recent years, there are many algorithms proposed for handwritten digits recognition. Boukharouba (2017) develops a new feature extraction technique for handwritten digit recognition based on support vector machines (SVM). During this method, the vertical

✉ Han Liu
LiuH48@cardiff.ac.uk
Hui-huang Zhao
happyday.huihuang@gmail.com

¹ College of Computer Science and technology, Hengyang Normal University, Hengyang 421008, China

² Hunan Provincial Key Laboratory of Intelligent Information Processing and Application, Hengyang 421008, China

³ School of Computer Science and Informatics, Cardiff University, Queen's Buildings, 5 The Parade, Cardiff CF24 3AA, UK

and horizontal directions of a digit image are combined with the famous Freeman chain code, and the approach does not require any normalization of digits. Mohebi and Bagirov (2014) presented a convolutional recursive modified self-organizing maps (SOM) and applied it to handwritten digits recognition. The results have shown that the proposed method can lead to an improvement of the recognition rate compared with other SOM-based algorithms.

In the machine learning context, it is commonly known that each standard learning algorithm usually shows different performance on different datasets. In other words, the use of an algorithm may lead to the production of strong classifiers on some datasets but the classifiers trained on other datasets using the same algorithm may be much weaker. In the case of handwritten digits recognition, a standard learning algorithm may be capable of learning some but not all specific characteristics of handwritten digits. Also, the same classifier may show different performance on different datasets, due to the different data distribution. In addition, instances of handwritten digits usually show very diverse characteristics due to different handwriting styles of different people, even if the instances belong to the same class (Ding et al. 2018).

To address the above issue, in this paper, we propose to adopt instance-based recognition of handwritten digits in the setting of ensemble learning, towards obtaining diverse classifiers trained using different learning algorithms. The whole procedure of recognition involves using convolutional neural network (CNN) for feature extraction, adopting a correlation-based feature subset selection method for obtaining diverse feature sets and setting multi-level fusion of classifiers trained on different feature sets.

The main contributions of this paper include: (1) the use of CNN to extract more diverse features from each handwritten digit image and different feature sets are prepared through filter-based feature selection; (2) an ensemble learning framework is proposed, which involves multi-level fusion of multiple classifiers trained on different feature sets using different learning algorithms.

The rest of this paper is organized as follows. In Sect. 2, we introduce some related work in this context. We describe the proposed approach in Sect. 3. Experimental results are presented in Sect. 4 and conclusion are shown in Sect. 5.

2 Related work

This section provides a review of the applications of convolutional neural networks for image classification, an overview of handwritten digits recognition and a review of traditional machine learning methods alongside potential improvements through the use of granular computing concepts.

2.1 Convolutional neural network

In machine learning, a convolutional neural network (CNN, or ConvNet) is a class of deep feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery (Plamondon and Srihari 2000). CNN is considered as an excellent tool for solving computer vision problems in a large number of fields. CNNs are widely used in modern AI systems but also bring challenges. For example, during the IEEE conference on computer vision and pattern recognition 2018, there are more than 10 papers that are based on CNN, e.g. Hui et al. (2018) proposed a state-of-the-art CNN named LiteFlowNet which can be used to improve flow estimation accuracy. To train a deep convolutional neural network with both low-precision weights and low-bit width activation, Zhuang et al. (2018) proposed to use a two-stage optimization strategy to improve its performance. Feng et al. (2018b) proposed a new loss function named ‘Wing loss’ for robust facial landmark localisation with CNNs. During 3D shape recognition, Feng et al. (2018a) proposed a GVCNN (group-view convolutional neural network) framework which can achieve a significant performance gain on the 3D shape recognition. And Zhang et al. (2018) proposed a new knowledge-based semisupervised deep CNN for Facial action unit intensity estimation, and it can achieve comparable or even better performance than some common methods with smaller datasets. Beside those, CNN is also used in energy-efficient reconfigurable accelerator (Chen et al. 2017), semantic image segmentation (Chen et al. 2018) and image fusion (Acharya et al. 2017a, b). CNNs involve relatively little pre-processing compared with other image classification algorithms. This independence from prior knowledge and human effort in feature design is a major advantage.

Convolutional architectures also seem to benefit extracting features from image data. In our approach, the image features of handwritten digits are extracted using the Convolutional Neural Network architecture.

2.2 Review of machine learning methods

There are many machine learning algorithms which are used in image recognition, and the most popular ones in machine learning mainly include multi-layer perceptron (MLP) (Mirjalili 2015), random forests (Biau et al. 2009), K nearest neighbour (KNN) (Vermeulen et al. 2017), Naive Bayes (NB) (Amor et al. 2004) and C4.5 decision tree (Quinlan 1996). Also, the above machine learning algorithms have been used popularly in handwritten digits recognition tasks.

A multilayer perceptron (MLP) network is a type of feed-forward artificial neural networks and it consists of

one or more fully connected layers. There are at least three layers in a MLP network (the input layer, a hidden layer and the output layer) (Yilmaz and Özer 2009). MLP does not specify the number of hidden layers, so it can choose the appropriate number of hidden layers according to their needs. There is no limit on the number of neurons in the output layer. It utilizes a kind of supervised learning techniques (Ravi et al. 2017).

Random forest can be understood as Cart tree forest, which is an integrated learning mode composed of multiple Cart tree classifiers. Among them, each Cart tree can be understood as a member, which trains a part of randomly put back from the sample set. In this way, multiple tree classifiers constitute a training model matrix. Then the samples to be classified are brought into this tree classifier, and the final classification of this sample is decided by the majority voting rule (Wager and Athey 2017). Random forest can easily identify the importance of each feature, but if there is a strong relationship between features A and B, that is to say, B can be deduced from A, then the importance of such feature is meaningless, because random forest often only gives a high value to A, and B will be much smaller (Scornet et al. 2015).

The K-nearest neighbor (KNN) algorithm is a famous statistical method for pattern recognition and occupies a considerable position in machine learning based classification algorithms. It is one of the simplest machine learning algorithms (Song et al. 2016). KNN is one of the most basic instance-based learning methods and one of the best text classification algorithms. The basic idea is that if the majority of the K instances closest to an unseen instance in the feature space (the nearest neighbor in the feature space) belong to a category, the instance also belongs to that category. The selected neighbors are instances that have been correctly classified (Zhang et al. 2017). A disadvantage of KNN is that it requires a large amount of calculation, because the distance between each instance to be classified and all known instances must be calculated to obtain its K nearest neighbors.

Bayes theorem is a very old statistical method (1763). Naive Bayes (NB) is a classification method based on Bayes theorem and independent hypothesis of characteristic conditions (Chen and Jahanshahi 2018). The step is to learn the joint probability distribution of input/output based on the “independent hypothesis of characteristic conditions”. According to this model, the output y with the maximum posterior probability is calculated by Bayesian theorem for the input. Naive Bayes (NB) has the advantages of simple implementation and high prediction efficiency. It can be used for large databases; The downside is that the prior probabilities have to be known (Amor et al. 2004).

C4.5 decision tree learning is an algorithm developed by Ross Quinlan to generate decision trees (Polat and Gne

2009). This algorithm is an extension of the ID3 algorithm developed by Ross Quinlan. The decision tree generated by the C4.5 algorithm can be used in classification problems of machine learning and data mining. Its goal is supervised learning: given a dataset, each tuple in it can be described by a set of attribute values, each of which belongs to a category from mutually exclusive ones. The goal of C4.5 is to learn to find a mapping relationship from attribute values to categories that can be used to classify new entities of unknown categories (Sathyadevan and Nair 2015).

3 Proposed framework

In this section, we provide a description of CNN based feature extraction and present an ensemble learning framework that involves multi-level fusion of multiple classifiers trained on different feature sets using different learning algorithms. We also justify how the design of the proposed framework involves the application of granular computing concepts.

3.1 CNN feature extraction

During our method, we use CNNLeNet-5 to obtain more diverse features from each handwritten digit image.

The proposed CNN feature extraction for handwritten digit images by LeNet-5 is illustrated in Fig. 1.

The LeNet architecture is considered as the first architecture for convolutional neural networks. We can easily see from the LeNet-5 in Fig. 1 that many feature maps are generated in each layer. So we can obtain more diverse features than using other common methods.

The LeNet-5 is an excellent architecture for handwritten digit recognition. The LeNet-5 has two parts, one is feature extraction, whereas the other one is classification which is used to classify objects. During our approach, we do not use the LeNet-5 to do classification (blue part in Fig. 1), and we only use it to extract features from images. During the classification, we use the proposed ensemble learning framework instead of a neural network that consists of fully connected layers.

Given an image of $32 \times 32 \times 1$, firstly, a convolution layer with six 5×5 filters with the stride of 1 is used and an output matrix of $28 \times 28 \times 6$ is generated. With the stride of 1 and no padding, the feature map is reduced from 32×32 to 28×28 . Then average pooling with the filter width of 2 and the stride of 2 is taken and the dimension is reduced by the factor of 2 and ends up with $14 \times 14 \times 6$. Furthermore, another convolution layer with sixteen 5×5 filters is used leading to an output matrix of $10 \times 10 \times 16$. Then another pooling layer is involved and ends up with an output matrix of $5 \times 5 \times 16$. Therefore, we extract sixteen 5×5 feature maps from each image, and each feature map (5×5)

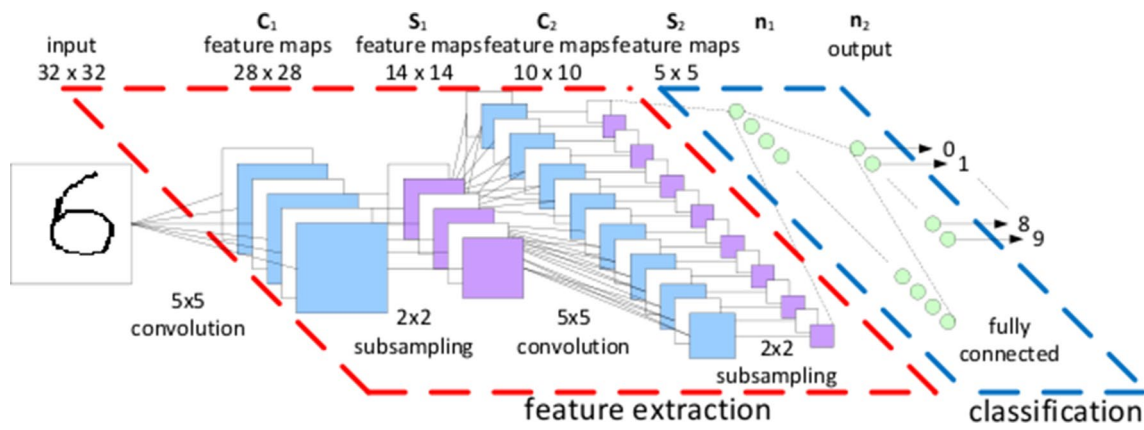


Fig. 1 CNN feature extraction for the handwritten digit image

is treated as a column vector (25×1). Overall, there are two convolution layers, two subsampling layers, and two fully connected layers in the LeNet-5.

3.2 Multi-level fusion of classifiers

The proposed ensemble learning framework involves multiple levels of fusion of diverse classifiers trained on different feature sets. The entire procedure of the proposed framework is illustrated in Fig. 2.

In particular, as shown in the feature preparation layer in Fig. 2, different feature sets can be prepared through feature extraction using different methods (but we only obtain one feature set extracted using CNN in this paper). Also, the

feature set extracted using a specific method can be further processed to obtain different feature subsets using different feature selection methods. In the third layer, about training of multiple classifiers, m learning algorithms are used to train base classifiers on each feature set F_i , therefore, a primary ensemble E_i is created on each of the n feature sets as shown in the primary fusion layer. Finally, the n primary ensembles created on the n feature sets are fused further to create the final ensemble, so that a final classification is made as the output of the final ensemble as shown in the final fusion layer.

In practice, the setting of ensemble learning can be achieved even in a more flexible way than the one shown in Fig. 2. For example, some base classifiers trained on a

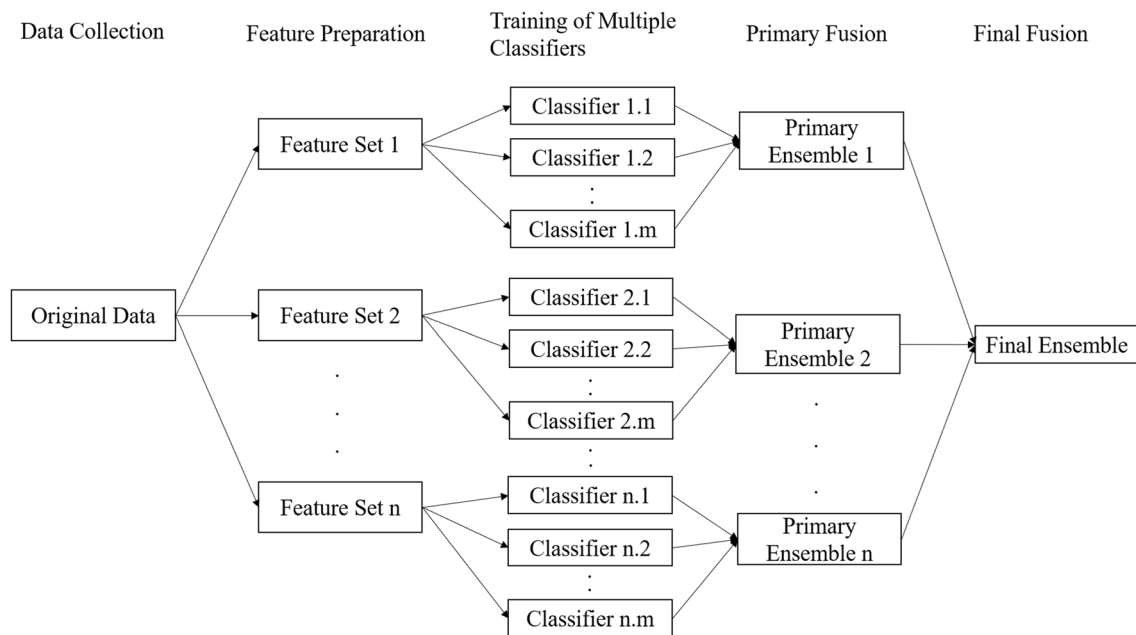


Fig. 2 Proposed framework of ensemble learning

feature set can be combined to make up a primary ensemble, which is combined further with the other base classifiers to make up a secondary ensemble. In this context, a secondary ensemble can be created on each feature set and some or all of the secondary ensembles can be fused further to make up a higher level ensemble or even the final ensemble. We will show this kind of setting of ensemble learning in Sect. 4.

The proposed ensemble learning framework is essentially designed in the setting of granular computing, which is a formalized paradigm of information processing (Pedrycz 2011; Pedrycz and Chen 2011, 2015a, b). In general, granular computing can be considered as a method of structural thinking at the philosophical level but can also be used as a strategy of structural problem solving at the practical level (Yao 2005b).

In theory, two main concepts of granular computing are referred to as granule and granularity (Liu and Cocea 2017, 2018; Liu et al. 2018). Granule is defined as a collection of smaller particles that can form a larger unit. In the context of ensemble learning, each ensemble can be viewed as a granule since it consists of multiple classifiers. While granules can be of very different sizes, the concept of granularity becomes highly needed to deal with the different sizes of different granules, that is, to involve different granules in different levels of granularity, according to the scale of their actual sizes. The proposed ensemble learning framework involves multiple levels of classifiers fusion, where each of the levels can be viewed as a specific level of granularity. In this context, a primary ensemble that only consists of base classifiers is viewed as a granule at the basic (bottom) level of granularity, whereas the final ensemble that may involve both base classifiers and lower level ensembles is viewed as a granule at the top level of granularity.

In practice, granular computing concepts are commonly used through taking one or both of the two operations, namely granulation and organization. The former operation is essentially decomposition of a whole into multiple parts in a top-down information processing manner, such as extraction of local features through the convolution layer of CNN, whereas the latter operation is essentially integration of multiple parts into a whole in a bottom-up information processing manner (Yao 2005a), such as fusion of multiple classifiers.

4 Experimental results and discussion

In this section, we report an experimental study conducted on the MNIST dataset, which is essentially a 10-class (0–9) classification task in the setting of machine learning.

The MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training

various image processing systems (Niu and Suen 2012). The database consists of a training set of 60,000 images and a test set of 10,000 images.

In this experimental study, the whole procedure involves feature extraction, feature selection and training and fusion of classifiers. During the CNN feature extraction, we input each digit image to the LeNet-5, and output its feature maps in the 3rd layer ($16 \times 10 \times 10$). Furthermore, those feature maps are changed into a single column. In terms of the setting of the CNN architecture, the activation function is set as sigmoid, and the loss function is set as a mean squared error, and the optimization function is set as l2 regularizer. In addition, the input Batch size is set as 1. There are 60,000 images in the MNIST and the Epoch is 1, therefore, there are 60,000 iterations in total.

In the feature selection stage, we apply the correlation-based feature subset selection method (Hall and Smith 1997) to obtain a reduced set of features. In this way, we use the reduced set of selected features alongside the original feature set extracted using CNN, such that diversity can be created through training classifiers on the two different feature sets.

In the classifiers training and fusion stage, we adopt KNN and RF for training base classifiers and primary ensembles, respectively, on the two feature sets. On each feature set, a secondary ensemble is obtained through combining the base classifier (trained using KNN) and the primary ensemble of decision trees (created using RF). The two primary ensembles created on the two feature sets are combined further to make up a larger ensemble for final fusion. The whole setting of the ensemble creation on each feature set is illustrated in Fig. 3.

In terms of parameters setting, the K value for KNN is set to 3 and the trained random forest consists of 100 decision trees. The KNN and RF classifiers are fused through averaging their hidden outputs (probability for each class), i.e. the mean rule of algebraic fusion (Zhou 2012). All the experiments are conducted using 10-fold cross-validation.

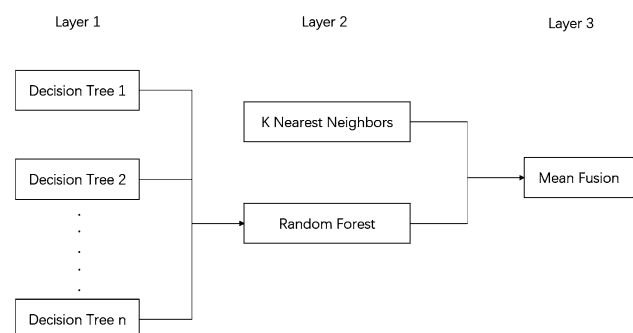


Fig. 3 Fusion of K nearest neighbor and random forest (Zhao and Liu 2018)

Table 1 Classification accuracy

Method	Full set of features	Selected set of features
KNN (Zhang 1992)	0.972	0.958
RF (Breiman 2001)	0.965	0.957
Fusion (KNN + RF) (Zhao and Liu 2018)	0.975	0.965
Final fusion (proposed)	0.981	

The results on the MINST dataset is shown in Table 1 in terms of classification accuracy. The results indicate that the nature of the KNN method through instance-based learning leads to the accuracy of $\geq 95.8\%$ on the two feature sets. Also, the RF method is generally very capable of training highly diverse decision tree classifiers on different training samples and feature subsets, which leads to the accuracy of $\geq 95.7\%$ on the two feature sets.

On the above basis, the further fusion of the base classifier (trained using KNN) and the decision tree (primary) ensemble (created using RF) leads to an improvement of the classification performance on each feature set, which indicates that the different learning strategies between the KNN and RF methods can really result in diversity between their trained classifiers. The final fusion of the above two secondary ensembles created on the two feature sets leads to a further improvement of the classification performance.

In addition, although feature selection may not necessarily lead to advances in the classification performance for each single classifier trained on the reduced feature set in comparison with using the full feature set, the fusion of classifiers trained on the two feature sets can lead to an improvement, which would indicate that the preparation of different feature sets through feature selection can effectively lead to the creation of diversity among classifiers trained on the different feature sets.

Overall, the experimental results suggest that multilevel fusion of classifiers through various ways of diversity creation is encouraged towards advances in the classification performance in a layer-by-layer manner.

5 Conclusions

In this paper, we have proposed a framework that involves CNN based feature extraction and multi-level fusion of diverse classifiers. In particular, we have designed to increase the diversity among classifiers through preparing different feature sets and using different learning algorithms for classifiers training. The experimental results show that our proposed ensemble approach can achieve the classification accuracy of $\geq 98\%$ using the MNIST dataset and

the results also indicate that the setting of ensemble learning which aims to train diverse classifiers is very useful to advance the overall performance of classification.

In future, we will investigate how to achieve optimal feature subsets selection to boost the performance further through using some optimization techniques (Chen and Chung 2006; Chen and Chien 2011; Chen and Kao 2013; Tsai et al. 2008, 2012). It is also worth to explore the effectiveness of the proposed framework in the setting of fuzzy ensemble learning (Nakai et al. 2003), where fuzzy set theory related techniques (Zadeh 1965; Wang and Chen 2008; Chen et al. 2009, 2012, 2013; Chen and Chen 2001, 2011; Chen and Tanuwijaya 2011; Chen and Chang 2011; Liu and Zhang 2018) are adopted to train base classifiers as the members of an ensemble (Liu and Chen 2018). Also, it is worth to investigate the effectiveness of adopting the proposed framework of ensemble learning in the context of multi-attribute decision-making (Xu and Wang 2016; Liu and You 2017; Chatterjee and Kar 2017; Lee and Chen 2008; Zulueta-Veliz and Garca-Cabrera 2018).

Acknowledgements This work was supported by National Natural Science Foundation of China (61503128), Science and Technology Plan Project of Hunan Province (2016TP102), Scientific Research Fund of Hunan Provincial Education Department (16C0226), Hengyang guided science and technology projects and Application-oriented Special Disciplines (Hengkefa [2018]60-31), Double First-Class University Project of Hunan Province(Xiangjiaotong [2018]469), Hunan Province Special Funds of Central Government for Guiding Local Science and Technology Development (2018CT5001) and Subject Group Construction Project of Hengyang Normal University(18XKQ02). The authors would also like to acknowledge support from the School of Computer Science and Informatics at the Cardiff University.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M (2017a) Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network. *Inf Sci* 405:81–90
- Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H (2017b) Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals. *Comput Biol Med* 100:270–278
- Amor NB, Benferhat S, Elouedi Z (2004) Naive bayes vs decision trees in intrusion detection systems. In: *ACM symposium on applied computing*, pp 420–424
- Biau G, Devroye L, Lugosi G (2009) Consistency of random forests and other averaging classifiers. *J Mach Learn Res* 9(1):2015–2033

- Boukharouba ABA (2017) Novel feature extraction technique for the recognition of handwritten digits. *Appl Comput Inform* 13(1):19–26
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Chatterjee K, Kar S (2017) Unified granular-number-based AHP-VIKOR multi-criteria decision framework. *Granul Comput* 2(3):199–221
- Chen SM, Chang YC (2011) Weighted fuzzy rule interpolation based on ga-based weight-learning techniques. *IEEE Trans Fuzzy Syst* 19(4):729–744
- Chen SJ, Chen SM (2001) A new method to measure the similarity between fuzzy numbers. In: *IEEE international conference on fuzzy systems*, Melbourne, pp 1123–1126
- Chen SM, Chen CD (2011) Handling forecasting problems based on high-order fuzzy logical relationships. *Expert Syst Appl* 38(4):3857–3864
- Chen SM, Chien CY (2011) Parallelized genetic ant colony systems for solving the traveling salesman problem. *Expert Syst Appl* 38(4):3873–3883
- Chen SM, Chung NY (2006) Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *Int J Inf Manag Sci* 17(3):1–17
- Chen FC, Jahanshahi RMR (2018) NB-CNN: deep learning-based crack detection using convolutional neural network and naive bayes data fusion. *IEEE Trans Ind Electron* 65(5):4392–4400
- Chen SM, Kao PY (2013) Taix forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines. *Inf Sci* 247:62–71
- Chen SM, Tanuwijaya K (2011) Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques. *Expert Syst Appl* 38(12):15,425–415,437
- Chen SM, Wang NY, Pan JS (2009) Forecasting enrollments using automatic clustering techniques and fuzzy logical relationships. *Expert Syst Appl* 36(8):11,070–011,076
- Chen SM, Munif A, Chen GS, Liu HC, Kuo BC (2012) Fuzzy risk analysis based on ranking generalized fuzzy numbers with different left heights and right heights. *Expert Syst Appl* 39(7):6320–6334
- Chen SM, Chang YC, Pan JS (2013) Fuzzy rules interpolation for sparse fuzzy rule-based systems based on interval type-2 gaussian fuzzy sets and genetic algorithms. *IEEE Trans Fuzzy Syst* 21(3):412–425
- Chen YH, Krishna T, Emer JS, Sze V (2017) Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks. *IEEE J Solid State Circuits* 52:127–138
- Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2018) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Mach Intell* 40(4):834–848
- Ding W, Wang X, Liu H, Hu B (2018) An empirical study of shape recognition in ensemble learning context. In: *International conference on wavelet analysis and pattern recognition*, Chengdu, pp 256–261
- Feng Y, Zhang Z, Zhao X, Ji R, Gao Y (2018a) GVCNN: group-view convolutional neural networks for 3d shape recognition. In: *The IEEE conference on computer vision and pattern recognition (CVPR)*
- Feng ZH, Kittler J, Awais M, Huber P, Wu XJ (2018b) Wing loss for robust facial landmark localisation with convolutional neural networks. In: *The IEEE conference on computer vision and pattern recognition (CVPR)*
- Hall MA, Smith LA (1997) Feature subset selection: a correlation based filter approach. In: *Proceedings of the 1997 international conference on neural information processing and intelligent information systems*. Springer, Berlin, pp 855–858
- Hui TW, Tang X, Loy CC (2018) Liteflownet: a lightweight convolutional neural network for optical flow estimation. In: *The IEEE conference on computer vision and pattern recognition (CVPR)*
- Lee LW, Chen SM (2008) Fuzzy multiple attributes group decision-making based on the extension of TOPSIS method and interval type-2 fuzzy sets. In: *Proceedings of the 2008 international conference on machine learning and cybernetics*, Kunming, vol 6, pp 3260–3265
- Liu H, Chen SM (2018) Multi-level fusion of classifiers through fuzzy ensemble learning. In: *International symposium on computational intelligence and design*, Hangzhou, pp 19–22
- Liu H, Cocea M (2017) Fuzzy information granulation towards interpretable sentiment analysis. *Granul Comput* 2(4):289–302
- Liu H, Cocea M (2018) *Granular computing based machine learning: a big data processing approach*. Springer, Berlin
- Liu P, You X (2017) Probabilistic linguistic TODIM approach for multiple attribute decision-making. *Granul Comput* 2(4):332–342
- Liu H, Zhang L (2018) Fuzzy rule-based systems for recognition intensive classification in granular computing context. *Granul Comput* 3(4):355–365
- Liu H, Cocea M, Ding W (2018) Multi-task learning for intelligent data processing in granular computing context. *Granul Comput* 3(3):257–273
- Mirjalili S (2015) How effective is the Grey Wolf optimizer in training multi-layer perceptrons. *Appl Intell* 43(1): 150–161. Springer, New York
- Mohebi E, Bagirov A (2014) A convolutional recursive modified self organizing map for handwritten digits recognition. *Neural Netw* 60(C):104–118
- Nakai G, Nakashima T, Ishibuchi H (2003) A fuzzy ensemble learning method for pattern classification. *J Jpn Soc Fuzzy Theory Intell Inform* 15(6):671–681
- Niu XX, Suen CY (2012) A novel hybrid CNN–SVM classifier for recognizing handwritten digits. *Pattern Recogn* 45(4):1318–1325
- Pedrycz W (2011) Information granules and their use in schemes of knowledge management. *Sci Iran* 18(3):602–610
- Pedrycz W, Chen SM (2011) *Granular computing and intelligent systems: design with information granules of higher order and higher type*. Springer, Heidelberg
- Pedrycz W, Chen SM (2015a) *Granular computing and decision-making: interactive and iterative approaches*. Springer, Heidelberg
- Pedrycz W, Chen SM (2015b) *Information granularity, big data, and computational intelligence*. Springer, Heidelberg
- Plamondon R, Srihari SN (2000) Online and off-line handwriting recognition: a comprehensive survey. *IEEE Trans Pattern Anal Mach Intell* 22(1):63–84
- Polat K, Gne S (2009) A novel hybrid intelligent method based on c4.5 decision tree classifier and one-against-all approach for multi-class classification problems. *Expert Syst Appl* 36(2):1587–1592
- Quinlan JR (1996) *Improved use of continuous attributes in C4.5*. AI Access Foundation, USA
- Ravi V, Pradeepkumar D, Deb K (2017) Financial time series prediction using hybrids of chaos theory, multilayer perceptron and multi-objective evolutionary algorithms. *Swarm Evolut Comput* 36:136–149
- Sathyadevan S, Nair RR (2015) *Comparative analysis of decision tree algorithms: ID3, C4.5 and random forest*. Springer, New Delhi
- Scornet E, Biau G, Vert JP (2015) Consistency of random forests. *Ann Stat* 43(4):1716–1741
- Song G, Rochas J, Beze L, Huet F, Magoules F (2016) K nearest neighbour joins for big data on mapreduce: a theoretical and experimental analysis. *IEEE Trans Knowl Data Eng* 28(9):2376–2392
- Tsai PW, Pan JS, Chen SM, Liao BY, Hao SP (2008) Parallel cat swarm optimization. In: *Proceedings of the 2008 international*

- conference on machine learning and cybernetics, Kunming, vol 6, pp 3328–3333
- Tsai PW, Pan JS, Chen SM, Liao BY (2012) Enhanced parallel cat swarm optimization based on the Taguchi method. *Expert Syst Appl* 39(7):6309–6319
- Vermeulen JL, Hillebrand A, Geraerts R (2017) A comparative study of k-nearest neighbour techniques in crowd simulation. *Comput Anim Virtual Worlds* 28(3–4):e1775
- Wager S, Athey S (2017) Estimation and inference of heterogeneous treatment effects using random forests. *Res Pap* 8(6):1831–1845
- Wang HY, Chen SM (2008) Evaluating students' answerscripts using fuzzy numbers associated with degrees of confidence. *IEEE Trans Fuzzy Syst* 16(2):403–415
- Xu Z, Wang H (2016) Managing multi-granularity linguistic information in qualitative group decision making: an overview. *Granul Comput* 1(1):21–35
- Yao J (2005a) Information granulation and granular relationships. In: *IEEE International Conference on Granular Computing*, Beijing, China, pp 326–329
- Yao Y (2005b) Perspectives of granular computing. In: *Proceedings of 2005 IEEE International Conference on Granular Computing*, Beijing, China, pp 85–90
- Yilmaz AS, Özer Z (2009) Pitch angle control in wind turbines above the rated wind speed by multi-layer perceptron and radial basis function neural networks. *Expert Syst Appl* 36(6):9767–9775
- Zadeh L (1965) Fuzzy sets. *Inf Control* 8(3):338–353
- Zhang J (1992) Selecting typical instances in instance-based learning. In: *Proceedings of the 9th international workshop on machine learning*, Aberdeen, pp 470–479
- Zhang X, Li Y, Kotagiri R, Wu L, Tari Z, Cheriet M (2017) KRNN: k rare-class nearest neighbour classification. *Pattern Recognit* 62:33–44
- Zhao H, Liu H (2018) Algebraic fusion of multiple classifiers for handwritten digits recognition. In: *International conference on wavelet analysis and pattern recognition*, Chengdu, pp 250–255
- Zhang Y, Dong W, Hu BG, Ji Q (2018) Weakly-supervised deep convolutional neural network learning for facial action unit intensity estimation. In: *The IEEE conference on computer vision and pattern recognition (CVPR)*
- Zhou ZH (2012) *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC, London
- Zhuang B, Shen C, Tan M, Liu L, Reid I (2018) Towards effective low-bitwidth convolutional neural networks. In: *The IEEE conference on computer vision and pattern recognition (CVPR)*
- Zulueta-Veliz Y, Garca-Cabrera L (2018) A choquet integral-based approach to multiattribute decision making with correlated periods. *Granul Comput* 3(3):245–256

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.