*Special Issue: Simulation of Digital Image Processing in Medical Applications*

*Simulation*

*Simulation: Transactions of the Society for Modeling and Simulation International*
2016, Vol. 92(9) 861–871

DOI: 10.1177/0037549716666962
sim.sagepub.com
SAGE SCS

# Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, *k*-nearest neighbors, and support vector machine

**Yudong Zhang[1,2], Siyuan Lu[1,3], Xingxing Zhou[1,4], Ming Yang[5], Lenan Wu[6], Bin Liu[7], Preetha Phillips[8] and Shuihua Wang[1,9]**

**Abstract**

In order to detect multiple sclerosis (MS) subjects from healthy controls (HCs) in magnetic resonance imaging, we developed a new system based on machine learning. The MS imaging data was downloaded from the eHealth laboratory at the University of Cyprus, and the HC imaging data was scanned in our local hospital with volunteers enrolled from community advertisement. Inter-scan normalization was employed to remove the gray-level difference. We adjust the misclassification costs to alleviate the effect of unbalanced class distribution to the classification performance. We utilized two-level stationary wavelet entropy (SWE) to extract features from brain images. Then, we compared three machine learning based classifiers: the decision tree, *k*-nearest neighbors (kNN), and support vector machine. The experimental results showed the kNN performed the best among all three classifiers. In addition, the proposed SWE + kNN approach is superior to four state-of-the-art approaches. Our proposed MS detection approach is effective.

**Keywords**

Multiple sclerosis, stationary wavelet entropy, decision tree, *k*-nearest neighbors, support vector machine, machine learning

## 1 Introduction

Multiple sclerosis (MS) is a chronic demyelinating disease, caused by damage to the insulating cover of nerve cells in central neural system.[1–4] The symptoms are different within individuals, mainly involving physical disability[5] and mental problems.[6] Specifically, the patients may be afflicted with vision disturbance,[7] retinal damage,[8] impaired color vision,[9] muscle weakness,[10] fatigue, depression,[11] etc.

With the conventional magnetic resonance imaging (MRI) technique it is difficult to detect severe tissue damage, due to the normal-appearing white matter (NAWM),[12–14] that is, the abnormal white matter (WM) areas appear normal. This clinic-radiological paradox is a challenge for neuroradiologists. There is a need for novel techniques that may offer a better diagnosis rate in monitoring MS than human interpretation.

Considering that computer vision (CV) performs better than human eyes in many fields, such as face recognition,

[1]School of Computer Science and Technology, Nanjing Normal University, China
[2]Hunan Provincial Key Laboratory of Network Investigational Technology, Hunan Policy Academy, China
[3]State Key Lab of CAD & CG, Zhejiang University, China
[4]Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, China
[5]Department of Radiology, Nanjing Children's Hospital, Nanjing Medical University, China
[6]School of Information Science and Engineering, Southeast University, China
[7]Department of Radiology, Zhong-Da Hospital of Southeast University, China
[8]School of Natural Sciences and Mathematics, Shepherd University, USA
[9]Department of Electrical Engineering, The City College of New York, USA

**Corresponding author:**
S Wang, School of Computer Science and Technology, Nanjing Normal University, 1 Wenyuan, Nanjing, Jiangsu 210023, China.
Email: wangshuihua@njnu.edu.cn

video security, and satellite surveillance, scholars have applied CV in detecting and monitoring the progress of MS. Murray et al.[15] developed a texture analysis system for MS. They extracted image features using the multiscale amplitude-modulation frequency-modulation (MAMFM) method. They used the support vector machine (SVM) as a classifier. Siddiqui et al.[16] developed an automated and intelligent medical decision support system for brain scan classification. They employed discrete wavelet transform (DWT), principal component analysis (PCA), and the least-square support vector machine (LS-SVM). Phillips et al.[17] proposed using wavelet entropy (WE) for pathological brain detection. They also proposed a **H**ybridization of **B**iogeography-based optimization and **P**article swarm optimization (HBP) for classifier training. Khotanlou and Afrasiabi[18] used the genetic algorithm (GA) to select effective features in MS lesions. The similarity index (SI) of the SVM classifier is determined as the fitness function of the GA. Tachinaga et al.[19] developed a computer-aided diagnostic system for detecting MS regions by tri-linear interpolation (TLI), *k*-means clustering (KMC), and the SVM. Deshpande et al.[20] used adaptive dictionary learning to detect lesions of MS. Nayak et al.[21] combined DWT, probabilistic PCA (PPCA), and the random forest (RF) for brain image classification.

WE is a relatively new texture analysis method that has been already used in neuroimaging for not only disease characterization and quantification, but also subtle signal intensity variations. Saritha et al.[22] combined WE with a spider-web plot to classify MR brain images. Phillips et al.[17] used WE and HBP in order to detect pathological brains. Zhou et al.[23] used WE and the feature selection method, and they detected abnormal brain images with an accuracy of 100.00%.

Nevertheless, WE calculates entropy values over each subband obtained by DWT. Du et al.[24] pointed out that DWT is of translational variance. Therefore, it is natural to replace DWT with stationary wavelet transform (SWT), which is of translational invariance. This means that even if the brain image was not registered correctly (maybe with one pixel error), the DWT coefficients may change significantly while SWT coefficients remain unchanged. This overwhelming advantage makes SWT outperform DWT in many fields. The combination of SWT and entropy is the so-called stationary wavelet entropy (SWE).

Our primary objective here is to detect MS in the brain by the use of SWE. In addition, we used three classifiers in the field of machine learning, for the aim of classification. Those three classifiers were the decision tree (DT), *k*-nearest-neighbor method, and SVM. The remainder of this paper is organized as follows: Section 2 describes the materials; Section 3 presents the methodology used; Section 4 gives the results; Section 5 discusses the results; Section 6 concludes this study and gives future directions.

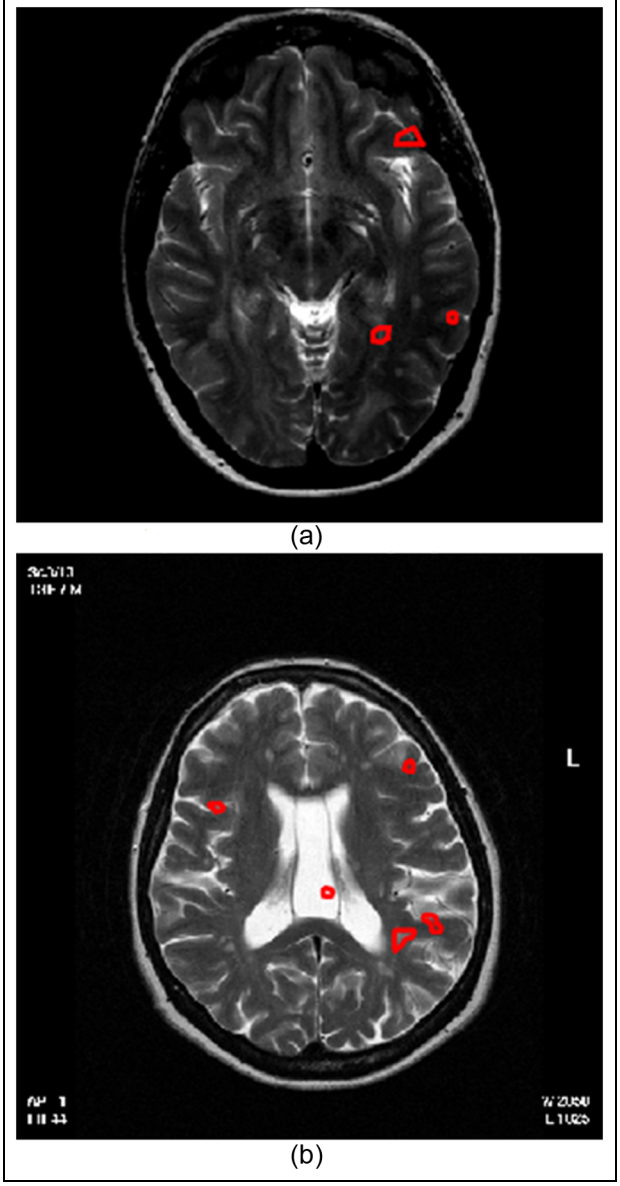


**Figure 1.** Illustration of multiple sclerosis patients: (a) a slice with three plaques; (b) a slice with five plaques (areas surrounded by red lines denote the plaque; color online only).

## 2 Materials

### 2.1 Source

The images used in this study come from two sources. The MS images were downloaded from the eHealth laboratory at the University of Cyprus (http://www.medinfo.cs.ucy.ac.cy/index.php/downloads/datasets). There are 38 patients (aged 34.1 $\pm$ 10.5 years, 17 males and 21 females) in the dataset. All brain lesions were identified by experienced MS neurologists, and were confirmed by radiologists. We selected the slices that were associated with plaques, and we obtained 676 slices in total. Figure 1 shows two slices with three and five plaques, respectively.

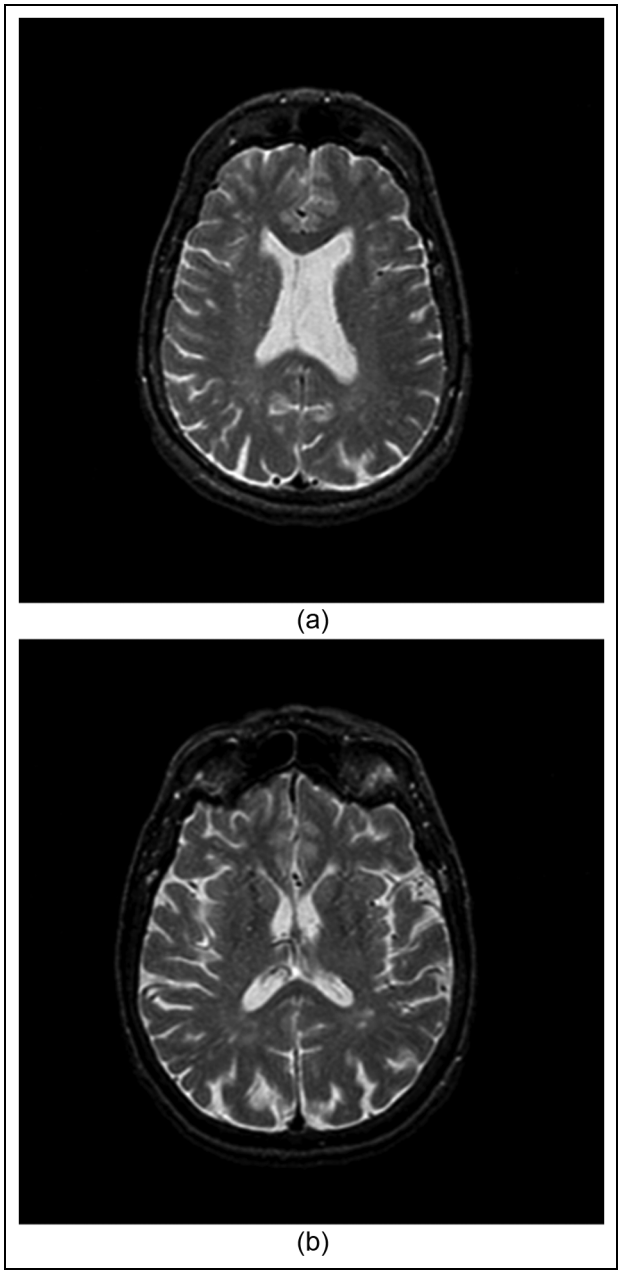


**Figure 2.** Illustrations of healthy controls.

For comparison, we enroll healthy controls (HCs) in the same range of age and gender distribution. The exclusion criteria for all volunteers were known neurological or psychiatric diseases, brain lesions, taking psychotropic medications, and contraindications to MR imaging. Figure 2 illustrates two pictures of HCs. Our study was approved by the Ethics Committee of Southeast University, and a signed informed consent form was obtained from every subject prior to entering this study. The brain extraction tool was not performed, since the MS images preserve scalps.

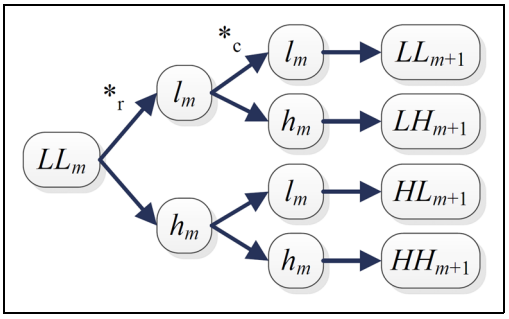


**Figure 3.** Block diagram of stationary wavelet transform. (*l* and *h* represent the low-pass and the high-pass filters, respectively, *m* represents the decomposition level, *L* and *H* represent the low-pass and high-pass result, respectively, $*_r$ represents the row-wise filter and $*_c$ represents the column-wise filter.)

### 2.2 Inter-scan normalization

Since the sources of the dataset are from different scanning machines, we need to match their image intensities so as to facilitate brain image comparability. Histogram stretching (HS)[25] was employed to increase the dynamic range of the original brain images by the following equation[26]:

$$h(x,y) = \frac{f(x,y) - f_{\min}}{f_{\max} - f_{\min}} \tag{1}$$

where $f$ represents the original brain image, $h$ is the intensity normalized image, $(x, y)$ are the coordinates of the pixel, and $f_{\min}$ and $f_{\max}$ represent the minimum and maximum intensity values, respectively.

## 3 Methodology

### 3.1 Stationary wavelet entropy

Nguyen et al.[27] developed the concept of SWE. Du et al.[24] applied SWE in pathological brain detection. In this study, we aimed to apply SWE in MS detection. SWE consists of two major processes. Firstly, it carries out SWT on a given image.

As Figure 3 shows, $h$ represents the high-pass filter (HPF) and $l$ the low-pass filter (LPF); the $m$-level SWT decomposition is implemented through bypassing all subbands through $h$ and $l$. The filters are recursively dilated at scale $2^m$, and then $2^{m-1}$ zeros are inserted among these coefficients.[28–30] Mathematically, the four subbands *LL, LH, HL*, and *HH* obtained as follows:

$$LL_{m+1} = (LL_m) *_r (g_m) *_c (g_m) \tag{2}$$

$$LH_{m+1} = (LL_m) *_r (l_m) *_c (h_m) \tag{3}$$

$$HL_{m+1} = (LL_m) *_r (h_m) *_c (l_m) \tag{4}$$

$$HH_{m+1} = (LL_m)*_r(h_m)*_c(h_m) \quad (5)$$

where $*_r$ represents the row-wise filter and $*_c$ represents the column-wise filter, since the rows and columns are filtered independently. The coefficients of different levels present multiscale, multi-resolution, and translational-invariant subbands.

In the next process, the randomness of the subbands is measured by the Shannon entropy $A$, defined as follows:

$$A = -\sum_i X_i^2 \log X_i^2 \quad (6)$$

Here $A$ represents the entropy and $X_i$ represents the $i$th element of a given subband. Note that for an $m$-level decomposition, there are in total $(3m + 1)$ subbands, and thus a vector of $(3m + 1)$ elements was formed.

## 3.2 Classification

We compared three classifiers in this study. Firstly, the DT was employed. The DT is a tree-like structure.[31] Each node represents a test over an attribute, each branch denotes its outcome, and each leaf node denotes a class label.[32–34] The path from the root node to the leaf node can be seen as a classification rule. The goal of the learning method for the DT is to create a DT model that predicts the value of a target class label on the basis of the input attributes.[35] C4.5 was used, which is an extension of the earlier ID3 algorithm.[36] C4.5 chooses at each node the attribute that most effectively splits the samples into subsets that are enriched in one class or the other. Here information gain was chosen as the splitting criterion. We chose the attribute with the largest normalized information gain to make the test. Then, C4.5 recurred on the smaller sub-lists. We set the maximum number of splits to four.

The second algorithm we used is the $k$-nearest neighbors (kNN) method. The input contains $k$ closet training data. The output is a class membership, by which a new instance is assigned to the class that is most common by a majority vote of its $k$ neighbors.[37–39] Figure 4 shows an illustration of kNN, where the new instance (blue circle) should be classified as the yellow square class, because there are three yellow squares and two green diamonds in the dashed circle.

kNN is a type of lazy learning, that is, all computation is deferred until the new-instance classification stage.[40] In this binary classification problem, it is beneficial to set $k$ to an odd number, so as to avoid tied votes.[41,42] The bootstrap method was used to set the optimal value of $k$. Euclidean distance was employed to measure the distance.

The last classifier is the SVM. It generates a hyperplane to achieve a desired separation.[43–45] It is commonly known that a larger margin will provide a lower generalization error.[46–48] In practical situations where linearly separable data are impossible, the soft margin concept was introduced so that the separation is as clear as possible.[49] Quadratic programming (QP) was used as the learning method.

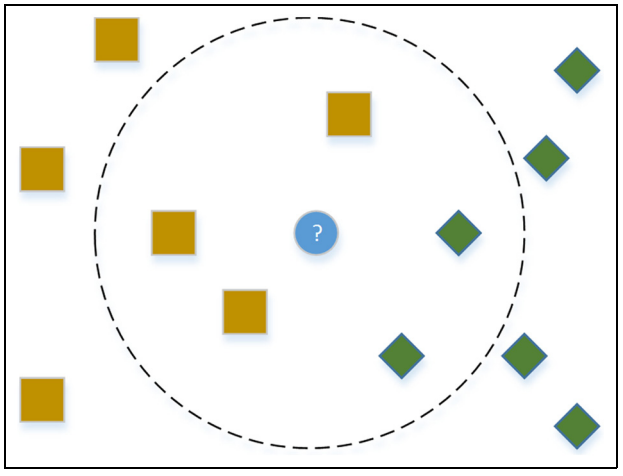


**Figure 4.** Illustration of $k$-nearest neighbors. (Color online only.)

The unbalanced class distribution is due to the difficulty of acquiring MS data and the ease of acquiring healthy brain data. To solve the effect of unbalanced class distribution to the classification performance, we adjust the misclassification costs.[50]

## 3.3 Statistical experiment

All the classifications were performed on a 10-fold cross-validation in order to avoid overfitting and achieve an out-of-sample estimation. We used 10-fold since it is the most commonly used. The flowchart of 10-fold cross-validation is shown in Figure 5. Therefore, the classification performances do not depend on the training data. All the programs were developed using the platform of MATLAB with ''wavelet design & analysis'' and ''classification learner'' apps.

Following general rules, the positive means were identified and the negative means rejected. Hence, the true positive (TP) means a MS patient is correctly identified as MS, the false positive (FP) means healthy people were incorrectly identified as MS, the true negative (TN) means healthy people were correctly identified as healthy, and the false negative (FN) means MS patients incorrectly identified as healthy. Four measures are used as follows:

$$\text{Sensitivity} = TP/(TP + FN) \quad (7)$$

$$\text{Specificity} = TN/(TN + FP) \quad (8)$$

$$\text{Precision} = TP/(TP + FP) \quad (9)$$

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN) \quad (10)$$

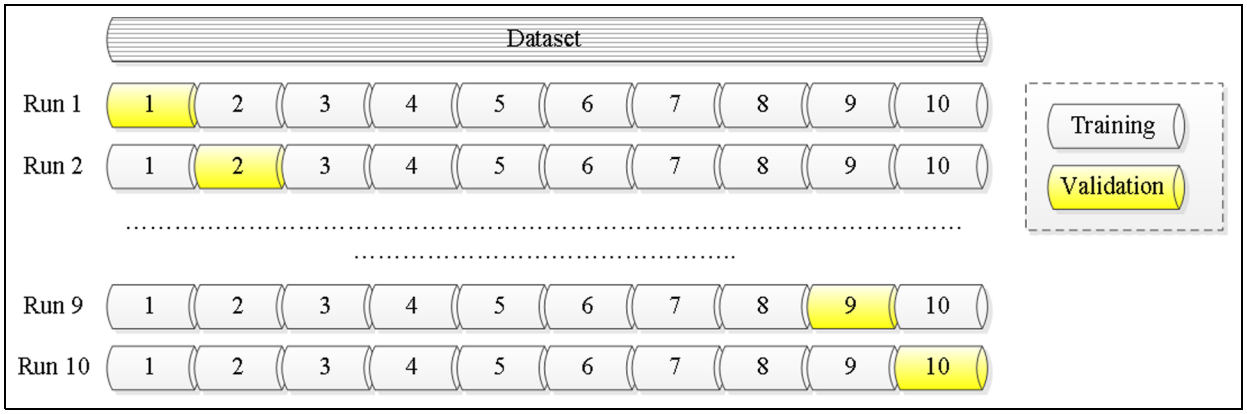


**Figure 5.** Flowchart of 10-fold cross-validation.

**Table 1.** Confusion matrix comparison.

| Method | Confusion matrix |
|---|---|
| DT | $\begin{bmatrix} 654 & 22 \\ 15 & 865 \end{bmatrix}$ |
| kNN | $\begin{bmatrix} 650 & 26 \\ 6 & 874 \end{bmatrix}$ |
| SVM | $\begin{bmatrix} 658 & 18 \\ 20 & 860 \end{bmatrix}$ |

DT: decision tree; kNN: *k*-nearest neighbors; SVM: support vector machine.

**Table 2.** Evaluation comparison (unit: %).

| Method | Sensitivity | Specificity | Precision | Accuracy |
|---|---|---|---|---|
| DT | 96.75 | 98.30 | 97.76 | 97.62 |
| kNN | 96.15 | **99.32** | **99.09** | **97.94** |
| SVM | **97.34** | 97.73 | 97.05 | 97.56 |

Bold means the best.
DT: decision tree; kNN: *k*-nearest neighbors; SVM: support vector machine.

## 4 Results

### 4.1 Stationary wavelet transform

Figure 6 shows the two-dimensional discrete wavelet transform (2D-DWT) result of a MS image. Here Figure 6(a) shows the original brain image and Figure 6(b) presents the DWT results. In contrast, Figure 7 shows the 2D-SWT result of a MS image. Figure 7(a) shows the original MS brain image. Figures 7(b)–(h) illustrate the seven subbands of two-level SWT.

### 4.2 Classifier comparison

The confusion matrix of DT, kNN, and SVM are listed in Table 1. Their performances measured by sensitivity, specificity, precision, and accuracy are shown in Table 2. Those four measures were defined in Section 3.3.

### 4.3 Comparison to state-of-the-art methods

In this experiment, we compared SWE with other state-of-the-art feature extraction methods, such as MAMFM + SVM,[15] DWT + PCA + LS-SVM,[16] WE + HBP,[17] and DWT + PPCA + RF.[21] Again, 10-fold cross-validation was used. The comparative analysis is listed in Table 3.

Table 3 shows that MAMFM + SVM[15] achieves a sensitivity of 94.08%, a specificity of 93.64%, a precision of 91.91%, and an accuracy of 93.83%. The DWT + PCA + LS-SVM method[16] achieves a sensitivity of 95.86%, a specificity of 96.48%, a precision of 95.43%, and an accuracy of 96.21%. The WE + HBP method[17] achieves a sensitivity of 96.15%, a specificity of 97.16%, a precision of 96.30%, and an accuracy of 96.72%. The DWT + PPCA + RF method[21] achieves a sensitivity of 96.01%, a specificity of 96.70%, a precision of 95.72%, and an accuracy of 96.40%. Finally, the proposed SWE + kNN method achieves a sensitivity of 96.15%, a specificity of 99.32%, a precision of 99.09%, and an accuracy of 97.94%.

### 4.4 Computation time

The computation time is also an important measure to evaluate different classification models. Our computation was implemented on a laptop with 2.20 GHz Intel® Core™ i3-2330M CPU and 8 GB RAM. The operating system was 64-bit Windows 7 Ultimate with service pack 1. The ''generate code'' button in the panel of the

**Table 3.** Comparative results (unit: %).

| Method | Sensitivity | Specificity | Precision | Accuracy |
|---|---|---|---|---|
| MAMFM + SVM[15] | 94.08 | 93.64 | 91.91 | 93.83 |
| DWT + PCA + LS-SVM[16] | 95.86 | 96.48 | 95.43 | 96.21 |
| WE + HBP[17] | **96.15** | 97.16 | 96.30 | 96.72 |
| DWT + PPCA + RF[21] | 96.01 | 96.70 | 95.72 | 96.40 |
| SWE + kNN (proposed) | **96.15** | **99.32** | **99.09** | **97.94** |

MAMFM: multiscale amplitude modulation frequency-modulation; SVM: support vector machine; DWT: discrete wavelet transform; PCA: principal component analysis; LS: least square; WE: wavelet entropy; HBP: hybridization of biogeography-based optimization and particle swarm optimization; PPCA: probabilistic PCA; RF: random forest; SWE: stationary wavelet entropy; kNN: *k*-nearest neighbors.

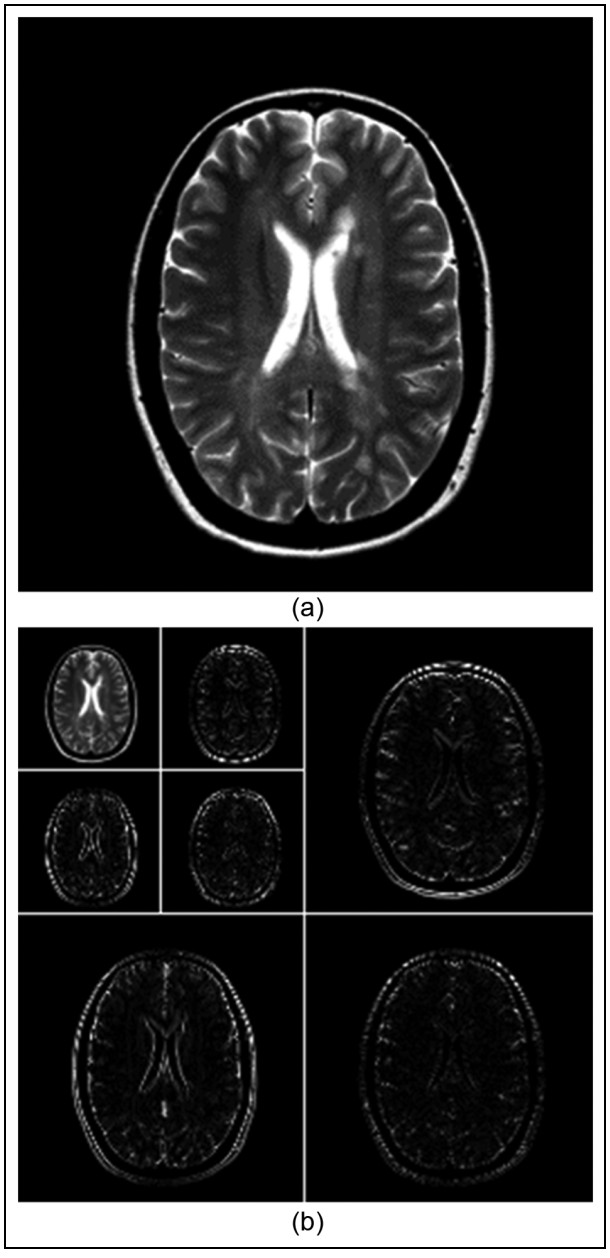


**Figure 6.** Two-dimensional discrete wavelet transform of a multiple sclerosis image: (a) original Image; (b) DWT result.

''classification learner'' app was clicked to automatically generate the MATLAB code for training classifiers and validation predictions. Two commands, ''tic'' and ''toc'', were employed to start and read elapsed time from a stopwatch timer. The results are illustrated in Figure 8. The computation times of the DT, kNN, and SVM are 0.7847 $\pm$ 0.1216, 1.1613 $\pm$ 0.2566, and 1.3031 $\pm$ 0.2788 seconds (mean $\pm$ standard deviation), respectively.

## 5 Discussion

After comparing Figures 6 and 7, we observed that SWT could extract more features than DWT, which is the main cogent reason why SWT performs better than DWT. Thus, every pixel in the seven bands represents an individual feature. To reduce the number of features, entropy was calculated over each subband. Finally, seven entropy values were obtained for each brain image.

The results in Tables 1 and 2 show clearly that the DT achieves a sensitivity of 96.75%, a specificity of 98.30%, a precision of 97.76%, and an accuracy of 97.62%. kNN achieves a sensitivity of 96.15%, a specificity of 99.32%, a precision of 99.09%, and an accuracy of 97.94%. The SVM achieves a sensitivity of 97.34%, a specificity of 97.73%, a precision of 97.05%, and an accuracy of 97.56%. In all, kNN performs the best in terms of specificity, precision, and accuracy, while the SVM performs the best in sensitivity. The DT performs the worst in all four measures. Thus, we can conclude that kNN yields the best classification performance among the three methods. This is because SWE can obtain a very low-dimensional feature space for the imaging data, and kNN is most suitable for processing this kind of data.[51,52] In this study, we used Euclidean distance; nevertheless, there are some other distance measures, for instance, the hamming distance, Pearson coefficient, Spearman coefficient, etc. Thus, we shall try to develop an automatic method to select the optimal distance measure.

From Table 3, Murray et al.[15] used MAMFM + SVM to detect lesions in MS images; in our experiment we used the same method, ''MAMFM + SVM,'' to detect MS brain images from healthy brain images. Thus, the results

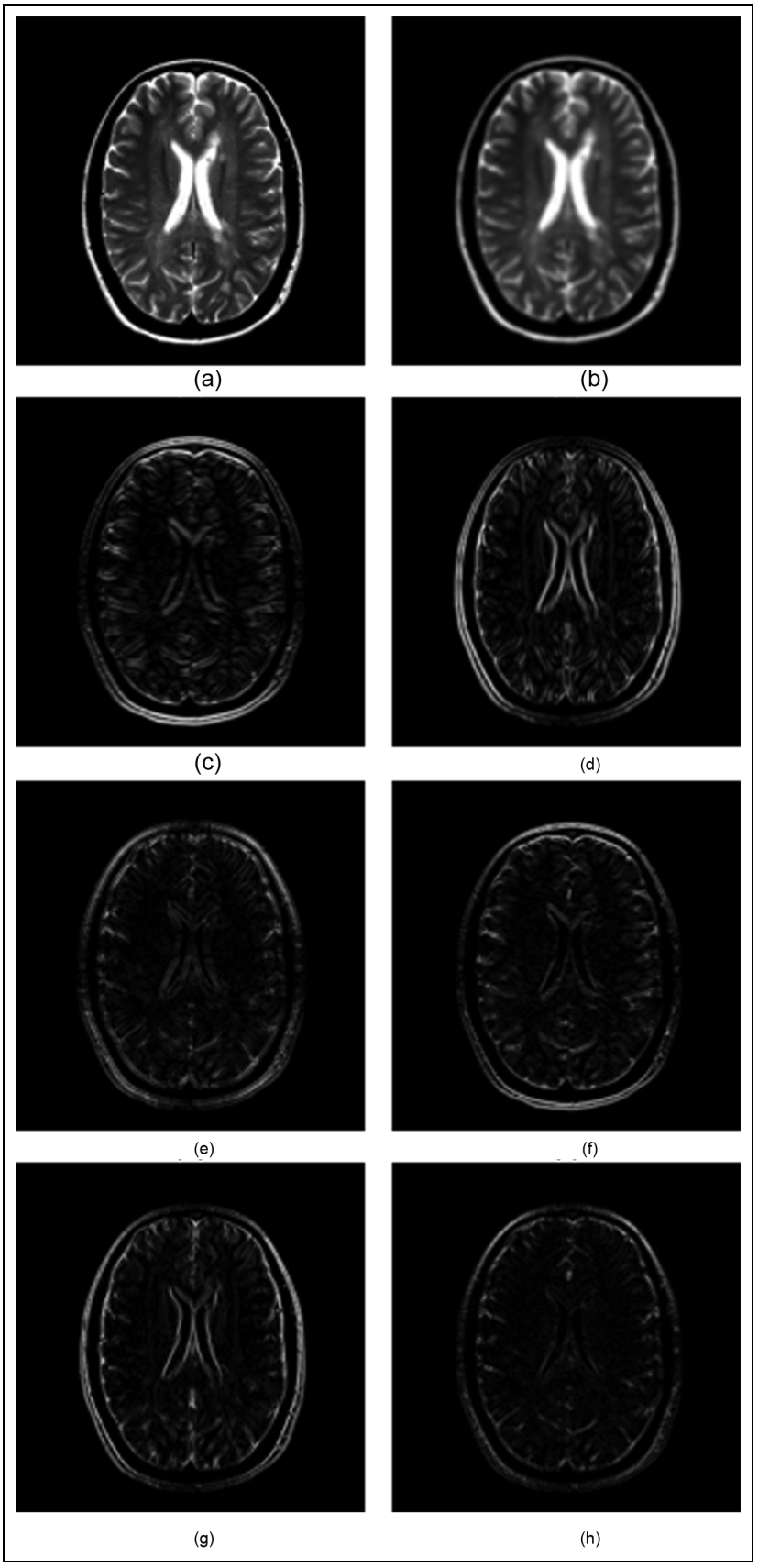


**Figure 7.** Two-dimensional discrete wavelet transform of a multiple sclerosis image: (a) original image; (b) $LL_2$; (c) $LH_2$; (d) $HL_2$; (e) $HH_2$; (f) $LH_1$; (g) $HL_1$; (h) $HH_1$.

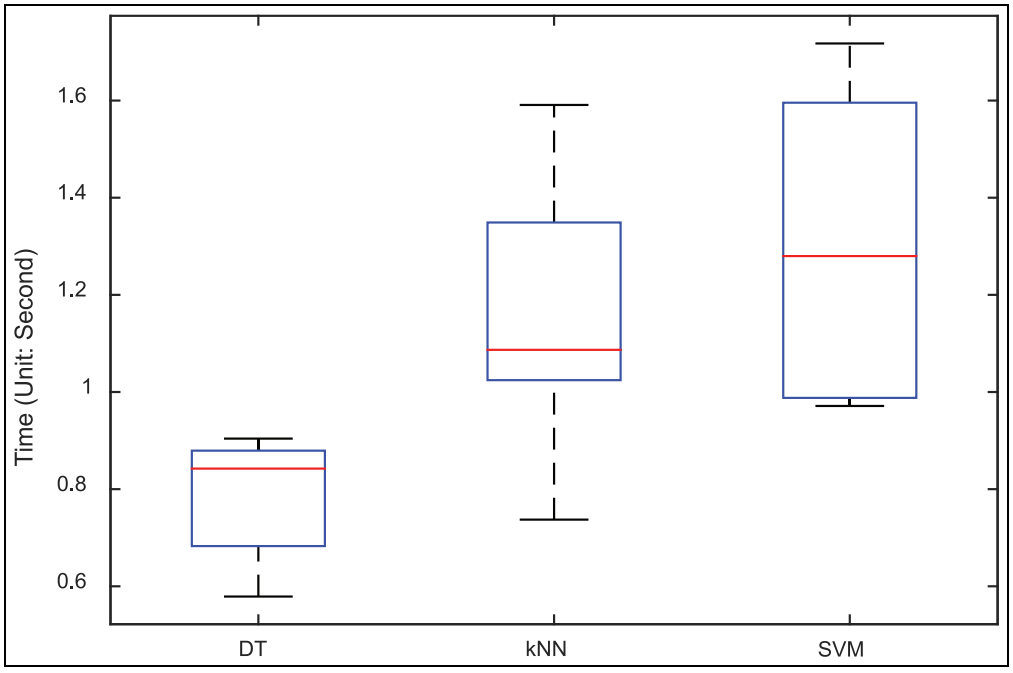


**Figure 8.** Computation time. DT: decision tree; kNN: *k*-nearest neighbors; SVM: support vector machine.

of MAMFM + SVM in this study are different from those of Murray et al.[15] In addition, we can observe that our SWE + kNN approach gives better performances in terms of all four measures than MAMFM + SVM,[15] DWT + PCA + LS-SVM,[16] WE + HBP,[17] and DWT + PPCA + RF.[21] This directly validates the effectiveness of our method. The reason may be two-fold. Firstly, we used an advanced feature extraction technique—SWE—that combined both SWT and Shannon entropy; thus, SWE provides our system with a more efficient feature. Secondly, kNN is a simple technique, but it presents almost perfect classification results due to the small-dimension (only seven features) classification task.

From Figure 8, we observe that the SVM takes the longest time of 1.3031 seconds, kNN takes the second longest time of 1.1613 seconds, and the DT costs the least time of 0.7847 seconds. The results are in line with our expectation. The training of the SVM needs to solve an optimization equation.[53] kNN does not need training time, since it only stores the feature vectors during the training phase.[54] Nevertheless, it needs to calculate the distance $k$ times in the validation phase.[55] For the DT, it has a simple tree-like structure and we also limit the maximum split number to four; in addition, the training algorithm C4.5 can prune branches to make a smaller tree. All of these help the DT to run as the fastest algorithm among the three. In practical use, the computation time will be reduced, since training will not be carried out for each new instance.

## 6 Conclusion and future directions

In this study, we developed a MS detection method based on SWE and kNN. The results showed its effectiveness.

In the future, we shall carry out the following researches: (i) acquire more MS imaging data to revalidate our algorithm; (ii) test other inter-scan normalization methods, such as histogram equalization; (iii) test other advanced entropic forms, for example, Tsallis entropy[56] and multiscale entropy[57]; (iv) test other advanced classifiers: the extreme learning machine,[58] kernel SVM,[59] probabilistic neural network,[60] and convolutional neural network[61]; (v) test an improved training algorithm for the DT, that is, C5.0 for Unix/Linux and See5 for Windows.[62]

### Author contributions

Yudong Zhang and Siyuan Lu conceived the study. Xingxing Zhou and Shuihua Wang designed the model. Ming Yang and Bin Liu acquired the data. Lenan Wu and Preetha Phillips analyzed and processed the data. Yudong Zhang and Shuihua Wang interpreted the data. Siyuan Lu and Xingxing Zhou developed the programs. Yudong Zhang, Ming Yang, and Shuihua Wang wrote the draft. All authors gave critical revisions and approved the submission.

### Funding

This work was supported by the Open Project Program of the State Key Lab of CAD&CG, Zhejiang University (grant number A1616),

the Open Fund of the Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University (grant number 93K172016K17), the Open Fund of the Key Laboratory of Statistical Information Technology and Data Mining, State Statistics Bureau (grant number SDL201608), the Natural Science Foundation of Jiangsu Province (grant number BK20150983), the Open Research Fund of Hunan Provincial Key Laboratory of Network Investigational Technology (grant number 2016WLZC013), the NSFC fund (61602250), and the Open Fund of Fujian Provincial Key Laboratory of Data Intensive Computing (grant number BD201607).

## References

1. Learmonth YC and Motl RW. Physical activity and exercise training in multiple sclerosis: a review and content analysis of qualitative research identifying perceived determinants and consequences. *Disabil Rehabil* 2016; 38: 1227–1242.
2. Bos SD, Berge T, Celius EG, et al. From genetic associations to functional studies in multiple sclerosis. *Eur J Neurol* 2016; 23: 847–853.
3. Raggi A, Covelli V, Schiavolin S, et al. Work-related problems in multiple sclerosis: a literature review on its associates and determinants. *Disabil Rehabil* 2016; 38: 936–944.
4. Sorensen PS, Sellebjerg F, Lycke J, et al. Minocycline added to subcutaneous interferon-1a in multiple sclerosis: randomized RECYCLINE study. *Eur J Neurol* 2016; 23: 861–870.
5. Tur C. Fatigue management in multiple sclerosis. *Curr Treat Opt Neurol* 2016; 18: Article ID 26.
6. Perrin PB, Panyavin I, Paredes AM, et al. A disproportionate burden of care: gender differences in mental health, health-related quality of life, and social support in Mexican multiple sclerosis caregivers. *Behav Neurol* 2015; 2015. DOI: 10.1155/2015/283958.
7. Costello F. Vision disturbances in multiple sclerosis. *Semin Neurol* 2016; 36: 185–195.
8. Kimbrough DJ, Sotirchos ES, Wilson JA, et al. Retinal damage and vision loss in African American multiple sclerosis patients. *Ann Neurol* 2015; 77: 228–236.
9. Longbrake EE, Lancia S, Tutlam N, et al. Impaired color vision as determined by Farnsworth Munsell 100 Hue testing is tightly associated with retinal thinning in multiple sclerosis. *Multiple Sclerosis J* 2014; 20: 358–358.
10. Hoang PD, Gandevia SC and Herbert RD. Prevalence of joint contractures and muscle weakness in people with multiple sclerosis. *Disabil Rehabil* 2014; 36: 1588–1593.
11. Burschka JM, Keune PM, Hofstadt-van Oy U, et al. Mindfulness-based interventions in multiple sclerosis: beneficial effects of Tai Chi on balance, coordination, fatigue and depression. *BMC Neurol* 2014; 14: Article ID 165.
12. Horvath A, Perlaki G, Toth A, et al. Increased diffusion in the normal appearing white matter of brain tumor patients: is this just tumor infiltration?. *J Neuro Oncol* 2016; 127: 83–90.
13. Magee R, Nolan YM and Downer EJ. An Assessment of inflammatory activity in human post-mortem cortical lesions and normal appearing white matter in multiple sclerosis. *Ir J Med Sci* 2016; 185: S19-S20.
14. Wegner C, Swiniarski A, Ott M, et al. Increased ongoing axonal injury in spinal normal-appearing white matter in multiple sclerosis as compared to neuromyelitis optica. *Multiple Sclerosis J* 2015; 21: 69–69.
15. Murray V, Rodriguez P and Pattichis MS. Multiscale AM-FM demodulation and image reconstruction methods with improved accuracy. *IEEE Trans Image Proc* 2010; 19: 1138–1152.
16. Siddiqui MF, Reza AW and Kanesan J. An automated and intelligent medical decision support system for brain MRI scans classification. *Plos One* 2015; 10: Article ID e0135875.
17. Phillips P, Dong Z and Yang J. Pathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization. *Progr Electromagn Res* 2015; 152: 41–58.
18. Khotanlou H and Afrasiabi M. Feature selection in order to extract multiple sclerosis lesions automatically in 3D brain magnetic resonance images using combination of support vector machine and genetic algorithm. *J Med Signal Sens* 2012; 2: 211–218.
19. Tachinaga S, Hiura Y, Kawashita I, et al. Development of a computer-aided diagnostic system for detecting multiple sclerosis using magnetic resonance images. *Nihon Hoshasen Gijutsu Gakkai zasshi* 2014; 70: 223–229.
20. Deshpande H, Maurel P and Barillot C. Classification of multiple sclerosis lesions using adaptive dictionary learning. *Comput Med Imag Graph* 2015; 46: 2–10.
21. Nayak DR, Dash R and Majhi B. Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. *Neurocomputing* 2016; 177: 188–197.
22. Saritha M, Joseph KP and Mathew AT. Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Pattern Recognit Lett* 2013; 34: 2151–2156.
23. Zhou XX, Ji GL, Yang JQ, et al. Detection of abnormal MR brains based on wavelet entropy and feature selection. *IEEJ Trans Electr Electr Eng* 2016; 11: 364–373.
24. Du S, Atangana A, Liu A, et al. Application of stationary wavelet entropy in pathological brain detection. *Multimedia Tool Appl*. Epub ahead of print 11 March 2016. DOI: 10.1007/s11042–016–3401–7.
25. Negi SS and Bhandari YS. A hybrid approach to image enhancement using contrast stretching on image sharpening and the analysis of various cases arising using histogram. In: *recent advances and innovations in engineering (ICRAIE)*, Jaipur, 2014.
26. Langarizadeh M, Mahmud R, Ramli AR, et al. Improvement of digital mammogram images using histogram equalization, histogram stretching and median filter. *J Med Eng Technol* 2011; 35: 103–108.
27. Nguyen N, Vo A, Choi I, et al. A stationary wavelet entropy-based clustering approach accurately predicts gene expression. *J Comput Biol* 2015; 22: 236–249.
28. Lenis G, Pilia N, Oesterlein T, et al. P wave detection and delineation in the ECG based on the phase free stationary wavelet transform and using intracardiac atrial electrograms as reference. *Biomed Eng Biomedizinische Technik* 2016; 61: 37–56.

29. Merah M, Abdelmalik TA and Larbi BH. R-peaks detection based on stationary wavelet transform. *Comput Meth Program Biomed* 2015; 121: 149–160.
30. Dong Z, Liu AJ, Wang S, et al. Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine. *J Med Imag Health Inform* 2015; 5: 1395–1403.
31. Langerudi MF, Rashidi TH and Mohammadian A. Individual trip rate transferability analysis based on a decision tree approach. *Transp Plann Technol* 2016; 39: 370–388.
32. Bernauer F, Hurkamp K, Ruhm W, et al. Snow event classification with a 2D video disdrometer - a decision tree approach. *Atmosph Res* 2016; 172: 186–195.
33. Zhang Y, Wang S, Phillips P, et al Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. *Knowl Base Syst* 2014; 64: 22–31.
34. Magniez F, Nayak A, Santha M, et al. Improved bounds for the randomized decision tree complexity of recursive majority. *Random Struct Algorithm* 2016; 48: 612–638.
35. Wylie CE, Shaw DJ, Verheyen KLP, et al. Decision-tree analysis of clinical data to aid diagnostic reasoning for equine laminitis: a cross-sectional study. *Vet Rec* 2016; 178: 8.
36. Sathyadevan S and Nair RR. Comparative analysis of decision tree algorithms: ID3, C4.5 and random forest. In: Jain LC, Behera HS, Mandal JK, et al. *Computational intelligence in data mining*. Vol. 31 of Smart Innovation Systems and Technologies. Berlin: Springer, 2015, pp.549–562.
37. An FW, Mihara K, Yamasaki S, et al. Highly flexible nearest-neighbor-search associative memory with integrated k nearest neighbor classifier, configurable parallelism and dual-storage space. *Jpn J Appl Phys* 2016; 55: Article ID 04ef10.
38. Chatzimilioudis G, Costa C, Zeinalipour-Yazti D, et al. Distributed in-memory processing of all k nearest neighbor queries. *IEEE Trans Knowl Data Eng* 2016; 28: 925–938.
39. Park Y, Hwang H and Lee SG. A novel algorithm for scalable k-nearest neighbour graph construction. *J Inform Sci* 2016; 42: 274–288.
40. Chon AT. Design of lazy classifier based on fuzzy k-nearest neighbors and reconstruction error. *J Korean Inst Intell Syst* 2010; 20: 101–108.
41. Gonzalez M, Bergmeir C, Triguero I, et al. On the stopping criteria for k-nearest neighbor in positive unlabeled time series classification problems. *Inform Sci* 2016; 328: 42–59.
42. Siriteerakul T, Boonjing V and Gullayanon R. Character classification framework based on support vector machine and k-nearest neighbour schemes. *Scienceasia* 2016; 42: 46–51.
43. Almasi ON and Rouhani M. Fast and de-noise support vector machine training method based on fuzzy clustering method for large real world datasets. *Turk J Electr Eng Comput Sci* 2016; 24: 219–233.
44. Yang X-J, Dong Z-C, Liu G, et al. Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine. *SpringerPlus* 2015; 4: Article ID 716.
45. Carrasco M, Lopez J and Maldonado S. A second-order cone programming formulation for nonparallel hyperplane support vector machine. *Exp Syst Appl* 2016; 54: 95–104.
46. Iosifidis A and Gabbouj M. Multi-class support vector machine classifiers using intrinsic and penalty graphs. *Pattern Recognit* 2016; 55: 231–246.
47. Wang S, Yang X, Zhang Y, et al. Identification of green, Oolong and black teas in China via wavelet packet entropy and fuzzy support vector machine. *Entropy* 2015; 17: 6663–6682.
48. Negri RG, Dutra LV and Sant'Anna SJS. Comparing support vector machine contextual approaches for urban area classification. *Remote Sens Lett* 2016; 7: 485–494.
49. Zhang Y-D, Chen S, Wang S-H, et al. Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine. *Int J Imag Syst Technol* 2015; 25: 317–327.
50. Newby D, Freitas AA and Ghafourian T. Coping with unbalanced class data sets in oral absorption models. *J Chem Inform Model* 2013; 53: 461–474.
51. Thomasian A, Li Y and Zhang LJ. Optimal subspace dimensionality for k-nearest-neighbor queries on clustered and dimensionality reduced datasets with SVD. *Multimedia Tool Appl* 2008; 40: 241–259.
52. Porta A, Castiglioni P, Bari V, et al. K-nearest-neighbor conditional entropy approach for the assessment of the short-term complexity of cardiovascular control. *Physiol Meas* 2013; 34: 17–33.
53. Lopez FJM, Puertas SM and Arriaza JAT. Training of support vector machine with the use of multivariate normalization. *Appl Soft Comput* 2014; 24: 1105–1111.
54. Chavan S, Abdelaziz A, Wiklander JG, et al. A k-nearest neighbor classification of hERG $K^+$ channel blockers. *J Comput Aid Mol Des* 2016; 30: 229–236.
55. Tanveer M, Shubham K, Aldhaifallah M, et al. An efficient regularized k-nearest neighbor based weighted twin support vector regression. *Knowl Base Syst* 2016; 94: 70–87.
56. Wu L. Pattern recognition via PCNN and Tsallis entropy. *Sensors* 2008; 8: 7518–7529.
57. Stosic D, Stosic D, Ludermir T, et al. Correlations of multiscale entropy in the FX market. *Phys A Stat Mech Appl* 2016; 457: 52–61.
58. Sajjadi S, Shamshirband S, Alizamir M, et al. Extreme learning machine for prediction of heat load in district heating systems. *Energ Build* 2016; 122: 222–227.
59. Zhou X-X, Yang J-F, Sheng H, et al. Combination of stationary wavelet transform and kernel support vector machines for pathological brain detection. *Simulation*. Epub ahead of print 2 March 2016. DOI: 10.1177/0037549716629227.
60. Zhang Y, Wu L, Naggaz N, et al. Remote-sensing image classification based on an improved probabilistic neural network. *Sensors* 2009; 9: 7516–7539.
61. Audhkhasi K, Osoba O and Kosko B. Noise-enhanced convolutional neural networks. *Neural Network* 2016; 78: 15–23.
62. Zhao JL and Huang WJ. Integrating Landsat TM imagery and see5 decision-tree software for identifying croplands: a case study in Shunyi District, Beijing. In: Gong Z, Luo XF, Chen JJ, et al. (eds) *Web information systems and mining*. (*Lecture Notes in Computer Science*, Vol. 6987). Berlin: Springer Press, 2011, pp.251–258.

## Author biographies

**Prof. Yudong Zhang** received his PhD degree in Signal and Information Processing from Southeast University in 2010. From 2010 to 2012, he worked at Columbia University as a postdoc under the supervision of Prof. Zhengchao Dong. From 2012 to 2013, he worked as a research scientist at Columbia University and New York State Psychiatric Institute under the supervision of Prof. Zhengchao Dong. At present, he is a full professor and doctoral advisor at the School of Computer Science and Technology at Nanjing Normal University. He also serves as the academic leader of the Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing. His research interests focus on computer-aided diagnosis and biomedical image processing.

**Siyuan Lu** received his BS from the Nanjing Normal University of Computer Science & Technology 2014. Now He is pursuing his MS at the Nanjing Normal University of Computer Science & Technology, majoring in image processing and machine learning. He is also affiliated with the State Key Lab of CAD & CG, Zhejiang University.

**Xingxing Zhou** is a PhD candidate studying at Nanjing Normal University. His research interest is biomedical image processing.

**Prof. Ming Yang** received her MD degree of medicine in radiology from Southeast University in 2011. Now she is the chief doctor at Nanjing Children's Hospital. Her research interests are functional MRI and cardiovascular imaging.

Prof. **Lenan Wu** was born in Quanzhou, China. He received his BS degree and MS in electronic engineering from the Nanjing University of Aeronautics and Astronautics Nanjing, in 1982 and 1987, respectively. He received his PhD degree in signal and information processing from Southeast University, Nanjing, in 1997. Now he works as a professor post-doctor in Southeast University. His research interests are multimedia information processing and communication signal processing. He has published over 290 journal papers and is the author or co-author of nine books.

**Prof. Bin Liu** obtained his MD from the Southeast University in 1990, and received his master's degree in Medical Imaging from Southeast University, Nanjing, China, in 2004. His research interests focus on functional MRI.

**Preetha Phillips** is pursuing a Bachelor of Science from Shepherd University. Her expected date of graduation is May 2016. She has conducted research on using magnetic resonance spectroscopy (MRS) in diagnosing panic disorder at Columbia Presbyterian Medical Center. She has also helped conduct research analyzing pain mechanisms in rats at Stony Brook University. She is an active member of the Biology National Honors Society (Beta Beta Beta), National Honors Society (Phi Kappa Phi), and Alpha Phi Omega.

**Dr Shuihua Wang** received her BS from Southeast University in 2008 and her MS from The City University of New York in 2012. She received her PhD from Nanjing University in 2016. At present, she works as an assistant professor in Nanjing Normal University. Her research interest is biomedical imaging.