Multimed Tools Appl
DOI 10.1007/s11042-015-2582-9

# Data and feature mixed ensemble based extreme learning machine for medical object detection and segmentation

**Wanzheng Zhu**[1] · **Weimin Huang**[2] · **Zhiping Lin**[1] · **Yongzhong Yang**[1] · **Su Huang**[2] · **Jiayin Zhou**[2]

Received: 27 September 2014 / Revised: 23 February 2015 / Accepted: 23 March 2015


**Abstract** Extreme learning machine (ELM) is a single-hidden layer feed-forward neural network with an efficient learning algorithm. Conventionally an ELM is trained using all the data based on the least square solution, and thus it may suffer from overfitting. In this paper, we present a new method of data and feature mixed ensemble based extreme learning machine (DFEN-ELM). DFEN-ELM combines data ensemble and feature subspace ensemble to tackle the overfitting problem and it takes advantage of the fast speed of ELM when building ensembles of classifiers. Both one-class and two-class ensemble based ELM have been studied. Experiments were conducted on computed tomography (CT) data for liver tumor detection and segmentation as well as magnetic resonance imaging (MRI) data for rodent brain segmentation. To improve the ensembles with new training data, sequential kernel learning is adopted further in the experiments on CT data for speedy retraining and iteratively enhancing the image segmentation performance. Experiment results on different testing cases and various testing datasets demonstrate that DFEN-ELM is a robust and efficient algorithm for medical object detection and segmentation.

✉ Weimin Huang
wmhuang@i2r.a-star.edu.sg

Wanzheng Zhu
zhuw0006@e.ntu.edu.sg

Zhiping Lin
ezplin@ntu.edu.sg

Yongzhong Yang
yyang8@e.ntu.edu.sg

Su Huang
huangs@i2r.a-star.edu.sg

Jiayin Zhou
jzhou@i2r.a-star.edu.sg

[1] School of Electrical and Electronic Engineering, 50 Nanyang Avenue, 639798, Singapore

[2] Institute for Infocomm Research, 1 Fusionopolis Way, #21-01, Connexis, 138632, Singapore

Published online: 14 May 2015

**Keywords** Extreme learning machine (ELM) · Ensemble learning · Overfitting · Classifier · Medical object detection and segmentation · Iterative learning

## 1 Introduction

Extreme learning machine (ELM) is a single-hidden layer feed-forward neural network (SLFN) with an efficient learning algorithm [13]. Instead of computing weights and biases by conventional optimization algorithms such as the gradient descent based back-propagation (BP) [9] method, it reduces the training time significantly by randomly assigning weights and biases for hidden nodes. ELM has been proved to have the ability to present good generalization performance. Many researchers have applied ELM to various applications, such as face recognition, human action recognition, landmark recognition and protein sequence classification [1, 3, 20, 32]. As a type of neural network, ELM is trained towards a minimum training error over a given training dataset. Inevitably it may suffer from the overfitting problem and cannot guarantee good results on unseen dataset. Zhou et al. showed that combinations of a set of neural networks might help mitigate the problem [30]. To avoid overfitting, Cao et al. presented a voting based extreme learning machine method [2] by performing multiple independent ELM trainings and making the final decision by majority voting [24]. In an online learning scheme, Lan et al. proposed an ensemble of online sequential ELM method [15] by constructing classifiers with new data in different incremental steps and averaging all classifiers' results, and Wang and Han [28] modified the kernel ELM for online sequential learning for time series prediction. To alleviate the two-class imbalance problem, Lin et al. studied multiple extreme learning machines [18] based on combining sampling technique, feature selection ensemble, extreme learning machine ensemble and decision tree to address the problem on the case of life cycle prediction. From another perspective, Liu and Wang proposed ensemble based extreme learning machine (EN-ELM) using cross-validation scheme on data ensemble [19]. Very recently, Mirza et al. proposed an ensemble of subset online sequential ELM method for class imbalance and concept drift [21].

In a recent work by our group (Huang et al. [11]), a random feature subspace ensemble ELM (FE-ELM) is proposed to avoid overfitting, using both one-class and two-class kernel-based ELM for tumor detection and segmentation from images. In this paper, we extend the work in [11] by further exploring both data ensemble and feature space ensemble and propose an improved algorithm called data and feature mixed ensemble based extreme learning machine (DFEN-ELM). This algorithm aims to make a strong classifier by combining several weak classifiers [7], while taking the advantage of fast speed of ELM [13] and overcoming the overfitting problem at the same time. The main idea is to randomly choose a subset of the whole training dataset to make up a new training dataset $I$, following a preset ratio of positive to negative samples; then for this new training dataset $I$, a subset of the whole feature spaces is chosen as a feature subspace $S$; finally, a classifier is constructed by training ELM using training dataset $I$ with feature subset $S$ and then applied to the testing datasets. The above mechanism is to construct one classifier, and it is repeated multiple times to fully exploit the data and feature spaces. These classifiers form the ensemble scheme, and the results of them are combined together to obtain the final detection or segmentation results by majority voting [24].

The proposed DFEN-ELM is then applied to medical image object detection and segmentation, which is a crucial step in the medical image analysis/interpretation workflow for clinical diagnosis, treatment planning, intra-operative guidance, and biomedical research. More

specifically, we will present the applications of ELM method in 1) CT-image based liver lesion detection and segmentation, and 2) magnetic resonance imaging (MRI) based rodent brain segmentation. To tackle the training data inefficiency (not representative, imbalanced), the sequential learning [28] approach is adopted to enable retraining of each classifier with additionally selected training data, and thus to help improve the overall performance. Results on CT image segmentation show that retraining with additionally selected data using this scheme takes much shorter time than the initial training.

In liver lesion detection and segmentation, many researchers have developed automatic or semi-automatic methods with reliable performance. A two-class support vector machine classifier with a propagational learning strategy [31] is proposed by Zhou et al. for liver tumor segmentation. In addition, Häme et al. proposed non-parametric intensity distribution estimation and a hidden Markov measure field model [8] for liver tumor segmentation. To tackle with the boundary leakage problem, Li et al. proposed a unified level set method [16]. Taking advantage of ensemble methods, Huang et al. proposed a random feature subspace ensemble ELM for liver tumor detection and segmentation [11]. In [4] the Gaussian mixture of the image intensity has been used as the basic classifier in tumor detection. In MRI brain extraction, Smith developed a surface deformation method known as Brain Extraction Tool (BET) for human brain image [26]. Besides the human brain research, rodent image analysis is attracting increasing attention in preclinical and pharmaceutical research. Several works on rodent brain extraction have been reported in recent years. A deformable model based method similar to BET [26] was proposed in [17] to extract rodent brain. An atlas-based method to extract the brain from rodent images was developed by Oguz, et al. [22]. Another rodent brain extraction work by Chou et al. is based on 3-D pulse-coupled neural networks [5]. What is more, many researchers have also tried other methods of image segmentation. For examples, Yushkevich et al. developed 3-D active contours to segment anatomical structures [29] and Veksler et al. proposed star shape prior for graph segmentation [27], etc. In the experiments, we apply DFEN-ELM method to liver tumor detection and segmentation from CT images and rodent brain segmentation from MRI T2 images. Comparisons of the DFEN-ELM with the baseline ELM, the data ensemble based ELM (DE-ELM) and the feature ensemble based ELM (FE-ELM) are presented. Results achieved are also benchmarked with other methods such as SVM [31] and 3D Pulse Coupled Neural Networks (3D PCNN) [5].

The remaining of this paper is organized as follows. Section 2 describes the relevant preliminaries such as ELM and related topics. Section 3 explains the proposed DFEN-ELM method. The experimental results are presented and discussed in Section 4. Lastly, the conclusion is drawn in Section 5.

## 2 Preliminaries

### 2.1 Extreme learning machine (ELM)

ELM is proposed as a single-hidden layer feed-forward neural network [13]. It has been shown that the learning speed is much faster than many iterative optimization methods such as conventional neural networks (NN). An ELM first generates its hidden layer weights and biases $\{W_i,b_i\}$ randomly, which map all training inputs into $L$-dimensional hidden layer nodes, and then it computes the output weights by least square solution [13]. Considering a training dataset as $\{X,Y\}=\{\{X_1,Y_1\}, \{X_2,Y_2\}, \ldots, \{X_N,Y_N\}\}$, for its $j^{th}$ training sample $\{X_j,Y_j\}$, where $X_j=[X_{j1}, X_{j2},\ldots, X_{jn}]$ is an $1 \times n$ input vector for input feature representation of a training

sample, $Y_j=[Y_{j1}, Y_{j2},\ldots, Y_{jm}]$ is an 1×m vector of the labelling of the j[th] sample, $j=1,2,\ldots,N$, the output of ELM with $L$ hidden nodes is formulated as:

$$f(X_j)=\sum_{i=1}^{L}\beta_i g(W_i\cdot X_j+b_i)=Y_j, j=1,2,\ldots,N \tag{1}$$

where $\beta_i=[\beta_{i1},\beta_{i2},\ldots,\beta_{im}]$ is an 1×m output weight vector connecting the $i^{th}$ hidden node and the output nodes, $i=1,2,\ldots,L$, $W_i=[W_{i1},W_{i2},\ldots,W_{in}]$ is an 1×n hidden layer weight vector connecting the input nodes and $i^{th}$ hidden node, $X_j$ is input training data, $W_i\cdot X_j$ denotes the inner product of $W_i$ and $X_j$, $b_i$ is hidden layer bias for node $i$, $g(W_i\cdot X_j+b_i)$ is a non-linear activation function, $f(X_j)$ is the ELM output, a 1×m vector. The equation above can also be written in a compact format as

$$f(\mathrm{X})=\mathrm{H}(\mathrm{X})\beta, \tag{2}$$

where $f(\mathrm{X})$ is $N\times m$ matrix for all training samples X, H(X) is the hidden layer output matrix of size $N\times L$, $\beta$ is of size $L\times m$, and

$$H_{ji}=g(W_i\cdot X_j+b_i) \tag{3}$$

is the output of each element in the matrix H (X).

The final output weight is solved by finding a least-square solution for $\mathrm{H}(\mathrm{X})\beta=\mathrm{Y}$

$$\beta=\mathrm{H}^{+}\mathrm{Y} \tag{4}$$

where $\mathrm{Y}=[Y_1^{'},Y_2^{'},\ldots,Y_N^{'}]^{'}$ is a N×m matrix, each row is an output vector, and $\mathrm{H}^{+}=(\mathrm{H}^{'}\mathrm{H})^{-1}\mathrm{H}^{'}$ with dimension of L×N is Moore-Penrose generalized inverse [25] matrix of H.

### 2.2 Kernel based extreme learning machine

Since ELM generates random hidden layer parameters, a limitation is that it may be stuck into a local optimum. Kernel-based Extreme Learning Machine is developed to establish a robust output using a regularized coefficient and kernel mapping [12]. The output function of Kernel Extreme Learning Machine is:

$$f(X)=h(X)\mathrm{H}^{'}\left(\frac{\mathrm{I}}{\lambda}+\mathrm{HH}^{'}\right)^{-1}\mathrm{Y}=\begin{bmatrix}K(X,X_1)\\ \vdots\\ K(X,X_N)\end{bmatrix}\left(\frac{\mathrm{I}}{\lambda}+\Omega_{ELM}\right)^{-1}\mathrm{Y}, \tag{5}$$

where $\{X_1,\ldots,X_N\}$ are the training samples, and $X$ is a testing sample.

$$h(X)=g(X), \tag{6}$$

$$\mathrm{H}=[h(X_1),\ldots,h(X_N)]^{'}, \tag{7}$$

$$\Omega_{ELM}=\mathrm{HH}^{'},\ \Omega_{ELMi,j}=h(X_i)h(X_j)=K(X_i,X_j) \tag{8}$$

H is the hidden-layer output matrix, in kernel ELM, of dimension N×1. $\Omega_{\mathrm{ELM}}$ is of size N×N, $g(X)$ is a hidden non-linear function that need not to be known to users, Y is the label of

training samples of size of N×m, I is the identity matrix, and $\lambda$ is the regularization coefficient. In this experiment, Gaussian kernel with $\gamma$=100 is used.

$$K(\mathrm{u},\mathrm{v}) = \exp\left(\gamma\|\mathrm{u} - \mathrm{v}\|^2\right) \tag{9}$$

### 2.3 Auto-encoder

Auto-encoder is usually used as an unsupervised learning [6]. It is a neural network trained with its output being the same as the input. Auto-encoder aims to transform its inputs into outputs with the least error of distortion. Meanwhile, it will capture the underline information of training dataset [6]. In our experiment, auto-encoder is used as a pre-training step towards an overall better performance and ELM is used to generate the auto-encoder because of its fast speed. Training input data X are mapped into $L$ hidden nodes, and then an ELM is trained to obtain $\beta$, so that X=H$\beta$. The auto-encoding of a sample $X$=[$x_1$, $x_2$,…, $x_n$] is shown in Fig. 1 below.

### 2.4 Data ensemble extreme learning machine

Zhou et al. proposed ensembling neural network with randomly assigned weights [30], which leads to the result of 'many could be better than all'. Using the similar algorithm, Cao et al. presented a voting based extreme learning machine method [2] by performing multiple independent ELM training. The main mechanism is as follows. Part of training data are chosen randomly from the whole training dataset, and then ELM training was conducted only based on those chosen data. The classifier is applied to the testing dataset to obtain a classification result. By changing the selected training data subset, multiple classifiers can be trained and multiple testing results can be obtained. The final result is calculated based on the majority voting scheme [24]. It shows that data ensemble ELM can outperform original ELM.

## 3 Data and feature mixed ensemble based ELM

ELM generates random hidden weights and biases to reduce the training time dramatically. However, due to its randomness, the parameters might be stuck into local optima. To overcome

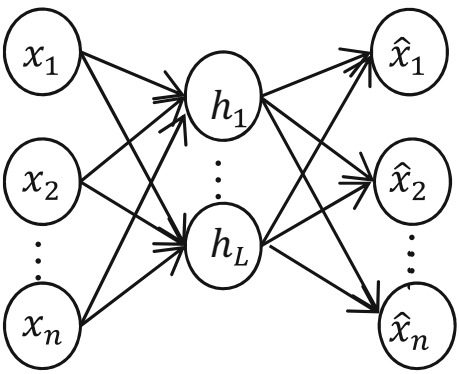


**Fig. 1** Auto-encoder

the limitation of traditional ELM, we propose data and feature ensemble based extreme learning machine (DFEN-ELM), which is developed as an extension on the basis of our previous work FE-ELM [11].

### 3.1 Baseline ELM for ensemble

For the auto-encoder ELM (AE-ELM), it considers output Y equal to input X for training, through solving the following equations.

$$\begin{aligned} h_j \triangleq h(X_j) &= \left[g(W_1 X_j + b_1), \ldots g(W_L X_j + b_L)\right], \\ \mathrm{H} &= \left[h_1', \ldots, h_N'\right]', \\ \beta &= \left(\frac{\mathrm{I}}{\lambda} + \mathrm{H}'\mathrm{H}\right)^{-1} \mathrm{H}'X, \end{aligned} \tag{10}$$

where $\beta$ is the $L \times n$ output weight matrix, Y=X is $N \times n$ matrix with each row being a training sample, H is the N×L hidden-layer output matrix, $W_i$ is th 1×L hidden layer weight vector, $b_i$ is the hidden layer bias, a scalar, $g(\cdot)$ is a non-linear function, $\{X_1, X_2, \ldots, X_N\}$ are input and output training data, I is the identity matrix, and λ is the regularization coefficient. With the AE-ELM,

$$\widehat{X} = f(X) = h(X)\beta \tag{11}$$

In the original AE-ELM [14], instead of using randomly generated input weights $W$ to the hidden layer, a conventional ELM is used with the coefficient matrix $\beta$ from the AE as $W$. As in our experiment, we observe that the conventional ELM does not work well for one-class training problem, while kernel based ELM which maps training data to another nonlinear space can solve the one-class problem. Based on the findings, the combination of auto-encoder ELM and kernel based ELM is selected as the baseline ELM for all the experiments in this study.

### 3.2 One-class extreme learning machine

For one-class ELM, tumor detection and segmentation problem can be treated as an anomaly detection problem. The training data consist of healthy liver voxels only. Ideally, by examining Eqs. (1), (3) and (4), a healthy testing sample should expect '$f(x)=1$' as its output. Thus in a feature space, a hyperplane can be constructed with all healthy liver voxels, and the difference $|f(x)-1|$ is the distance of any point (a sample in either class) to this hyperplane constructed by the ELM. Since the hyperplane is learned to represent one class samples of liver, tumor voxels should have an output $f(x)$ far away from '1'. A distance threshold can be set by a threshold selection method [23] in order to distinguish tumor from healthy liver tissues.

### 3.3 Feature ensemble based extreme learning machine

Our previous work on feature ensemble based ELM [11] is also further studied in this paper for comparison. In feature ensemble, one randomly selected feature subspace is used to create one independent classifier. Multiple classifiers are constructed through training on different feature subspaces, and testing results are obtained from those classifiers. Subsequently, all the testing results are combined for the final decision by majority voting [24].

### 3.4 Data and feature mixed ensemble based ELM (DFEN-ELM)

The data ensemble method can perform well if training data are well-diversified and representative. Meanwhile, it may not help if the data are less representative even though features are well-representative. The feature ensemble method can also encounter the same problem. Therefore, data and feature mixed ensemble method is proposed in this paper, which can help achieve better performance. We present the details of the proposed DFEN-ELM method as follows.

The main idea of DFEN-ELM is to randomly choose a subset $I$ of the whole training data set according to a data choosing rate. Then for this new training dataset $I$, a subset $S$ of the whole feature set is chosen based on a feature choosing rate. Lastly, the data subset $I$ embedded with the feature subset $S$ is used to train a classifier. The above steps are repeated to obtain different classifiers. All classifiers are applied to a testing data set to obtain multiple classification results, based on which, the majority voting [24] is used to get the final result. The method can be summarized as Algorithm 1 and shown as follows.

**Algorithm 1** DFEN-ELM

**Algorithm:**

**Input:** training dataset and testing dataset, data choosing rate, feature choosing rate, number of classifiers M.

**For** Each classifier $r$=1, 2, …, M

① Image pre-processing,

② Normalize training dataset and testing dataset,

③ Randomly choose a portion ***I*** from training dataset based on a data choosing rate,

④ For the chosen portion ***I***, construct training dataset by choosing a feature subspace ***S*** from the whole features based on the feature choosing rate,

⑤ Perform ELM auto-encoding using the chosen portion ***I*** of training data embedded with the chosen portion of feature subspace ***S***,

⑥ ELM training using the chosen portion ***I*** of training data embedded with the chosen portion of feature subspace ***S***,

⑦ ELM testing for all testing data sets.

**End**

⑧ Make the final decision using the majority voting scheme of all classifiers.

⑨ Image post-processing.

**End of Algorithm**

The diagram of the main DFEN-ELM algorithm is shown in Fig. 2. It will be applied to both one-class and two-class training depending on the training data available.

### 3.5 Iterative learning

Even with the data and feature mixed ensemble learning, prior to the testing, it may be hard to know the quality of the training data in some cases. Therefore, the initial training data might not be representative or might be imbalanced. In such applications, we can add additional training data based on the previous testing result, and then re-train each classifier again using

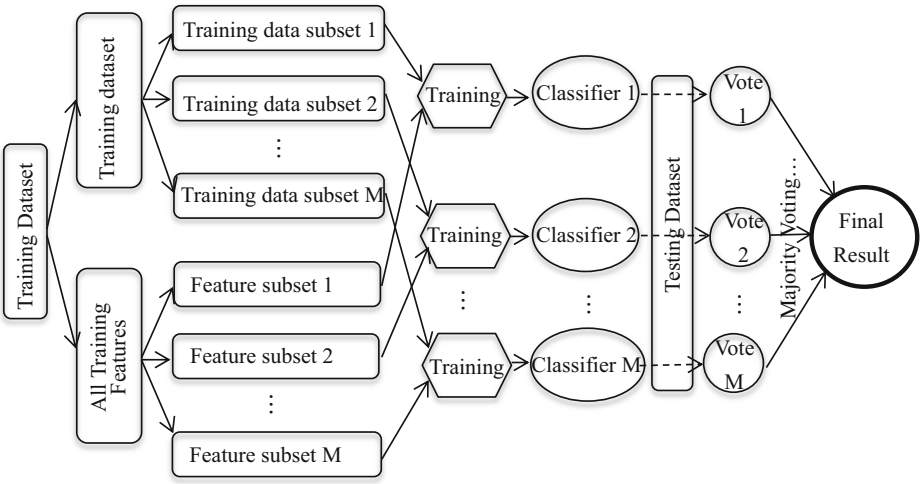


**Fig. 2** Diagram of the main DFEN-ELM algorithm

the intermediate result together with the newly selected training data to obtain improved performance. Note that we need not redo the auto-encoding. Only the classifiers for the ensemble to be retrained with the new data. Here the sequential learning method for kernel ELM updating [28], originally used for time series prediction, is adopted for a fast retraining.

The iterative learning scheme is conducted only on human liver tumor segmentation experiment, from the application point of view so that it will not cost users too much effort to select the data. The main idea of iterative learning scheme is illustrated in Fig. 3 and explained as follows.

Firstly, the intermediate segmentation results are observed and some pixels from obviously misclassified regions are selected manually from three orthogonal slices. Secondly, each classifier is retrained by using the updated training data containing the original training data and the newly added training data. Lastly, classification results are obtained from each classifier and combined to make the decision using the majority voting approach. This process

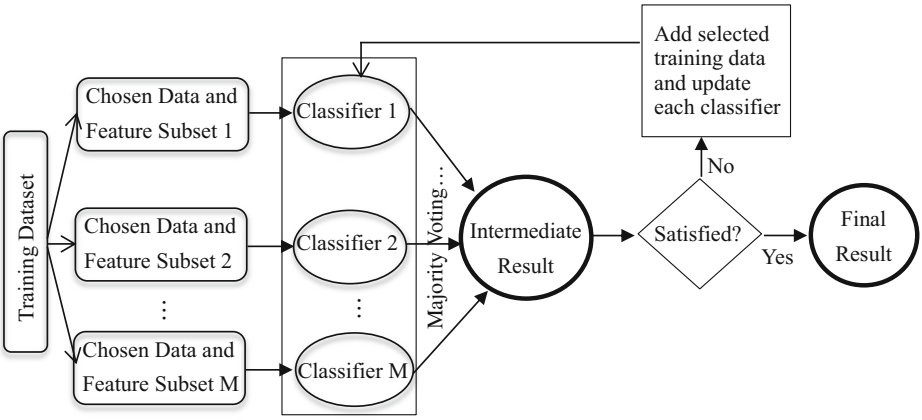


**Fig. 3** Diagram of the iterative learning algorithm

can be repeated if necessary. Usually one or two iterations will be enough to achieve a much improved result.

### 3.6 Discussion on the ensemble of classifiers

Feature and data ensembles affect the final result in different way. In the baseline kernel ELM (5), the feature mapping K ($X_i$, $X_j$) and $\Omega_{ELM}$ will be different when the features are selected differently. Thus the feature ensemble is actually the voting by combination of the results from different feature spaces. In data ensemble, by selecting different training data, the kernel ELM is actually learning the distribution of the subset of the data after the kernel mapping. Thus the mixed ensemble presented in the paper could take the advantage of the data generalization from both feature selection and data selection. Finally the iterative learning provides an approach to adjust the data distribution for better characterization.

## 4 Experiment and discussion

The testing of the proposed DFEN-ELM is conducted on human liver CT data and rodent brain MRI data. The results are compared with the outputs from the baseline ELM, data ensemble ELM and feature ensemble ELM, as well as propagational SVM [31] and 3D PCNN [5] for various experiments. The additional iterative learning scheme is conducted only on human liver tumor segmentation experiment, from the application point of view. The original CT/MRI images may have different intensity contrasts and some of the slices may contain noises. Pre-processing such as Gaussian filtering, anisotropic diffusion and histogram equalization are applied to the images. Among them, Gaussian filtering and anisotropic diffusion can help reduce noises, and histogram equalization can normalize the intensity of the organ of interests and enhance the contrasts of different objects. Normalization of the whole dataset is also adopted since patients' data can be quite different from each other. Post-processing operations such as morphological erosion and dilation are also applied in the experiment. Erosion is a fundamental morphological operator, and the basic effect is to erode away the boundaries of regions. Dilation is used together with erosion, and it can recover the boundary parts eroded away. Erosion and dilation mixed together can remove small misclassified regions or pixels. Image features used for each voxel include $3\times3\times3$ gray level mean intensity, $5\times5\times3$ gray level mean intensity, $3\times3\times3$ gray level power, $5\times5\times3$ gray level power, $3\times3$ edge magnitude, variance, energy, correlation, entropy, contrast, homogeneity, cluster shade, cluster prominence and 9 Law's texture features [10]. All these features are extracted in local windows in order to encode more information into the training data. Then these extracted features are pooled into a vector for each voxel, and multiple feature vectors are combined into a matrix which is fed into a base classifier to learn the representation of tumor and non-tumor voxels.

Liver tumor detection is performed based on CT images from 11 patients, and there are 27 tumors marked in total. Liver tumor segmentation is performed based on the tumors with segmentation ground truth on 19 tumors from the same 11 patients. The resolutions of CT-scans range from $0.5859\times0.5859\times1$ to $0.7891\times0.7891\times1$ ($mm^3$). The size of every 2-D slice is $512\times512$ pixels. The number of slices for each patient differs, ranging from 150 to 300. The 3-D tumor size of all the patients also differs, ranging from 153 to 23236 voxels according to the ground truth which is manually labelled. Kernel ELM with $\gamma=100$ together with an ELM auto-encoder is used as a baseline ELM for liver tumor detection and segmentation.

### 4.1 Liver tumor detection

Automatic or semi-automatic tumor detection can save clinicians' large amount of time to screen CT data for diagnosis. The liver tumor detection can be treated as an anomaly detection problem if there is no prior tumor model obtained, which is called one-class anomaly detection. On the other hand, if the tumor model can be established, two-class classification can also be adopted for fully automatic tumor detection. In this work, both one-class and two-class tumor detections are tested, together with DFEN-ELM to be compared with baseline ELM.

#### *4.1.1 One-class liver tumor detection*

For one-class liver tumor detection, training data for a patient is obtained by only choosing healthy liver voxels from 2 to 3 slices of the same patient's CT data. In this case, the tumor detection has become an anomaly detection problem. All the voxels in the liver region of that test patient will be scanned to detect tumor tissue using the trained classifier. The performance measurement is based on the ground truth labelled by medical experts. In liver tumor detection experiment, we obtain the results by keeping the sensitivity of tumor detection at 100 and 93 % by using different number of erosion and dilation operations as well as adjusting threshold values on ELM output. 100 % sensitivity means all 27 tumors were detected as true positives and none of the tumor is omitted while 93 % sensitivity means 2 tumors are not detected. A false positive indicates that a healthy tissue or region is misclassified as a tumor. One-class liver tumor detection results are summarized in Table 1.

It can be seen that when the sensitivity is kept at 100 %, ensemble methods can outperform the baseline ELM itself, and DFEN-ELM method has the least false positive tumors. DFEN-ELM reduces most false positive tumors since it has the least overfitting problem and it can represent training data and predict testing data well. However, when the sensitivity is kept at 93 %, the ensemble methods do not help. One of the possible reasons is that the false positive is quite low and too many erosion operations have wiped out most of the small false positive tumors, and it does not eliminate false positive tumors further.

#### *4.1.2 Two-class liver tumor detection*

For two-class liver tumor detection, instead of getting training data from the same patient's CT data, we obtain training data by using other patients' CT data, and using healthy liver voxels as negative samples and lesion voxels as positive training samples. In this way, our two-class

**Table 1** One-class detection error (with the training data from the same patient)

| | Mean false positive per patient (Sensitivity=100 %) | Mean false positive per patient (Sensitivity=93 %) |
|---|---|---|
| Baseline ELM | 5.27 | 1.45 |
| Feature ensemble based ELM (FE-ELM) | 4.36 | 1.45 |
| Data ensemble based ELM (DE-ELM) | 4.00 | 1.45 |
| Data and feature mixed ensemble based ELM (DFEN-ELM) | 3.45 | 1.45 |

liver tumor detection method becomes a fully automatic approach for a new testing CT. The training and testing procedures of ensemble are the same as those of one-class scheme (Fig. 2). The detection is also conducted at 100 and 93 % sensitivities for comparison. The results for two-class liver tumor detection are listed in Table 2.

It can be seen that the overall trend for two-class tumor detection is very similar to that of one-class tumor detection. When the sensitivity is kept at 100 %, ensemble methods can outperform the baseline ELM itself, and DFEN-ELM method has the least false positive tumors. But when the sensitivity is kept at 93 %, the ensemble methods do not help very much.

Since the two-class tumor detection is based on the training data totally independent of that used in one-class training, and it is for fully automatic tumor detection, the results in Table 2 cannot be compared directly to the one-class detection result in Table 1. One example is from the patients shown in Fig. 4. The left image in Fig. 4 shows an example where the voxels of the tumor have brighter intensity, while the right figure in Fig. 4 shows a different type of tumor of which the voxels have darker intensity. In two-class training, since the training data are obtained from other patients, a classifier trained using the right image of Fig. 4 as the training data cannot generate a good detection result of the left image of Fig. 4. However the improvement of the results within the same training scheme (one-class or two-class) is observed, which shows that mixed ensemble learning DFEN-ELM outperforms the baseline ELM, the data ensemble ELM and the feature ensemble ELM.

Figure 5 shows two CT images with only the true positive detection results (top left) and two images with the false positive errors (top right). Yellow labelled boxes are true positive tumors and red labelled boxes are false positive tissues. ELM may misclassify some blood vessels (brighter part in a liver) as tumor, or misclassify the liver boundary part as tumor. One possible reason for misclassifying the blood vessels is that the training data contain few training samples in bright blood vessel part, which is not enough to form a representative of the blood vessel voxels for the one-class training. Therefore, it may misclassify a big and brighter blood vessel as anomaly (such as the top red box in the fourth image in Fig. 5. The misclassification of inferior vena cava (IVC) (the red box in central part of the third and fourth images in Fig. 5) is due to that the training and testing are conducted on voxel level. Since all the voxels are smoothed by a Gaussian filter to reduce noise, the voxels similar to lesion tissue in intensity may be classified by ELM as a possible lesion. The four bottom images in Fig. 5 show intermediate output results from ELM. From the images of the intermediate result, it can be seen that liver boundaries are very likely to be classified as tumor, but most of the boundary parts will be removed after post-processing such as average filtering, image erosion and image dilation.

**Table 2** Two-class detection error (with the training data from other patients)

| | Mean false positive per patient (Sensitivity=100 %) | Mean false positive per patient (Sensitivity=93 %) |
|---|---|---|
| Baseline ELM | 4.36 | 1.73 |
| Feature ensemble based ELM (FE-ELM) | 4.18 | 1.73 |
| Data ensemble based ELM (DE-ELM) | 3.82 | 1.73 |
| Data and feature mixed ensemble based ELM (DFEN-ELM) | 3.64 | 1.64 |

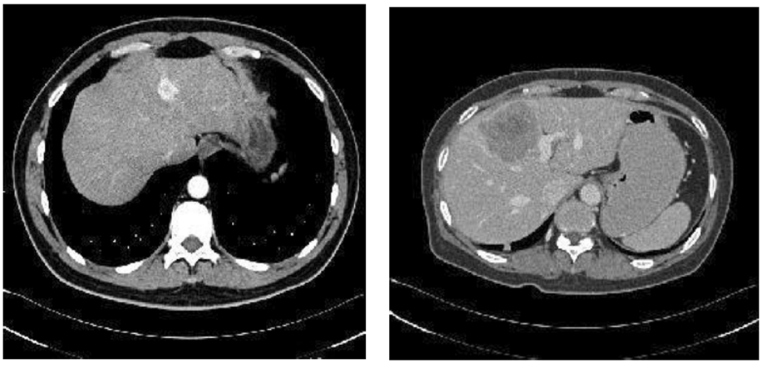

**Fig. 4** *Left* figure indicates a tumor example. *Right* figure indicates a different type of tumor

We also compared our results with the work in [4], where Mixture of Gaussian is used to filter the tumor tissues first, followed by a tumor model detailing the size and the tumor shape (sphere like and symmetrical). The result shows a comparable performance, for [4], the false alarm is around 1 tumor per patient in average, while we achieved 1.45 tumors per patient. The good performance of [4] is mainly due to the tumor shape model, which can be combined with our ELM classifier to improve our result. For example, in the fourth image in Fig. 5, the false tumor detected is of elongated shape, which can be efficiently filtered by the tumor shape information.

### 4.2 Two-class liver tumor segmentation

Once a tumor is detected, segmentation will be the next crucial step for diagnosis and planning for treatment. Since boundary voxels of livers may affect the overall liver tumor segmentation due to noise and unclear lesion boundary, it is difficult to segment by fully automatic method.

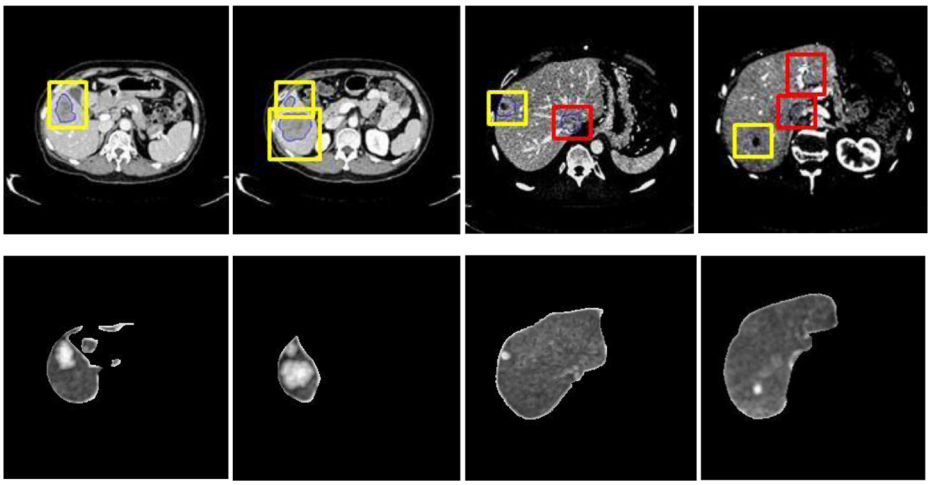

**Fig. 5** One-class liver tumor detection. *Top row* shows the true detection in *yellow box*, and false detection in *red box*. The *bottom row* shows the intermediate output of the ELM output. Compared to the original CT, the tumor regions have been highlighted with higher image intensity

Human intervention is often required to help select a region of interest (ROI) in order to improve the performance. The two-class ensemble ELMs are tested in this experiment, with an iterative approach to adjusting the training data. The overall performance is evaluated using Jaccard index (volume overlap) which is defined as below:

$$Volume\ Overlap = \frac{S \cap G}{S \cup G} \tag{12}$$

where $S$ represents the segmentation result and $G$ is the respective ground truth of the lesion.

In the experiment, we test the semi-automatic segmentation approach using the proposed method and we also test the baseline, the feature ensemble based ELM and the data ensemble based ELM for comparison. Besides, we apply the semi-automatic SVM method [31] to compare with the baseline ELM. For each patient's CT scan, the training data for a patient is obtained by manually choosing the samples from some slices of the CT by drawing a rough region (e.g., circular or rectangle region) to get positive voxels in the lesion and negative voxels in the healthy liver parenchyma region. The segmentation work is performed within a region of interest drawn manually around the tumor. All the experiments for two-class tumor segmentation by different methods use the same training data for each patient, the same testing data and the same ROI.

Table 3 shows volume overlap of tumors by using different two-class classification methods. Baseline ELM results in about 73 % volume overlap of lesion segmentation and the ground truth, and DFEN-ELM method can increase by about 2.2 % on top of that. It can be seen that DFEN-ELM method is about 1 and 1.3 % better than data ensemble method and feature ensemble method respectively. In comparison, the baseline ELM is comparable to the semi-automatic SVM based segmentation on the same data set.

Figure 6 shows the data ensemble and feature ensemble results for two-class liver tumor segmentation. The red dashed line (top one) indicates the data ensemble results and the solid blue line indicates the feature ensemble results. The horizontal dashed line in black indicates the performance of the baseline method without ensemble. It clearly shows that the performance of data ensemble is going up with the data choosing rate increasing until the rate reaches 25 to 30 %, and then the performance would drop with the data choosing rate increasing further and finally it reaches the baseline performance. The peak volume overlap value is about 74.3 %. For feature ensemble, the trend is quite similar. The performance is also going up with the increasing of the feature choosing rate until it reaches 50 to 60 %, and then it drops to the baseline performance. We find that the number of classifiers does not change the overall performance much as long as the number of classifiers is over a certain number (10–20 in our experiments). When the number of classifiers is too low, it will perform worse even than

**Table 3** Two-class liver lesion segmentation

| | Volume overlap (%) |
|---|---|
| SVM | 72.92 |
| Baseline ELM | 73.09 |
| Feature ensemble based ELM (FE-ELM) | 73.92 |
| Data ensemble based ELM (DE-ELM) | 74.31 |
| Data and feature mixed ensemble based ELM (DFEN-ELM) | 75.26 |

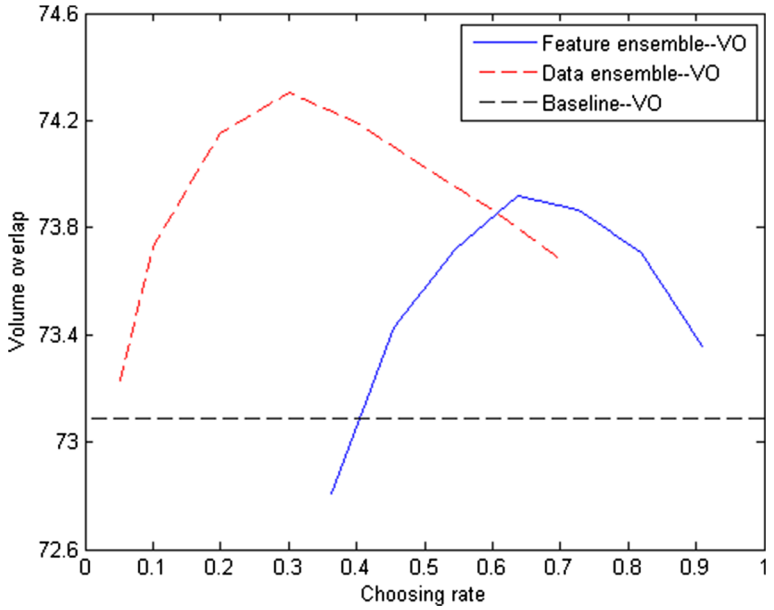


**Fig. 6** Volume overlap of the two-class segmentation and the ground truth, with different ensemble methods, under different selection ratios

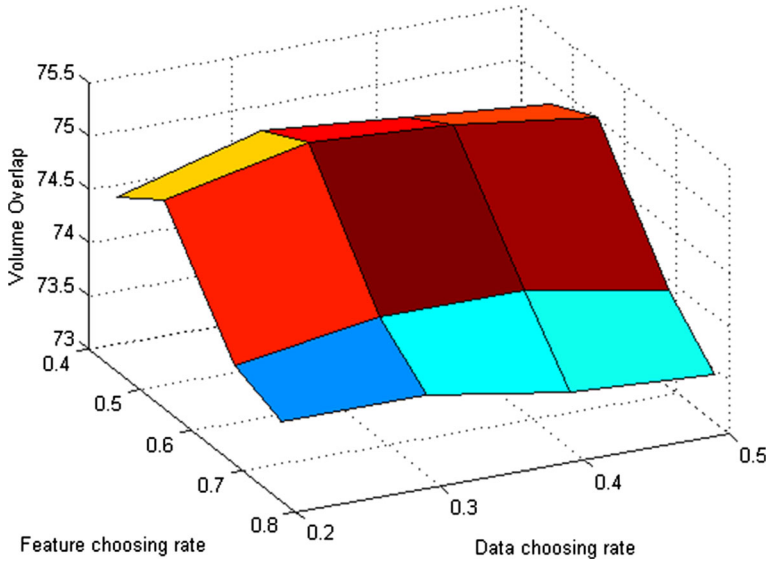


**Fig. 7** Volume overlap of the DFEN-ELM two-class segmentation of tumors and the ground truth, under different selection ratios

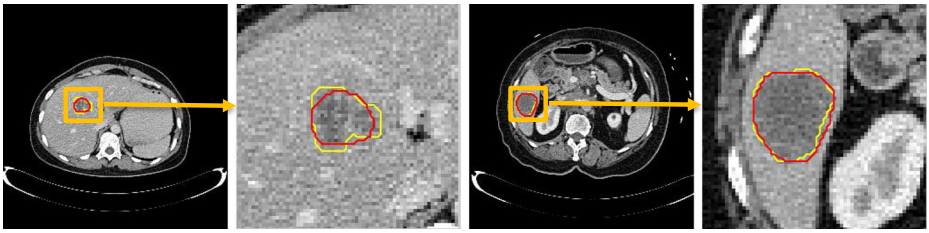

**Fig. 8** Two-class DFEN-ELM liver tumor segmentation. *Yellow lines* indicate our segmentation results and red lines indicate the ground truth labelled by experts

the baseline ELM as the data or feature is too sparse to represent the real data distribution. The experiment results shown in the rest of the paper are based on 15 classifiers.

Figure 7 shows the data and feature mixed ensemble results for two-class classification. The peak point occurs at about the 30 % data choosing rate and 60 % feature choosing rate. If the feature choosing rate reaches 100 %, which means all features are used, the result is consistent with data ensemble curve (The peak value is about 74.3 %). Similarly, if data choosing rate reaches 100 %, which means all data are used, the result is consistent with feature ensemble curve (The peak value is about 73.9 %).

Figure 8 shows some results of two-class DFEN-ELM segmentation. Although it may have some misclassifications, the overall lesion segmentation performs well, the boundaries are well preserved.

### 4.3 Two-class liver tumor segmentation with iterative learning

With the data and feature mixed ensemble method, we can achieve a better result based on the given training data. However due to the initial training data are not optimal, the results may not be satisfying. Figure 9 below shows such a case with bad initial segmentation and how the new training data are selected for re-learning. Due to the positive training data are usually much fewer than the negative data, in the iterative learning, all the new positive data will be used in each classifier before ensemble.

After adding new training data, we re-train each classifier to obtain the overall results as shown in Table 4. The experiment was conducted using 15 classifiers.

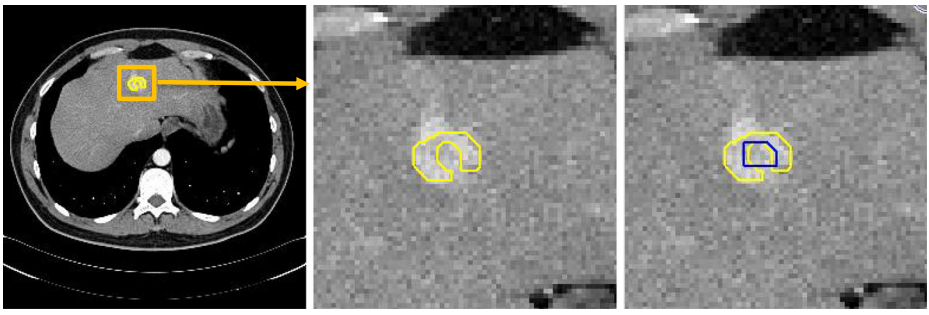

**Fig. 9** *Yellow regions* in the left one and the middle one show the first segmentation result. *Blue region* in the right one shows the added training data

**Table 4** Two-class liver lesion segmentation

| | Volume overlap (%) |
|---|---|
| Data and feature mixed ensemble based ELM (DFEN-ELM) | 75.26 |
| Data and feature mixed ensemble based Iterative ELM | 76.61 |

Table 4 shows that the overall volume overlap can now increase by about 1.35 % compared to the one using initial training data only. From the segmentation results, it can be seen that iterative learning can effectively improve its overall performance. In total 19 tumors, 2 tumors get improvement by over 8 % (up to 12 %) and there are other tumors getting improvement of 2~3 %. Fig. 10 shows an example of a tumor segmentation. The initial segmentation (yellow contour) does not perform well probably due to the appearance of the tumor and the training

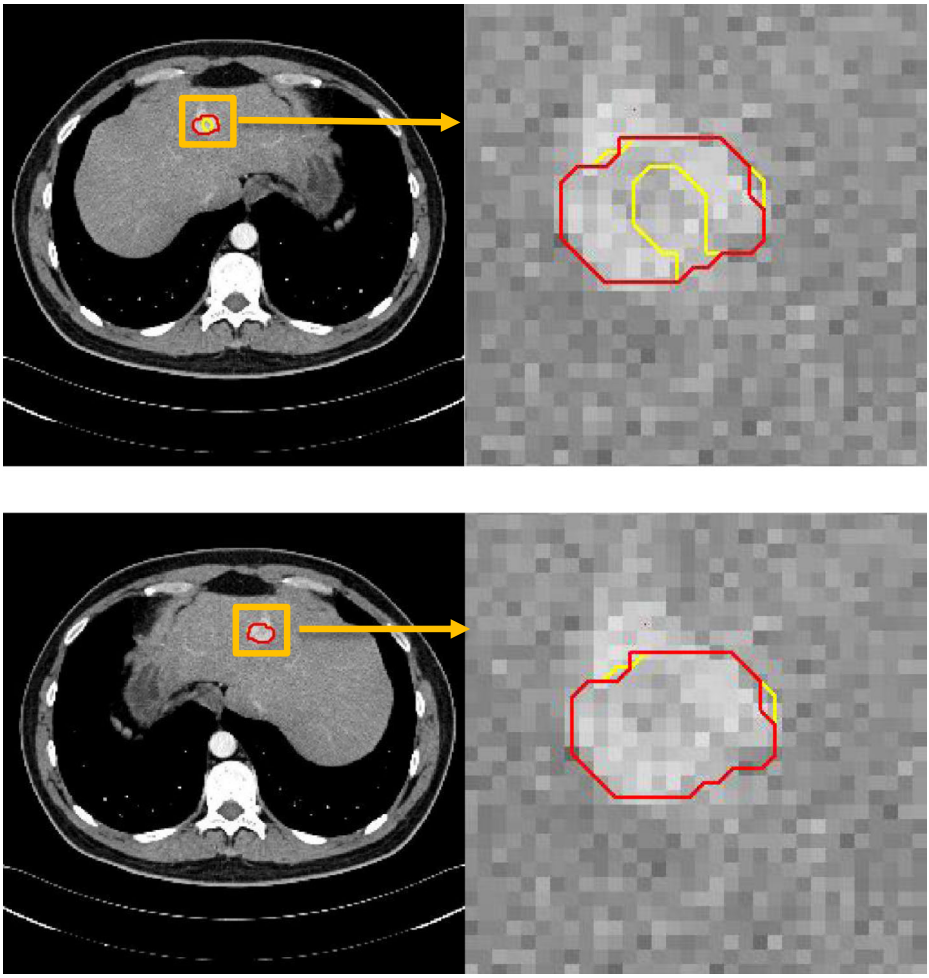

**Fig. 10** *Top two* figures show the first-time segmentation and *bottom two* figures show the sequential segmentation results. *Yellow lines* indicate our segmentation results and *red lines* indicate the ground truth labelled by experts

**Table 5** Two-class liver lesion segmentation training time

| | Training time (s) |
|---|---|
| First-time learning | 7~60 |
| Iterative learning | 0.1~1 |

data selection. After adding additional training data and conducting re-training, most of the misclassified voxels were corrected and the result shows significant improvement.

In addition to improving the quality of training data, Iterative learning scheme also has advantages in terms of computational efficiency. The trade-off is with more human intervention in the process. The majority time of training is to calculate the inverse matrix in (5). In our experiment, instead of calculating a new inverse matrix from scratch, the sequential method is adopted to compute the matrix inverse when new data (columns) are added. The training time for each ensemble is shown in Table 5. For each ensemble (with 15 classifiers), the updating learning only takes 1–2 % of the first training time, depending on the new added data size.

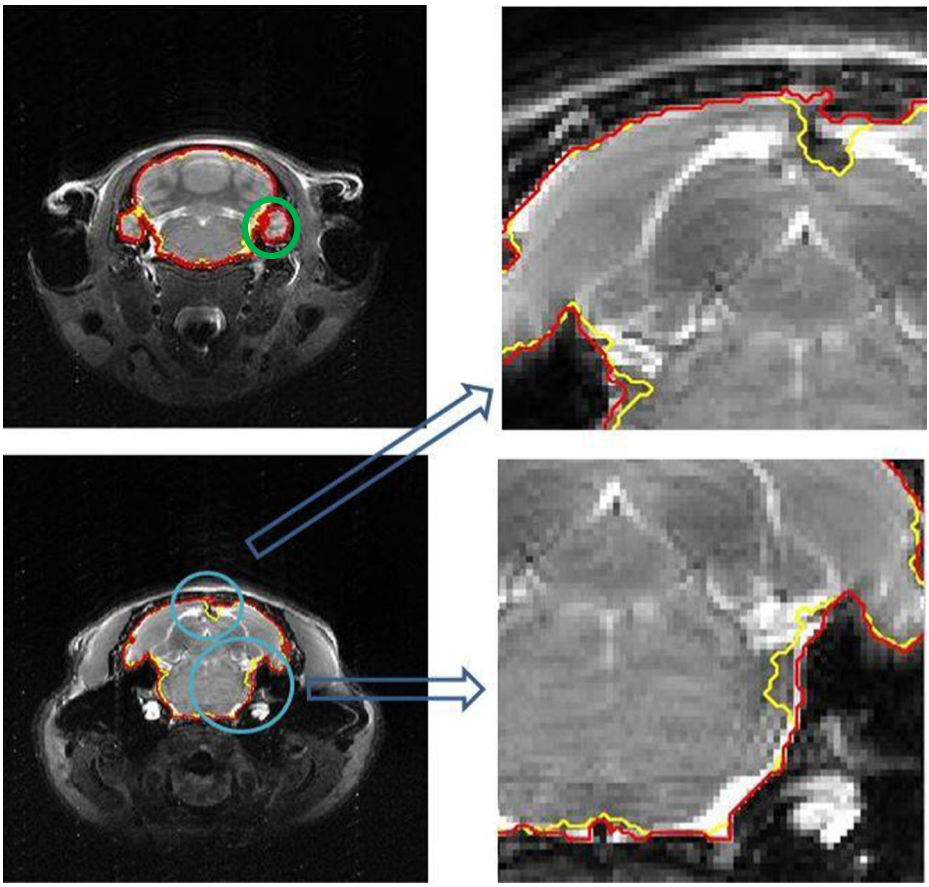

**Fig. 11** Segmentation results. *Yellow line* indicates PCNN results. *Red line* indicates the improvement by DFEN-ELM

**Table 6** Brain extraction result of rodent 1

| | Volume overlap (%) |
|---|---|
| PCNN [5] | 92.68 |
| ELM | 93.80 |
| F-ELM | 94.01 |
| D-ELM | 94.01 |
| DFEN-ELM | 94.07 |

### 4.4 Rodent brain segmentation from MRI images

So far, we have discussed using CT-scans for liver tumor detection and segmentation. It turns out DFEN-ELM outperforms baseline ELM in medical object detection and segmentation. In this section we apply the DFEN-ELM to another application for MRI rodent brain extraction, which is an important step in translational research for drug and disease study.

The rodent brain images are MRI T2 generated images, from a Bruker 9.4 T animal MRI system. The first rodent brain has $0.1367 \times 0.1367 \times 1$ mm voxel distance, and the size of every 2-D slice is $256 \times 256$ (pixel). The second rodent brain has $0.25 \times 0.25 \times 0.25$ mm voxel distance, and the size of every 2-D slice is $100 \times 100$ (pixel). This experiment is conducted based on the segmentation results using 3-D pulse-coupled neural networks (PCNN) by Chou et al. [5]. With the PCNN results, we train different ELMs to further improve the rodent brain segmentation. We applied morphological erosions on the PCNN results to get an inner boundary, and then an ELM will consider voxels inside the inner boundary as positive samples for brain. The same mechanism is applied to get negative samples, i.e., the ELM will consider voxels outside the outer boundary as negative samples after applying morphological dilation on the PCNN results. In our study, all rodent brain 2-D images are assumed being symmetrical, by taking advantage of that, our method would repair some missing brain parts such as the highlighted part in the top left figure shown in Fig. 11.

Tables 6 and 7 show the volume overlap for each of the two rats respectively. Basically, applying ELM on the PCNN results would help to increase about 1 % in volume overlap. Data and feature mixed ensemble extreme learning machine can improve another 0.3 %. The improvement is not significantly high since the inner part of the brain takes the majority volume and the original segmentation works for most regions well. However, from the details of the segmentation, the DFEN-ELM shows an obvious improvement in the boundary in terms of accuracy.

The example images in Fig. 11 show the PCNN results and the improvement with DFEN-ELM for some rodent brain slices, where yellow lines shows the results from PCNN and red lines shows the DFEN-ELM results. Top-left image shows that the DFEN-ELM method can repair some missing part in the rodent brain generated from the PCNN (the green circle part).

**Table 7** Brain extraction result of rodent 2

| | Volume overlap (%) |
|---|---|
| PCNN [5] | 91.58 |
| ELM | 92.43 |
| F-ELM | 92.61 |
| D-ELM | 92.58 |
| DFEN-ELM | 92.77 |

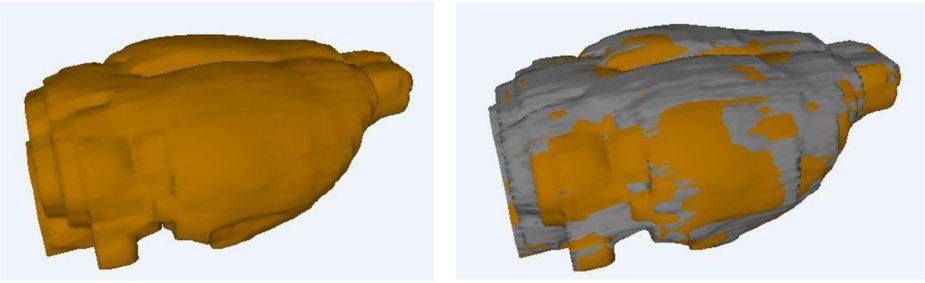

**Fig. 12** 3D visualization of a segmented rodent brain (*left image*), and the overlap on the ground truth (*right image*)

Shown in the right column of Fig. 11, in the enlarged regions, it can be seen that the DFEN-ELM results in a better boundary of the brain segmentation. For a better illustration we show a 3D model reconstructed from the proposed method in Fig. 12. The left one is 3-D model reconstructed based on the segmentation results of DFEN-ELM. The right one is the overlapping visualization where the yellow part is the segmentation results of DFEN-ELM and the white part is manually labelled ground truth. It can be seen that the overall shape of DFEN-ELM is quite close to the ground truth.

## 5 Conclusion

In this paper, we have proposed data and feature mixed ensemble based ELM (DFEN-ELM) for medical object detection and segmentation from CT and MRI images. Taking the advantages of ensemble based strategy, DFEN-ELM can alleviate overfitting problem. Compared with the baseline ELM, both data ensemble based ELM and feature ensemble based ELM can lead to a better performance while DFEN-ELM shows a further improvement than data ensemble based ELM or feature ensemble based ELM alone. Moreover in data ensemble ELM, our experiments show that the best data choosing rate is about 25–30 % and for feature ensemble ELM, the best feature choosing rate is about 60–80 %. For DFEN-ELM, our experiments show that the best performance occurs at 30 % data choosing rate and 60 % feature choosing rate. For all the ensemble methods, it is found that the number of classifiers does not change the overall performance as long as the number of classifiers is over a certain number (10–20 in our experiments), and thus the overall performance is maintained at a relatively stable value. To further tackle the possible bad initial training data issue, we adopt the sequential learning for iteratively updating the classifiers in CT image segmentation. It might improve the overall quality of the training data significantly and thus improve the overall performance. Meanwhile, using sequential learning scheme, the retraining can be sped up significantly.

Experiment results on different testing cases and various testing datasets demonstrate that DFEN-ELM is more accurate and more stable than the basic ELM itself, data ensemble based ELM and feature ensemble based ELM. Although the testing shows promising improvement, more validation with other large scale datasets from different applications will be our future work. Data imbalance can be another interesting problem for the DFEN-ELM method as we often observed that the data of different classes were severely unbalanced, for example when the tumor is small, or in some cases for iterative learning where the new training data selection cannot be easily

done. For the later, it is observed in both applications of rodent brain extraction and tumor segmentation due to the relatively small deviation of the initial segmentation from the real boundary.

**Acknowledgments** We wish to acknowledge the funding support for this project from Nanyang Technological University under the Undergraduate Research Experience on CAmpus (URECA) programme and the funding support by Agency for Science, Technology and Research (A*STAR) Joint Council Office under DP Grant 1334 k00084.

## References

1. Cao JW, Chen T, Fan J (2014) Fast online learning algorithm for landmark recognition based on BoW framework. Proceedings of the 9th IEEE Conference on Industrial Electronics and Applications, Hangzhou, pp 9–12
2. Cao J, Lin Z, Huang GB, Liu N (2012) Voting based extreme learning machine. Inf Sci 185(1):66–77
3. Cao J, Xiong L (2014) Protein sequence classification with improved extreme learning machine algorithms. BioMed Res Int (2014) Article ID 103054, 12 p, doi: 10.1155/2014/103054.
4. Chi Y, Zhou J, Venkatesh SK, Huang S, Tian Q, Hennedige T, Liu J (2013) Computer-aided focal liver lesion detection. Int J Comput Assist Radiol Surg 8(4):511–525
5. Chou N, Wu J, Bai Bingren J, Qiu A, Chuang KH (2011) Robust automatic rodent brain extraction using 3-D pulse-coupled neural networks (PCNN). IEEE Trans Image Process 20(9):2554–2564
6. Erhan D, Bengio Y, Courville A, Manzagol PA, Vincent P, Bengio S (2010) Why does unsupervised pre-training help deep learning? J Mach Learn Res 11:625–660
7. Freund Y, Schapire R, Abe N (1999) A short introduction to boosting. J Japan Soc Artif Intell 14(771–780):1612
8. Häme Y, Pollari M (2012) Semi-automatic liver tumor segmentation with hidden Markov measure field model and non-parametric distribution estimation. Med Image Anal 16(1):140–149
9. Haykin S (1999) Neural networks: a comprehensive foundation, 2nd edn. Prentice Hall, Upper Saddle River
10. Huang W, Li N, Lin Z, Huang GB, Zong W, Zhou J, Duan Y (2013) Liver tumor detection and segmentation using kernel-based extreme learning machine. In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp. 3662–3665
11. Huang W, Yang Y, Lin Z, Huang GB, Zhou J, Duan Y, Xiong W (2014) Random feature subspace ensemble based extreme learning machine for liver tumor detection and segmentation. IEEE Conf Eng Med Biol Soc (EMBC) 4675–4678
12. Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B Cybern 42(2):513–529
13. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501
14. Kasun LLC, Zhou H, Huang GB, Vong CM (2013) Representational learning with ELMs for big data. IEEE Intell Syst 28(6):31–34
15. Lan Y, Soh YC, Huang GB (2009) Ensemble of online sequential extreme learning machine. Neurocomputing 72(13):3391–3395
16. Li BN, Chui CK, Chang S, Ong SH (2012) A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images. Expert Syst Appl 39(10):9661–9668
17. Li J, Liu X, Zhuo J, Gullapalli RP, Zara JM (2011) A deformable surface model based automatic rat brain extraction method. Biomed Imaging: From Nano to Macro, 2011 I.E. Int Symp 1741–1745
18. Lin SJ, Chang C, Hsu MF (2013) Multiple extreme learning machines for a two-class imbalance corporate life cycle prediction. Knowl-Based Syst 39:214–223
19. Liu N, Wang H (2010) Ensemble based extreme learning machine. IEEE Signal Process Lett 17(8):754–757
20. Minhas R, Baradarani A, Seifzadeh S, Jonathan Wu QM (2010) Human action recognition using extreme learning machine based on visual vocabularies. Neurocomputing 73(10):1906–1917
21. Mirza B, Lin Z, Liu N (2015) Ensemble of subset online sequential extreme learning machine for class imbalance and concept drift. Neurocomputing 149:316–329
22. Oguz I, Lee J, Budin F, Rumple A, McMurray M, Ehlers C, Styner M (2011) Automatic skull-stripping of rat MRI/DTI scans and atlas building. In SPIE Med Imaging 796225–796225
23. Otsu N (1975) A threshold selection method from gray-level histograms. Automatica 11(285–296):23–27
24. Polikar R (2006) Ensemble based systems in decision making. IEEE Circuits Syst Mag 6(3):21–45
25. Rakha MA (2004) On the Moore–Penrose generalized inverse matrix. Appl Math Comput 158(1):185–200
26. Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17(3):143–155

27. Veksler O (2008) Star shape prior for graph-cut image segmentation. Proc of 10th European Conference on Computer Vision 2008. Springer Berlin Heidelberg, pp 454–467
28. Wang X, Han M (2014) Online sequential extreme learning machine with kernels for nonstationary time series prediction. Neurocomputing 145:90–97
29. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3):1116–1128
30. Zhou ZH, Wu J, Tang W (2002) Ensembling neural networks: many could be better than all. Artif Intell 137(1):239–263
31. Zhou J, Xiong W, Tian Q, Qi Y, Liu J, Leow WK, Han Y, Venkatesh S, Wang SC (2008) Semi-automatic segmentation of 3D liver tumors from CT scans using voxel classification and propagational learning. In MICCAI Workshop (41) p. 43
32. Zong W, Huang GB (2011) Face recognition based on extreme learning machine. Neurocomputing 74(16): 2541–2551

**Mr Wanzheng Zhu** is currently working towards the B.Eng. degree in School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore. During his undergraduate period, he received full scholarship awarded by Ministry of Education of Singapore. His current research interests include computer vision, machine learning, deep learning and ensemble learning.

**Dr Weimin Huang** received his Ph.D in Pattern Recognition and Intelligent Control from Tsinghua University, Beijing, China, in 1996. He is currently a Senior Scientist in the Institute for Infocomm Research, Singapore. His research interests include pattern recognition, image processing, computer vision, human computer interaction

and statistical learning, in area of object segmentation, detection, tracking, as well as medical image analysis and modeling. Dr Huang is PI and Co-PI for various research grants on video analysis, healthcare, image based diagnosis, surgical planning, simulation and image guided intervention.

**Prof Zhiping Lin** received the Ph.D. degree in information engineering from the University of Cambridge, England in 1987. Since Feb. 1999, he has been an Associate Professor at Nanyang Technological University, Singapore. Dr. Lin is currently serving as the Editor-in-Chief of Multidimensional Systems and Signal Processing (MSSP). He is also a Subject Editor of the journal of the Franklin Institute and the Lead Guest Editor of a special issue in Mathematical Problems in Engineering. He was an Associate Editor of IEEE Transactions on Circuits and Systems - II for 2010-2011 and General Chair of the 9th International Conference on Information, Communications and Signal Processing (ICICS), 2013. His research interests include MSSP, statistical and biomedical signal processing, and machine learning. He is the co-author of the 2007 Young Author Best Paper Award from the IEEE Signal Processing Society and Distinguished Lecturer of the IEEE Circuits and Systems Society for 2007-2008.

**Mr Yongzhong Yang** is currently a research officer at Bioinformatics Institute, Singapore. Before that, he received a B.Eng. from Nanyang Technological University, Singapore in 2014. His research interests include computer vision and machine learning.

**Mr. Su Huang** received his BSc degree in computational Mathematics from Zhongshan University, China in 1982, and MSc degree in Computer Science from Peking University, China in 1989. He has being involved in medical imaging research since 1998, and he is a principal research engineer now in Institute for Infocomm Research, Agency for Science, Technology and Research(A*STAR), Singapore. His research interests include medical image process, analysis and visualization algorithms and tools development.

**Dr Jiayin Zhou** received his Bachelor's and Master's degrees in biomedical engineering from Zhejiang University, Zhejiang, China in 1997 and 2000, respectively, and the Ph.D. degree from Nanyang Technological University (NTU), Singapore, in 2005. From 2004 to 2009, he was with NTU and then in the National University of Singapore as a Researcher for several medical image processing and analysis projects. Since 2009, he has been with the Institute for Infocomm Research, A*STAR, Singapore. His current research interests include medical image processing and analysis, quantitative imaging metrics, computer-aided surgery planning and simulation.