DEEP LEARNING FOR EXTRACTING WATER BODY FROM LANDSAT IMAGERY

Liu Yang[1,2,3], Shengwei Tian[1,2,3], Long Yu[2,3,4], Feiyue Ye[2,3,*]
Jin Qian[2,3] and Yurong Qian[1]

[1]School of Software
Xinjiang University
No. 499, Northwest Rd., Saybagh District, Urumqi, P. R. China
{ yangliu_xju; tianshengwei }@163.com

[2]Key Laboratory of Cloud Computing and Intelligent Information Processing of Changzhou City
[3]College of Computer Engineering
Jiangsu University of Technology
No. 1801, Zhongwu Rd., Changzhou 213001, P. R. China
*Corresponding author: yfy@jstu.edu.cn; qjqjlqyf@163.com

[4]Network Center
Xinjiang University
No. 14, Shengli Road, Tianshan District, Urumqi 830046, P. R. China
yul_xju@163.com

ABSTRACT. *There are regional limitations in traditional methods of water body extraction. For different terrain, all the methods rely heavily on carefully hand-engineered feature selection and large amounts of prior knowledge. Due to the difficulty and high cost in acquiring, the labeled data of remote sensing is relatively small. Thus, there exist some challenges in the classification of huge amount of high dimension remote sensing data. Deep Learning has a good capacity of hierarchical feature learning from unlabeled data. Stacked sparse autoencoder (SSAE), one deep learning method, is widely investigated for image recognition. In this paper, a new water body extraction model based on SSAE is established. At first, current useful features (NDWI, NDVI, NDBI and so forth) are collected to construct unique feature matrix for each pixel. Next, a Feature Expansion Algorithm (FEA) is designed by taking account of the influence of neighboring pixels to expand feature matrixes. Setting the expansion features as inputs, SSAE is trained to extract water body. The experimental results showed that the proposed model outperformed Support Vector Machine (SVM) and traditional neural network (NN). Meanwhile, the proposed FEA explored more distinct features of water body so that the accuracy of water body extraction was improved to a great extent.*
Keywords: Unsupervised feature learning, Deep Learning, Water body extraction, Feature Expansion Algorithm, Stacked sparse autoencoders

1. **Introduction.** Water is the vital natural resources for the human survival and development, as well as an important restriction factor of eco-environment. Accurate water body recognition is crucial to many applications including environmental monitoring [1], resource survey, flood assessment and drought detection [2]. Satellite-based remote sensing technology provides continuous snapshots of Earth's surface over long periods [3]. Remote sensing technology, which can be used for water body extraction, is becoming a research hotspot in recent years.

At present, numerous methods have been proposed to abstract water bodies from satellite images. The most commonly used methods can be divided into three categories:

single-band threshold method, multiband spectral threshold method and water index method.

(1) Single-band threshold method: Single-band threshold method utilizes the spectral characteristic differences between water and other objects. It distinguishes water by setting up suitable threshold [4]. Through analysis of the image histogram, Jupp used threshold method on band 7 [5]. Moller-Jensen achieved water extraction by setting thresholds on band 4 and 5 [6]. Lu and Li utilized the method of density slice on band 5. However, the proposed algorithm could only recognize water whose area is more than 4000 square meters [7]. Braud and Feng delineated coastline of Louisiana with setting threshold on each band separately before accessing corresponding performance, which showed that setting threshold on band 5 was the most appropriate method [8]. By this token, single-band threshold method is simple but cannot extract small linear water area and remove shadows.

(2) Multiband spectral relationship method: Barton and Bathols calculated the ratio image of channel 2 and channel 1 of AVHRR data and used it to recognize water from satellite images [9]. Du and Zhou found that only water has the feature of (B2 + B3) > (B4 + B5) [10]. Multiband spectral relationship method has better effect on water extraction but still cannot remove the shadows. Wang et al. combined (B2 + B3) > (B4 + B5) and B4/B2 < 0.9 to extract water body and found that the proposed method was suitable for extracting water of mountainous terrain [11].

(3) Water index method: This method is the most widely used method. McFeeters proposed the Normalized Difference Water Index (NDWI) which can inhibit part of background noise but mistake built-up land for water bodies [12]. The equation of NDWI is as follows: (Green – NIR)/(Green + NIR), where Green is green band and NIR is near-infrared band. Considering the defect of NDWI, Xu put forward a modified NDWI (MNDWI) that uses mid-infrared (MIR) instead of NIR. MNDWI can differentiate between buildings and water and be given by: (Green – MIR)/(Green + MIR) [13].

In different terrain, there are various shadows or background noise to be removed. For example, compared to arid area, humid area possesses abundant vegetation that adds a task about removing information of vegetation from water. In mountainous land, due to the overlap region between water body and shadow spectral ranges, the extracted water may always be mixed with shadows of mountain. Therefore, Shen et al. adopted object-oriented analysis for extracting more comprehensive features of water in mountain area to build the decision ruleset. Through that, the extraction accuracy increases to 95% and even higher in cloud-free case [14]. In 2013, a method of MNDWI combined with K-T transformation to distinguish water from mountain shadows was proposed [1]. In the same year, Wang et al. used three methods to remove the impact of the shadow and get the best one for extracting water in mountains [15]. In urban areas, the types of surface targets are complex. The primary task of extracting water is to remove built-up land and shadows of buildings. Based on Normalized Difference Built-up Index (NDBI), a method was proposed to automate extracting built-up areas for mapping [16]. So NDBI can be used for removing built-up land from water bodies. After the resolution fusion of QuickBird PAN and Multispectral images, Xu et al. set up a repository including spectral features, topological rules, shape features, size information and so on to extract water types [17]. An automatic method was proposed which combined the use of the multiscale extraction and the SMA technique to extract water in urban was more accurate compared to satellite images of moderate resolution [18].

Most of the mentioned methods need to be given a suitable threshold. The accuracy of abstracting water has more to do with this threshold [19]. If the threshold is too low, other objects may be extracted incorrectly. On the contrary, small linear water area may not be

extracted. To extract water more precisely, pattern recognition methods have been widely used in classification of Landsat Imagery [20-22]. The most commonly adopted methods of pattern recognition are Support Vector Machines (SVM) and traditional neural networks (NN). In 2008, Dixon used Support Vector Machines (SVMs) and artificial neural network (ANN) to classify land use from remote sensing images. Zhang et al. classified spectrally uniform water bodies with a distinct boundary in a salt farm by neural network and maximum likelihood classifiers. He found that neural network is superior to maximum likelihood method for detailed mapping of salty water bodies [23]. Liu et al. proposed a novel semi-supervised SVM model by using self-training approach to address the problem of remote sensing land cover classification from Landsat TM images [24].

In 2006, Hinton proposed a new approach which accomplished the dimensionality reduction and classification of data by means of Deep Belief Network (DBN) [25-30]. And it serves as a catalyst for a new wave of Deep Learning (DL) in machine learning. Since then, Deep Learning has become a burgeoning research direction and was applied to image recognition, natural language processing, speech recognition and information retrieval. In the field of remote sensing, Deep Learning establishes a compelling rationale for remote images' classification. In 2010, Deep Learning was utilized on detecting roads in high-resolution aerial images for the first time [31]. Therewith, Lv et al. put forward a remote sensing image classification method based on DBN model [32]. That was one of the first attempts to urban detailed classification by Deep Learning approach. Up to now, however, we have not found related papers which aim specially at water body extraction by Deep Learning approach.

Stacked autoencoder is one of Deep Learning methods and has a strong personality to learn higher features from the lower level, due to which it is widely used in various areas [33-35]. Thus, we proposed a model for extracting water bodies from Landsat images based on stacked autoencoder. And a Feature Expansion Algorithm (FEA) was also proposed to enrich the features of water bodies. Selecting the Landsat 8 images about three areas as input, the proposed model had been validated and evaluated. A series of experiments were carried out to show that our model could extract water bodies from other objects more effectively than SVM and NN. In particular, the constructed model established the feasibility of learning features of water body with limited or few labeled samples. It is significant for easing the work of collection, dispose and field measurements of training data. The description of the whole model is shown in Figure 1.

2. Stacked Sparse Autoencoder.

2.1. **Sparse autoencoder.** An autoencoder neural network is an unsupervised learning algorithm which utilizes back-propagation algorithm, letting the target values equal to the inputs, such as $y^{(i)} = x^{(i)}$ [36]. A typical structure of autoencoder is shown in Figure 2.

The autoencoder tries to learn a function $h_{w,b}(x) \approx x$. Like Figure 2, it attempts to learn an approximation to the identity function, namely, making output $\hat{x}$ be similar to x. And by placing constraints on the network, connotative structure of the data will be discovered.

For instance, we can limit the number of hidden units to explore connotative features. As a typical example, assume the inputs x are the gray value of pixel for a 20×20 image, so $n = 400$. Meanwhile, we can set the number of hidden layer to 100, which forces the network to learn a condensed representation of the input which contains a vector of 100 dimensions instead of a 400-pixel input. In this way, the algorithm will find some
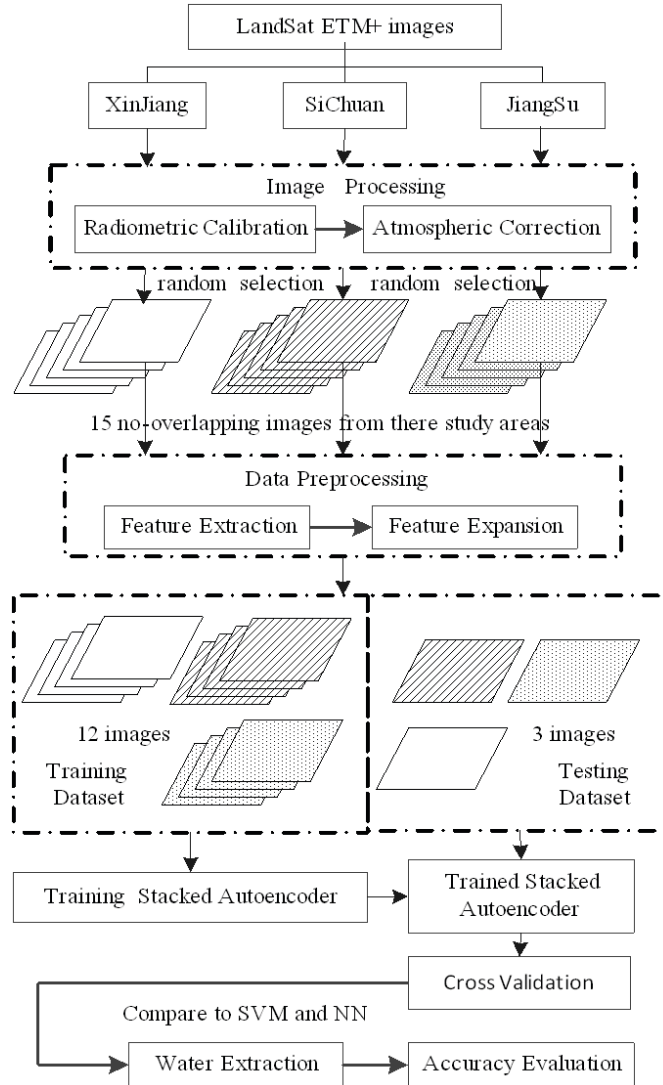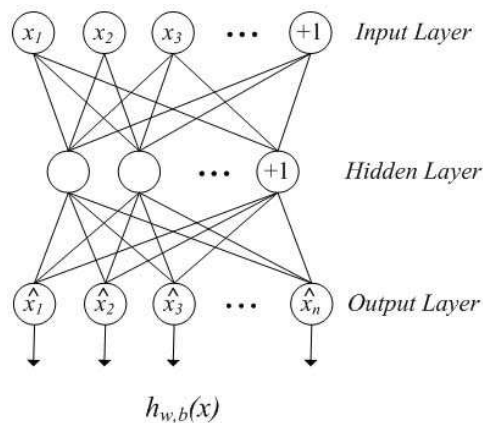
FIGURE 1. Flowchart of the whole model

FIGURE 2. Structure of an autoencoder

relevance of the input features. To sum up, an autoencoder can learn a low-dimensional representation of data which is exactly similar to PCAs (principal component analysis).

However, when the number of hidden units is large (perhaps even greater than the number of input pixels), other constraints must be imposed on the network. On this occasion, we add a sparsity constraint on the hidden units. We usually consider a neuron as being "active" if its output value is close to 1 or "inactive" if its output value is close to 0. And the sparsity constraint causes the neuron to be inactive most of the time. Select sigmoid as the activation function. Assume that $a_j^{(2)}$ signifies the activation of hidden unit j in the autoencoder. And we let $a_j^{(2)}(x)$ denote the activation of hidden unit for a given specific input x. Next, let $\hat{\rho}_j$ (Equation (1)) be the average activation of hidden unit j (take the average on the training set).

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} \left[a_j^{(2)} \left(x^{(i)} \right) \right] \tag{1}$$

Then we set ρ as the anticipant degree about sparsity of hidden layer. To each j, the following equation should be satisfied.

$$\hat{\rho}_j = \rho \tag{2}$$

where, ρ is called sparsity parameter, which is a small value close to zero. In order to achieve the expected sparsity, the activations of hidden unit must be near to 0. So an extra penalty term will be taken into account that does the penalty for some situation where $\hat{\rho}_j$ has great difference with ρ. The penalty term can be defined in Equation (3).

$$\sum_{j=1}^{s_2} \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{(1 - \rho)}{(1 - \hat{\rho}_j)} \tag{3}$$

Here, s_2 stands for the number of neurons in hidden layer, and each j corresponds a neuron in hidden layer. Combining with knowledge of Kullback-Leibler divergence (KL divergence), the above equation is able to be rewritten:

$$\sum_{j=1}^{s_2} KL \left(\rho || \hat{\rho}_j \right) \tag{4}$$

Here, $KL \left(\rho || \hat{\rho}_i \right) = \sum_{j=1}^{s_2} \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{(1-\rho)}{(1-\hat{\rho}_j)}$. If $\hat{\rho}_j = \rho$, $KL(\rho||\hat{\rho}_j) = 0$. Meanwhile, $KL(\rho||\hat{\rho}_j)$ increases monotonically as $\hat{\rho}_j$ diverges from ρ. Therefore, the overall cost function of network is shown as follows, where β controls the weight of the sparsity penalty term [2].

$$J_{sparse(W,b)} = J(W,b) + \beta \sum_{j=1}^{s_2} KL \left(\rho || \hat{\rho}_j \right) \tag{5}$$

2.2. **Softmax regression.** Softmax regression is a supervised algorithm that is often used for multi-class classification [37,38]. For a given testing dataset $\left\{ \left(x^{(1)}, y^{(1)} \right), \ldots, \left(x^{(m)}, y^{(m)} \right) \right\}$, the hypothesis $h_\theta \left(x^{(i)} \right)$ and cost function $J(\theta)$ of softmax regression take the form:

$$h_\theta \left(x^{(i)} \right) = \begin{bmatrix} p \left(y^{(i)} = 1 \right) | x^{(i)}; \theta \\ p \left(y^{(i)} = 2 \right) | x^{(i)}; \theta \\ \vdots \\ p \left(y^{(i)} = k \right) | x^{(i)}; \theta \end{bmatrix} = \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{6}$$

$$J\left(\theta\right) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\left\{y^{(i)} = j\right\}\log\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right] \tag{7}$$

In the hypothesis function, θ is the parameter of the model. It is common to set θ as a k-by-$(n+1)$ matrix, so that

$$\theta = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ \vdots \\ \theta_k^T \end{bmatrix} \tag{8}$$

2.3. **Stacked autoencoder.** A stacked autoencoder is a deep neural network which consists of multiple layers of sparse autoencoders. It employs greedy layer-wise training method to execute pre-training. The outputs of each layer serve as inputs of the successive layer. In Figure 3, a stacked autoencoder with two hidden layers and a final softmax classifier layer is used for water extraction.
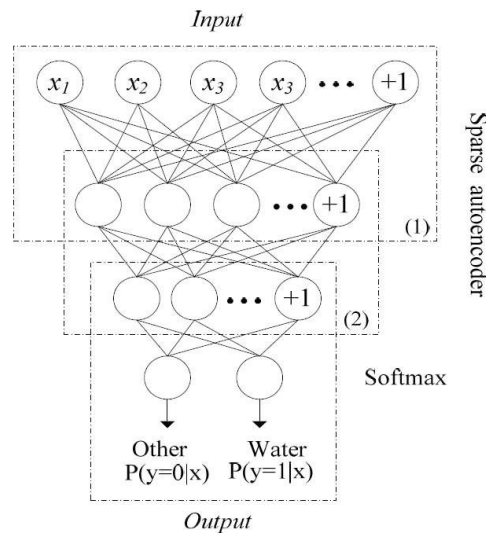
FIGURE 3. The structure of a stacked autoencoder for water body extraction

From top to bottom, the first sparse antoencoder is trained on the raw inputs to learn the primary features firstly. Then these outputs from the first sparse autoencoder would be regarded as inputs to the second sparse autoencoder. As well, secondary features are learned on afferent primary features. What is more, secondary features are obtained through the trained second sparse autoencoder. And these secondary features would act as "raw inputs" to a softmax classifier for training. Finally, a three-layer model which is able to classify water body and other object is established.

3. **Method.**

3.1. **Study areas.** Three study areas in China with different water body types were chosen to evaluate the robustness of the proposed method. Each of them possesses their typical regional features. Table 1 shows the details of the three study areas.

The first area is called Xinjiang (Xj) which covers the area of $73°40'E \sim 96°18'$E, $34°25' \sim 48°10'$N. It is located in Eurasia inland and shows up a typical half-arid, arid climate. And there are about 500 rivers in Xinjiang, including Tarim River, Ebinur Lake, Yili River and so on. Xinjiang possesses Taklimakan desert which is the largest desert in

TABLE 1. The view of the study areas and data source

No	Study Area	Path/Row	Acquisition Date	Climate
1	Xinjiang	140/34	7 June 2014	half-arid, arid
2	Jiangsu	131/36	23 July 2013	humid
3	Sichuan	118/39	13 June 2014	humid

China. Therefore, the key to enhance extractable accuracy of water body is the distinction between water and arid soil.

Jiangsu (Js), the second area, covers the area of $116°18'E \sim 121°57'E$, $30°45'N \sim 35°20'N$. Jiangsu is located in the middle of the coastal provinces in East China, and also lies in the lower reaches of the Yangtze River and Huaihe River. Thus, it has abundant water resources. In 2013, the Yangtze River delta urban agglomeration (including Jiangsu, Shanghai and Zhejiang) becomes a number of the six word-class international urban agglomeration. Apparently, we should put our center of gravity on distinguishing shadows of buildings from water body.

The last area is Sichuan (Sc), which covers the area of $97°21'E \sim 108°33'E$, $26°03'N \sim 34°19'N$. Sichuan is a mountainous province. The east of Sichuan includes Ridge and Valley Province of Chuandong and Hills of Chuanzhong; the west is West Sichuan Plateau. Thus, the shadows of mountains must be removed from water body.

3.2. **Data preprocessing.** The whole process of data preprocess is divided into two steps. First, radiometric calibration and atmospheric correction are executed for each image. Then for each study area, five images are selected randomly for experiments. The last column describes the corresponding validation samples for each image. And the whole validation samples were obtained by executing supervised classification and likelihood classification on high-resolution images.

3.3. **Feature extraction.** Water possesses a stronger ability to absorb sunlight than other objects. So it has a weaker reflectance. Especially in NIR and MIR, the reflectance of water is almost zero [1]. Aiming to learn and master the characteristics of water, a variety of methods are proposed by several scholars, including Water Index Method (WI), IHS (Intensity, Hue, and Saturation) Transformation and so on. These methods provide a probability to obtain more precise extraction results. Far from increasing, however, the extract accuracy will decrease with inappropriate methods. Hence, we analyzed all the useful methods (each method is regarded as a feature, such as MNDWI, and NDWI). Some typical methods of water body extraction were shown in Table 3.

Then for each pixel, well-chosen features are computed to construct a feature matrix. Figure 4 shows the mapping relationship from pixel to feature matrix.
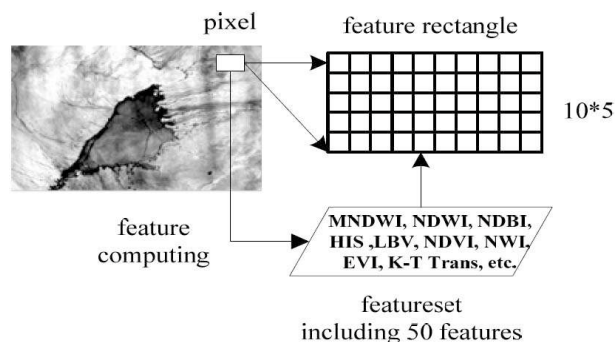

FIGURE 4. The view of a pixel and its homologous feature rectangle

L. YANG, S. TIAN, L. YU, F. YE, J. QIAN AND Y. QIAN

TABLE 2. View of the selected images of study areas
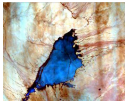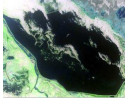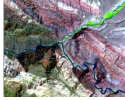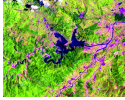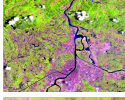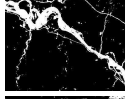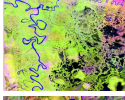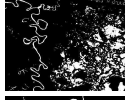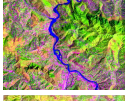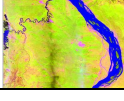
No	Name	Width × Height	Pixels	Image	Validation Samples
1	Xj_1	375×310	116250		
2	Xj_2	669×342	228798		
3	Xj_3	370×200	74000		
4	Xj_4	400×302	120800		
5	Xj_5	277×208	57616		
6	Js_1	400×375	150000		
7	Js_2	434×325	141050		
8	Js_3	362×304	110048		
9	Js_4	482×393	189426		
10	Js_5	579×554	320766		
11	Sc_1	403×299	120497		
12	Sc_2	392×397	156418		
13	Sc_3	373×338	126074		
14	Sc_4	352×309	108768		
15	Sc_5	408×375	153000		

TABLE 3. The elaborate description about typical methods of water body extraction

Methods	Examples	Equation	Advantages
Single-band threshold	Band 5	Gray value < threshold	Simple and easy to implement
Multiband spectral relationship	b2, b3, b4, b5.	(b2 + b3) > (b4 + b5)	It can extract small linear water area better than single-band method.
Water index	NDVI	(b4 − b3)/(b4 + b3)	It can expand the difference between water, vegetation and soil.
	NDWI	(b2 − b4)/(b2 + b4)	Restrains the information of vegetation and highlight water body's information.
	MNDWI	(b2 − b5)/(b2 + b5)	It is good at extracting water body in urban environments.
	NDBI	(b5 − b4)/(b5 + b4)	The difference between water and building is expanded.
Spectrum transform	K-Thomas transformation	It can show six parts, such as soil brightness, greenness, and yellow stuff.	It eliminates the influence of shadow from cloud.
	HIS	It is able to extract Hue, Intensity and Saturation from images of remote sensing.	It is suitable for extracting small linear water bodies.

3.4. **Feature Expansion Algorithm.** Each image can be regarded as numerous grids consisting of pixels. In an image, each pixel is not self-existent. And adjacent pixel points may have some effects on target pixels. Thus, we proposed a Feature Expansion Algorithm (FEA) to dig neighborhood information. First of all, each grid was mapped into a pixel point. Next, we chose a point as a target point. Meanwhile, we drew a circle whose center and radius are respectively target point and r. Then the distance between object point named o and adjacent point named $o_{i,j}$ was computed. Afterwards, for each target point, we got a related dataset which was made up of adjacent points where $dist\,(o, o_{i,j}) \leq r$. Finally, for given features, we got the mean of the related dataset as a new feature. Meanwhile, we can adjust the length of r to find the best expansion distance. As is showed in Figure 5, there are some situations in which $r = 1$, $\delta = 0.5$. Among the following pictures, related datasets comprised some grids with shadow. The detailed process and pseudo-code are described in Table 4.

In order to get better effects, we added weight coefficient to Feature Expansion Algorithm, namely related rank (RR). As is shown in Figure 6, we set different values for RR according to the distances between target point and its adjacent points. The nearest point is given the highest related rank with lower weight for distant point on the contrary.

Therefore, the detailed pseudo-code of modified feature expansion is showed in Table 5.

(a) when $r = 1$　　　　(b) when $r = 1.5$

(c) when $r = 2$　　　　(d) when $r = 2.5$
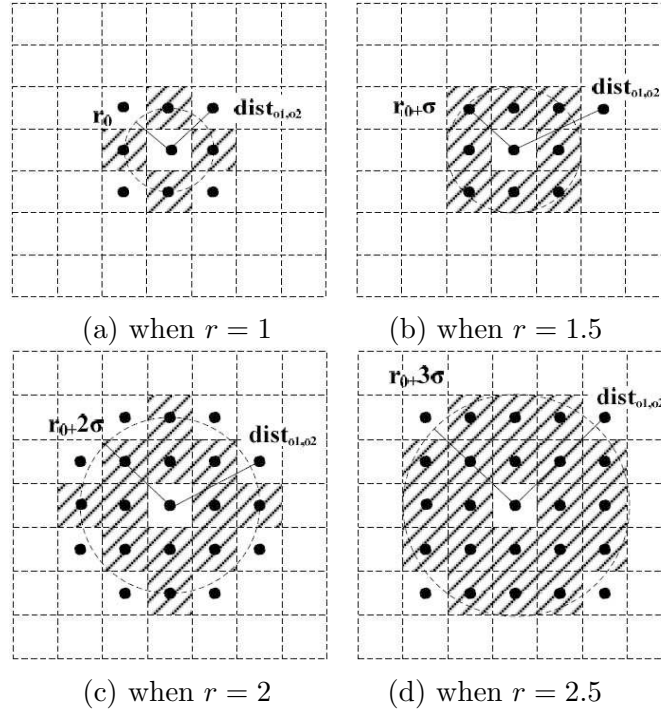
FIGURE 5. Four examples of the Feature Expansion Algorithm

TABLE 4. Key codes of the Feature Expansion Algorithm

Algorithm: Feature Expansion
Input: Feature matrix $f[][]$
Output: Mean of new feature matrix $newf[][]$
1. Confirm a point $f[a][b]$ as a target point named o;
2. As the centre of o, draw a circle with given radius $(r + \delta)$.
3. Compute distance between o and its adjacent points $o_{i,j}$
4. for each $o_{i,j}$
5.   if $(dist\,(o, o_{i,j}) \leq r + \delta)$
6. $total+ = f[i][j]$;
7.   count $+ +$;
8.   end if
9. $mean = total/count$;
10.   if $(dist\,(o, o_{i,j}) \leq r + \delta)$
11. $total+ = (f[i][j] - mean)^2$;
12.   num $+ +$;
13.   end if
14. end for each
15. Return Mean

4. **Experiments and Results Analysis.** Based on stacked sparse autoencoder (SSAE), a water body extraction model was established. Through several experiments, parameters from Table 6 were set for the initiation of sparse autoencoder and softmax. And we chose MATLAB (Matrix & Laboratory) for simulation experiments and adopted Feature Extraction Algorithm to enhance accuracy.

4.2	3.6	3.2	3	3.2	3.6	4.2
3.6	2.8	2.2	2	2.2	2.8	3.6
3.2	2.2	1.4	1	1.4	2.2	3.2
3	2	1	▨	1	2	3
3.2	2.2	1.4	1	1.4	2.2	3.2
3.6	2.8	4	2	4	2.8	3.6
4.2	3.6	3.2	3	3.2	3.6	4.2

9	8	7	6	7	8	9
8	5	4	3	4	5	8
7	4	2	1	2	4	7
6	3	1	▨	1	3	6
7	4	2	1	2	4	7
8	5	4	3	4	5	8
9	8	7	6	7	8	9

FIGURE 6. The contrast of related rank and distance

TABLE 5. The key codes of modified Feature Expansion Algorithm

---

Algorithm: $ComputeRR\left(dist\left(o,o_{i,j}\right),weight,\delta\right)$

---

1. for each $o_{i,j}$
2.   if $(dist\left(o,o_{i,j}\right) \leq r+\delta)$
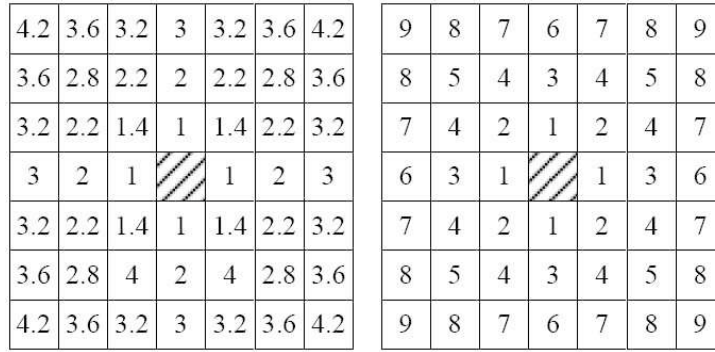3.   if $(dist\left(o,o_{i,j}\right) \leq 1 \&\& dist(o,o_{i,j}) > 0)$
4. $total+ = weight * f[i][j]$;
5. $count + +$;
6.   end if
7.   if $(dist\left(o,o_{i,j}\right) \leq 1.5 \&\& dist(o,o_{i,j}) > 1)$
8. $total+ = (weight+\delta) * f[i][j]$;
9. $count + +$;
10.   end if
11.   Other situation can be done in the same manner ...
12.   end if
13. $mean = total/count$;
14. end for each

TABLE 6. The parameter initialization for SSAE

Parameter Initialization for SAE	
Input Size	50
Hidden Size	50
Sparsity parameter	0.1
Lambda	$3e^{-3}$
Beta	5
Iterations	800
Parameter Initialization for Softmax	
Lambda	$1e^{-4}$
Iterations	900

To present a detailed evaluation of our extraction model for water body, four experiments were designed to validate the availability and performance of this model. The details of the experiments and result analysis were shown below.

4.1. **Training and testing dataset preprocessing.** As shown in Table 2, fifteen images were selected uniformly from three different study areas. At first, four images from

TABLE 7. View of the experimental dataset

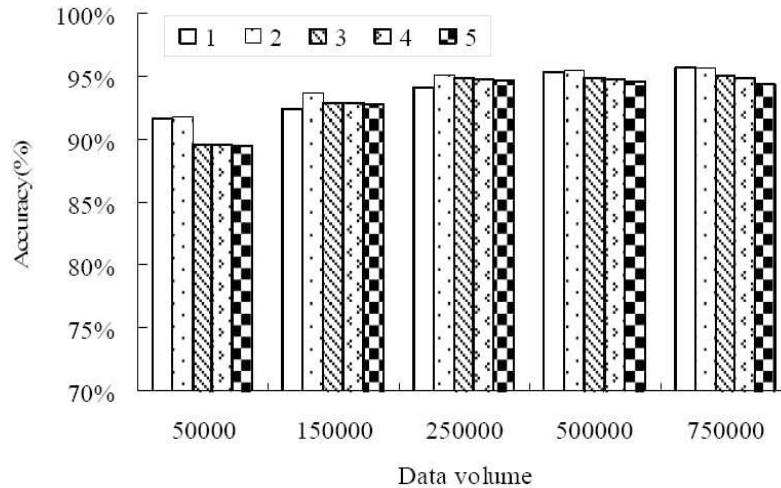| No	Data Volume	Training Data	Testing Data
1	50000	40000	10000
2	150000	120000	30000
3	250000	200000	50000
4	500000	400000	100000
5	750000	600000	150000

FIGURE 7. The influence of the number of layers

each study area were chosen to make up a dataset for training. The rest of images were used for testing. Then we get unique feature matrixes one by one through the process of feature extraction. As shown in Table 7, there were several datasets of different volumes for experiments.

4.2. **The impacts of different number of layers on extraction accuracy.** The number of layers about stacked sparse autoencoder (SSAE) has more to do with accuracy. Thus, an experiment is executed to find the optimal number of layers. The experiment results were shown in Figure 7.

When the layer number ranged from one to two, extraction accuracy was enhanced. However, the accuracy had a declining trend when the number was greater than 2. Obviously, the proposed model got the highest accuracy when the number of layers was 2. With the increase of data volume, the whole accuracy raised steady. When the volume of dataset was 750000, accuracy increased to 95.74%. Therefore, we set the number of layers of sparse autoencoder to 2.

4.3. **Comparison with SVM and NN.** To evaluate the performance of the proposed model, we selected two frequently-used models as contrasts. Before the experiment, several testing experiments were carried out to find suitable parameters for SVM and NN.

Figure 8 showed the results of the experiment based on which several conclusions were summarized:

(1) From high to low, the accuracy of three models can be arranged as: SSAE, NN and SVM, whose average accuracies are 90.32%, 91.97%, 94.35%, respectively. The proposed model has marked unique advantages that can learn higher level features from the lower ones. It can select not much but optimal features which may not be found from the
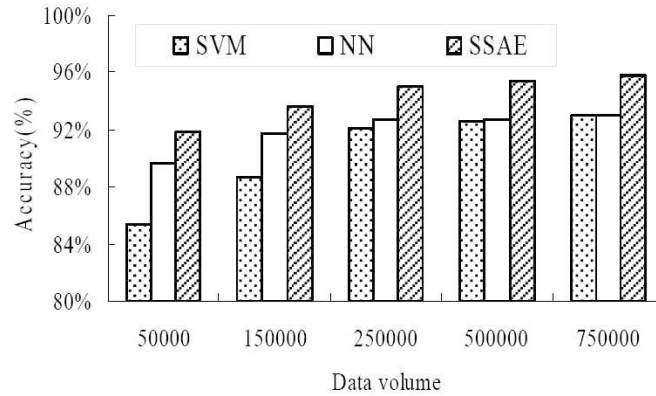
FIGURE 8. The influence of different number of layers

TABLE 8. Initialization parameters of SAE and softmax

Parameter Initialization for SAE	
Input Size	100
Hidden Size	100
Sparsity param	0.1
Lambda	$3e^{-3}$
Beta	5
Iterations	800
Parameter Initialization for Softmax	
Lambda	$1e^{-4}$
Iterations	900

given features; (2) Extraction accuracy is under 90% except the proposed model when data volume is 50000. From a small amount of dataset, SVM and NN cannot learn full-scale features of the whole water types. Other than surface learning approach, SSAE adopts greedy layer-wise training method. The original training data is mapped into a new feature space by means of layer-by-layer feature transformation. These new features express substantive characteristics of water preferably. As we can see that the accuracy of our model is 6.51% higher than SVM; (3) Extraction accuracy maintains a steady increasing trend along with the growth of data volume. Through large amount of data, further features are learned to extract water precisely.

4.4. **The impacts of the size of radius on extraction accuracy.** Considering the neighborhood characteristics for every pixel, we put forward a Feature Expansion Algorithm. Here, we re-process the former dataset and expand the number of features from 50 to 100. The initialization of parameters of our model is shown in Table 8.

For Feature Expansion Algorithm, the size of radius has effects on extractable accuracy. Thus, we design an experiment to find the best size of radius.

Apparently, in Figure 9, with the increase of radius, the accuracy keeps up a decreasing tendency. This phenomenon corresponds to our strategy about setting rank for neighbor pixels in the light of pixels' position. The farther the distance of adjacent pixels is, the less the target pixel is affected. Therefore, we expand the former 50 features to 100 by making use of Feature Expansion Algorithm.
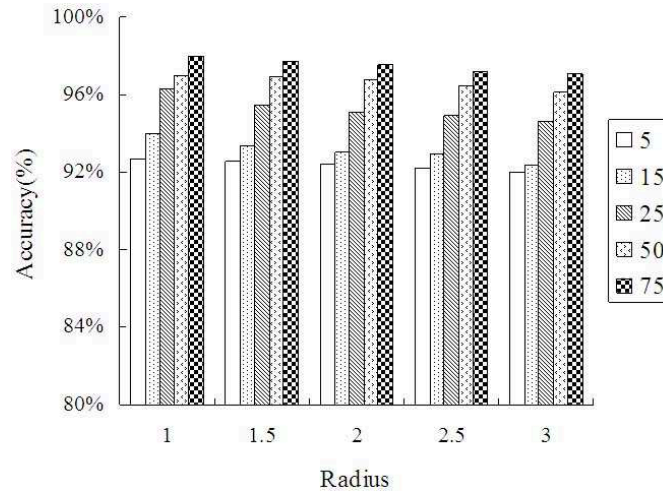
FIGURE 9. The influence of the size of radius

TABLE 9. The increases of accuracy using FEA

Data Volume	SVM	NN	SSAE	Total
50000	2.08%	0.59%	0.9%	3.57%
150000	2.13%	1.18%	1.0%	4.31%
250000	2.18%	2.17%	1.24%	5.59%
500000	2.3%	2.51%	1.54%	6.35%
750000	2.32%	2.6%	2.24%	7.16%
Average	2.20%	1.81%	1.38%	5.39%

4.5. **Qualitative evaluation of feature expansion algorithm.** A new experiment was conducted on the new dataset for qualitative evaluation of feature expansion algorithm. Similarly, SVM and NN model were regarded as competitors. In the experiment, each original feature matrixes was expanded to a new matrix with 100 features. Under different volumes of dataset, five line charts are delineated in Figure 10. In Figure 10, "50" stands for the number of the original features, and "100" is the number of the new feature matrix.

In Figure 10, from (a) to (e), we conclude that the Feature Expansion Algorithm has better ability to describe the characteristics of water body. Among the results, the highest accuracy is 97.98% (using 50 features, the corresponding accuracy is 95.54%). Table 9 expresses the situation of increases in detail.

5. **Conclusions.** Water body extraction is still confronted with a long-standing challenge in removing various shadows (including shadows of mountains, buildings and clouds, etc.) and eliminating noise within water. To figure out that, we proposed a model for water body extraction based on the stacked sparse autoencoder. Many features were concatenated for training model. Meanwhile, we put forward a Feature Expansion Algorithm (FEA) to dig more characteristics for water body.

The experiment results showed that the proposed model has outstanding capability of feature learning and gets better results by contrasting with the Support Vector Machine (SVM) and traditional neural network (NN). In particular, because of unsupervised feature learning, the proposed model provides a lead for using limited number or few training samples and getting higher accuracy. It makes much sense to decrease the cost of preparing training data. Getting accurate information of water body by the proposed model,

(a) Data volume: 50000

(b) Data volume: 150000

(c) Data volume: 250000

(d) Data volume: 500000
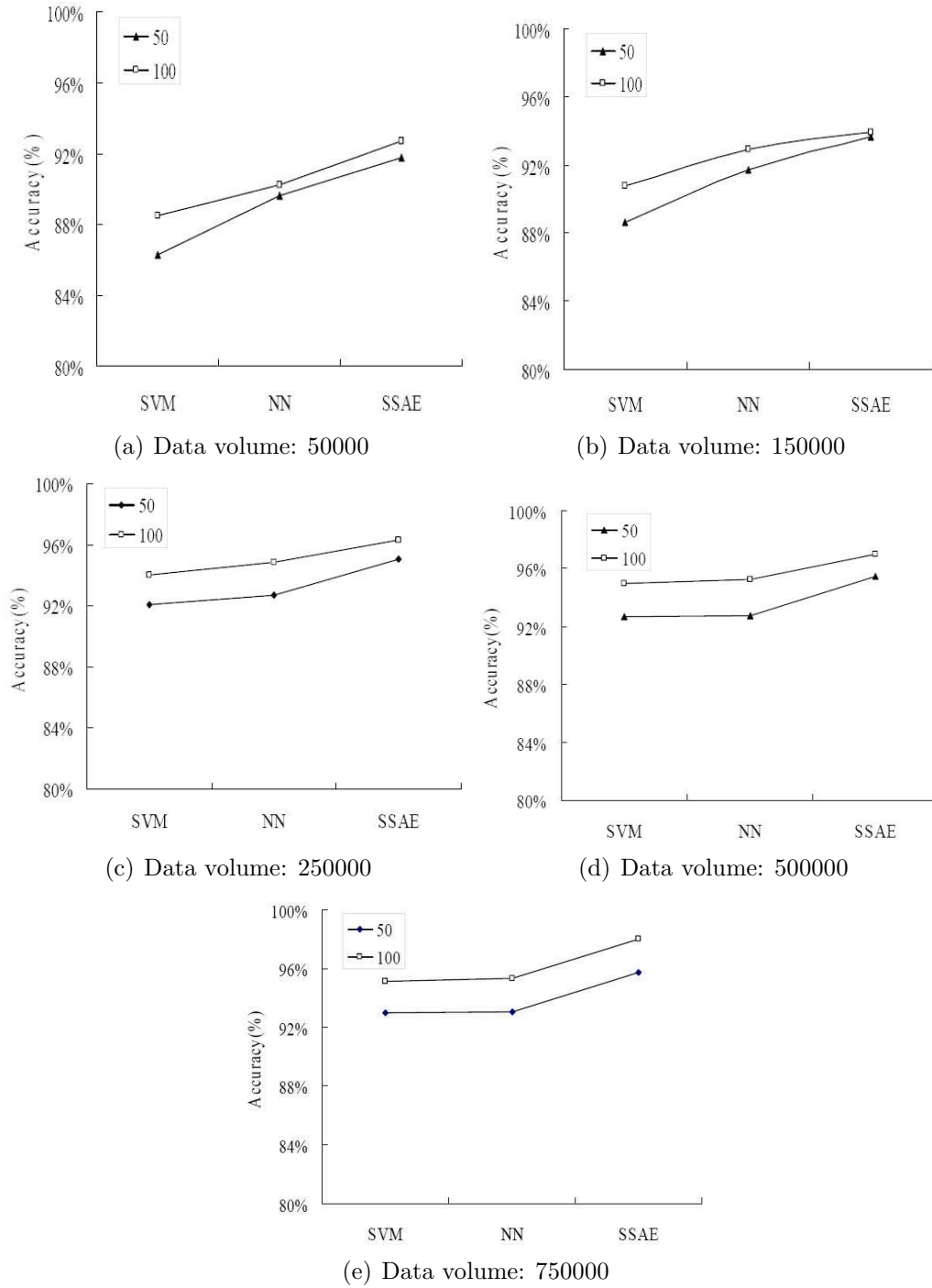

(e) Data volume: 750000

FIGURE 10. The contrast of SVM and NN using Feature Expansion Algorithm on the different volumes of dataset

we can monitor the status of water resources fast and timely. It is of great immediate significance for protecting the environment and achieving sustainable development.

REFERENCES

[1] Y. N. Wang, F. Huang and Y. C. Wei, Water body extraction from LANDSAT ETM+ image using MNDWI and KT transformation, *The 21st International Conference on Geoinformatics*, Kaifeng, China, 2013.

[2] C. Senaras, E. Gedik and Y. Yardimci, A novel dynamic thresholding and categorizing approach to extract water objects from VHR satellite images, *Geoscience and Remote Sensing Symposium (IGARSS)*, Quebec City, Canada, 2014.

[3] H. Jiang, M. Feng, Y. J. Zhu, N. Lu, J. X. Huang and T. Xiao, An automated method for extracting rivers and lakes from landsat imagery, *Remote Sensing*, vol.6, no.6, pp.5067-5089, 2014.

[4] H. F. Chen, J. L. Wang, Z. Chen, L. Yang and W. Xi, Comparison of water extraction methods in mountainous plateau region from TM image, *Remote Sensing Technology and Application*, vol.19, no.6, pp.479-484, 2004.

[5] D. L. B. Jupp, *Landsat Based Interpretation of the Cairns Section of the Great Barrier Reef Marine Park, Natural Resources Series/Division of Water and Land Resources*, Commonwealth Scientific and Industrial Research Organization, 1985.

[6] L. Moller-Jensen, Knowledge-based classification of an urban area using texture and context information in landsat TM imagery, *Photogrammetric Engineering and Remote Sensing*, pp.899-904, 1990.

[7] J. J. Lu and S. H. Li, Improvement of the techniques for distinguishing water bodies from TM data, *Remote Sensing of Environment*, vol.7, no.1, pp.17-23, 1992.

[8] D. H. Braud and W. Feng, Semi-automated construction of the Louisiana coastline digital land/water boundary using landsat thematic mapper satellite imagery Louisiana applied oil spill research and development program, *OSRAPD Technical Report Series*, vol.97, no.2, 1998.

[9] I. J. Barton and J. M. Bathols, Monitoring floods with AVHRR, *Remote Sensing of Environment*, vol.30, no.1, pp.89-94, 1989.

[10] Y. Y. Du and C. H. Zhou, Automatically extracting remote sensing information for water bodies, *Journal of Remote Sensing*, vol.2, no.4, pp.264-269, 1998.

[11] J. Wang, Y. Zhang and G. Kong, The application of multi-band spectral relationship method in water body extraction, *Mine Surveying*, vol.4, pp.30-32, 2004.

[12] S. K. McFeeters, The use of normalized difference water index (NDWI) in the delineation of open water features, *International Journal of Remote Sensing*, vol.17, no.7, pp.1425-1432, 1996.

[13] H. Xu, Fast information extraction of urban built-up land based on the analysis of spectral signature and normalized difference index, *Geographical Research*, vol.9, no.5, pp.589-595, 2005.

[14] J. Shen, L. Yang and X. Chen, A method for object-oriented automatic extraction of lakes in the mountain area from remote sensing image, *Remote Sensing for Land & Resources*, vol.3, pp.84-87, 2012.

[15] W. Wang, X. Zhu and Y. Sun, Water extraction method based on ETM image of mountain, *Journal of System Simulation*, vol.9, no.25, pp.2196-2205, 2013.

[16] Y. Zha, J. Gao and S. Ni, Use of normalized difference built-up index in automatically mapping urban areas from TM imagery, *International Journal of Remote Sensing*, vol.24, no.3, pp.583-594, 2003.

[17] T. Xu, Z. Tan and X. Yan, Extraction techniques of urban water bodies based on object-oriented, *Geospatial Information*, vol.3, no.8, pp.64-66, 2010.

[18] Y. N. Zhou, J. C. Luo, Z. F. Shen and X. D. Hu, Multiscale water body extraction in urban environments from satellite images, *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol.7, no.10, pp.4301-4312, 2014.

[19] X. Jian, H. Chen, Z. Y. Xing, T. Fang and L. Y. Yin, Effect of intercalated LDHs on the mechanical and thermal stability of the epoxy/cyanate nanocomposites, *Science Technology and Engineering*, vol.14, no.18, pp.267-270, 2014.

[20] W. X. Sun, V. Heidt, P. Gong and G. Xu, Information fusion for rural land-use classification with high-resolution satellite imagery, *IEEE Trans. Geosci. Remote Sens.*, vol.41, no.4, pp.883-890, 2003.

[21] Z. X. Qi, A. G. O. X. Yeh, X. Li and Z. Lin, A novel algorithm for land use and land cover classification using RADAR-SAT-2 polarimetric SAR data, *Remote Sensing of Environment*, pp.21-39, 2012.

[22] B. Dixon and N. Candade, Multispectral landuse classification using neural networks and support vector machines: One or the other, or both? *International Journal of Remote Sensing*, vol.29, no.4, pp.1185-1206, 2008.

[23] Y. Zhang, J. Gao and J. Wang, Detailed mapping of a salt farm from landsat TM imagery using neural network and maxi-mum likelihood classifiers: A comparison, *International Journal of Remote Sensing*, vol.28, no.10, pp.2077-2089, 2007.

[24] Y. Liu, B. Zhang and L. M. Wang, A self-trained semisupervised SVM approach to the remote sensing land cover classification, *Computers & Geosciences*, pp.98-107, 2013.

[25] G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science*, vol.313, no.5786, pp.504-507, 2006.

[26] D. Yu and L. Deng, Deep learning and its applications to signal and information processing [exploratory DSP], *Signal Processing Magazine*, vol.28, no.1, pp.145-154, 2011.

[27] I. Arel, D. C. Rose and T. P. Karnowski, Deep machine learning – A new frontier in artificial intelligence research [research frontier], *Computational Intelligence Magazine*, vol.5, no.4, pp.13-18, 2010.

[28] X. Hu and J. Zhu, Deep learning: New hotspot in machine learning, *Communications of the CCF*, vol.9, no.7, pp.64-69, 2013.

[29] Y. Bengio, A. Courville and P. Vincent, Representation learning: A review and new perspectives, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.35, no.8, pp.1798-1828, 2013.

[30] K. Yu, L. Jia, Y. Q. Chen and W. Xu, Deep learning: Yesterday, today, and tomorrow, *Journal of Computer Research and Development*, vol.50, no.9, pp.1799-1804, 2013.

[31] V. Mnih and G. E. Hinton, Learning to detect roads in high-resolution aerial images, *Computer Vision – ECCV 2010, Lecture Notes in Computer Science*, vol.6316, pp.210-223, 2010.

[32] Q. Lv, Y. Dou, X. Niu, J. Q. Xu and F. Xia, Remote sensing image classification based on DBN model, *Journal of Computer Research and Development*, vol.51, no.9, pp.1911-1918, 2014.

[33] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio and P. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, *The Journal of Machine Learning Research*, pp.3371-3408, 2010.

[34] H. C. Shin, M. R. Orton, D. J. Collins, S. J. Doran and M. O. Leach, Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.35, no.8, pp.1930-1943, 2013.

[35] J. Deng, Z. Zhang, E. Marchi and B. Schuller, Sparse autoencoder-based feature transfer learning for speech emotion recognition, *Humaine Association Conference on Affective Computing and Intelligent Interaction*, Geneva, 2013.

[36] N. Andrew, Sparse autoencoder, *CS294 A Lecture Notes*, vol.72, 2011.

[37] K. Gimpel and N. A. Smith, Softmax-margin CRFs: Training log-linear models with cost functions, *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp.733-736, 2010.

[38] A. Tuerk, Implicit softmax transforms for dimensionality reduction, acoustics, *IEEE International Conference on Speech and Signal Processing*, Las Vegas, 2008.